An Asymmetric Protocol for Argumentation Games in Defeasible Logic

Jenny Eriksson Lundström[1], Guido Governatori[2], Subhasis Thakur[2] and Vineet Padmanabhan[3]

[1] Dept. of Information Science, Computer Science Division, Uppsala University, Sweden
[2] School of ITEE, The University of Queensland, Brisbane, Australia
[3] Dept. of Computer & Information Sciences, University of Hyderabad, India
email: jenny.eriksson@dis.uu.se, {guido, subhasis}@itee.uq.edu.au, vineetcs@uohyd.ernet.in

Abstract. Agent interactions where the agents hold conflicting goals could be modelled as adversarial argumentation games. In many real-life situations (e.g., criminal litigation, consumer legislation), due to ethical, moral or other principles governing interaction, the burden of proof, i.e., which party is to lose if the evidence is balanced [21], is *a priori* fixed to one of the parties. Analogously, when resolving disputes in a heterogeneous agent-system the unequal importance of different agents for carrying out the overall system goal need to be accounted for. In this paper we present an asymmetric protocol for an adversarial argumentation game in Defeasible Logic, suggesting Defeasible Logic as a general representation formalism for argumentation games modelling agent interactions.

1 Introduction

Adversarial situations arise when agents in pursuit of their goals interact with other agents pursuing goals of a conflicting nature. In a setting where issues need to be resolved, the agent interaction could be modelled as an adversarial argumentation game. Argumentation games are defeasible, meaning that an argument put forward by one of the agents in support of a conclusion could be defeated by contrary evidence and arguments put forward by the other agent. Thus, the agents are to resolve the dispute by putting forward the arguments that will enable the best outcome for their cases. Using a symmetrical protocol an argumentation game between homogeneous (equally strong) parties could be modelled. However, as in many real-life settings, also in agent systems disputes arise where the claims of the agents involved in the interaction (e.g., regarding distribution of a scarce resource) are of unequal importance to the overall goal of the agent system, and thus need to be handled accordingly. In addition, as in many real-life situations the evidence presented by the parties of a dispute may be inconclusive and the accompanying arguments incoherent. Thus, a majority of the disputes has to be resolved by higher-level principles guiding the interaction. One important principle is referred to as the burden of proof, *cf. e.g.* [21].

To accommodate for a correct outcome of argumentation games in heterogeneous agent systems, we present an asymmetric protocol for adversarial argumentation games.

A. Ghose and G. Governatori (eds)
10 Pacific Rim International Workshop on Multi-Agents (PRIMA 2008)
LNAI 5044, 2008.

The paper is organized as follows: In Section 2 we present argumentation games and their setup. In Section 3 we highlight the most relevant features of defeasible logic. We discuss the formalization of argumentation games using defeasible logic as presented in [24] in Section 4. Section 5 presents the asymmetrical protocol for argumentation games. We use a criminal litigation setting to illustrate and discuss some of the benefits of the model. Section 6 presents some related work. In Section 7 we conclude.

2 Argumentation Games

Consider an adversarial argumentation (dialogue) game as an interaction between two parties, the *Proponent* and the *Opponent*. The two parties debate over a topic. Each equipped with a set of arguments the parties take turn in putting forward a subset of these arguments, i.e., *move*, with the sole purpose of justifying their claim. The game is governed by a protocol for admissible moves and the winning conditions. For the proponent a basic protocol for an argumentation game is that the arguments of the move attack the previous move of the adversary, and that the main claim follows from the arguments assessed as currently valid. For the opponent goes that an admissible move has to attack the previous move of the adversary, and that the main claim is not derivable. Even though more complex winning criteria could be devised, by a basic protocol, a player wins the argumentation game when the other party is out of admissible moves.

3 Defeasible Logic

Defeasible Logic (DL) [17,1] is a simple, flexible, rule based non-monotonic formalism able to capture different and sometimes incompatible facets of non-monotonic reasoning [2], and efficient and powerful implementations have been proposed [16,11].

Knowledge in DL can be represented in two ways: facts and rules. *Facts* are represented either in form of states of affairs (literal and modal literal) and actions that have been performed. Facts are represented by predicates. For example, "Tweety is a penguin" is represented by $\mathit{Penguin}(\mathit{Tweety})$. A *rule* describes the relationship between a set of literals (premises) and a literal (conclusion), and we can specify how strong the relationship is. As usual rules allow us to derive new conclusions given a set of premises. We distinguish between *strict rules*, *defeasible rules* and *defeaters* represented, respectively, by expressions of the form $A_1,\ldots,A_n \rightarrow B$, $A_1,\ldots,A_n \Rightarrow B$ and $A_1,\ldots,A_n \rightsquigarrow B$, where $A_1,\ldots,A_n$ is a possibly empty set of prerequisites (causes) and B is the conclusion (effect) of the rule. We only consider rules that are essentially propositional. Rules containing free variables are interpreted as the set of their ground instances.

Strict rules are rules in the classical sense: whenever the premises are indisputable then so is the conclusion. Thus they can be used for definitional clauses. An example of a strict rule is "Penguins are birds", formally: $\mathit{Penguin}(X) \rightarrow \mathit{Bird}(X)$.

Defeasible rules are rules that can be defeated by contrary evidence. An example of such a rule is "Birds usually fly": $\mathit{Bird}(X) \Rightarrow \mathit{Fly}(X)$. The idea is that if we know that X is a bird, we may conclude that X can fly *unless other evidence suggest she may not*.

Defeaters are a special kind of rules. They are used to prevent conclusions, not to support them. For example: $\mathit{Heavy}(X) \rightsquigarrow \neg \mathit{Fly}(X)$. This rule states that if something is heavy then it might not fly. This rule can prevent the derivation of a "fly" conclusion. On the other hand it cannot be used to support a "not fly" conclusion.

DL is a "skeptical" non-monotonic logic, meaning that it does not support contradictory conclusions. Instead DL seeks to resolve conflicts. In cases where there is some support for concluding A but also support for concluding $\neg A$, DL does not conclude either of them (thus the name "skeptical"). If the support for A has priority over the support for $\neg A$ then A is concluded. No conclusion can be drawn from conflicting rules in DL unless these rules are prioritized. The *superiority relation* among rules is used to define priorities among rules, that is, where one rule may override the conclusion of another rule. For example, given the defeasible rules

$$r : Bird(X) \Rightarrow Fly(X) \qquad r' : Penguin(X) \Rightarrow \neg Fly(X)$$

which contradict one another, no conclusive decision can be made about whether a Tweety can fly or not. But if we introduce a superiority relation $\succ$ with $r' \succ r$, then we can indeed conclude that Tweety cannot fly since it is a penguin.

We now give a short informal presentation of how conclusions are drawn in DL. Let D be a theory in DL (i.e., a collection of facts, rules and a superiority relation). A *conclusion* of D is a tagged literal and can have one of the following four forms:

$+\Delta q$ meaning that q is definitely provable in D (*i.e.*, using only facts and strict rules).
$-\Delta q$ meaning that we have proved that q is not definitely provable in D.
$+\partial q$ meaning that q is defeasibly provable in D.
$-\partial q$ meaning that we have proved that q is not defeasibly provable in D.

Strict derivations are obtained by forward chaining of strict rules, while a defeasible conclusion p can be derived if there is a rule whose conclusion is p, whose prerequisites (antecedent) have either already been proven or given in the case at hand (i.e., facts), and any stronger rule whose conclusion is $\neg p$ has prerequisites that fail to be derived.

Formally a DL theory (as formalized by [4]) is a structure $D = (F, R, \succ)$ where F is a finite set of factual premises, R a finite set of rules, and $\succ$ a superiority relation on R. Given a set R of rules, we denote the set of all strict rules in R by R_s, the set of strict and defeasible rules in R by R_{sd}, the set of defeasible rules in R by R_d, and the set of defeaters in R by R_{dft}. $R[q]$ denotes the set of rules in R with consequent q. In the following $\sim p$ denotes the complement of p, that is, $\sim p$ is $\neg q$ if $p = q$, and $\sim p$ is q if p is $\neg q$. For a rule r we will use $A(r)$ to indicate the body or antecedent of the rule and $C(r)$ for the head or consequent of the rule. A rule r consists of its antecedent $A(r)$ (written on the left; $A(r)$ may be omitted if it is the empty set) which is a finite set of literals, an arrow, and its consequent $C(r)$ which is a literal.

Provability is based on the concept of a derivation (or proof) in D. A derivation is a finite sequence $P = (P(1), \dots, P(n))$ of tagged literals. Each tagged literal satisfies some proof conditions. A proof condition corresponds to the inference rules corresponding to one of the four kinds of conclusions we have mentioned above. $P(1..i)$ denotes the initial part of the sequence P of length i. Here we state the conditions for strictly and defeasibly derivable conclusions (see [1] for the full presentation of the logic):

If $P(i+1) = +\Delta q$ then

(1) $q \in F$, or
(2) $r \in R_s[q]$, $\forall a \in A(r) : +\Delta a \in P(1..i)$.

If $P(i+1) = +\partial q$ then
(1) $+\Delta q \in P(1..i)$, or
(2) (2.1) $\exists r \in R_{sd}[q] \forall a \in A(r) : +\partial a \in P(1..i)$ and
(2.2) $-\Delta{\sim}q \in P(1..i)$ and
(2.3) $\forall s \in R[{\sim}q]$ either
(2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
(2.3.2) $\exists t \in R_{sd}[q]\ \forall a \in A(t) : +\partial a \in P(1..i)$ and $t \succ s$.

4 Dialogue Games in Defeasible Logic – A Symmetric Protocol

In [24] we presented a model for an argumentation game in DL using a basic symmetric protocol for adversarial dispute. We parse a dialogue into defeasible rules utilizing time of the dialogue as the time of the rule. In order to resolve the dispute, the agents take turn in putting forward arguments from a private knowledge base, i.e., a finite set of (defeasible) arguments in support of their claim. At each time step, an agent is allowed to put forward any of its arguments (rules) that has precedence over any contradictory defeasible rule of the previous steps.

In this symmetric protocol we assume that if at time t_2 we have a *valid* rule $w \in R_{sd}^{t_2}$ which contradicts a defeasible rule $s \in R_d^{t_1}$ of time t_1 and $t_2 > t_1$ then the strength of w is greater than s. The expression $a@t$ denotes that the expression a being put forward or upgraded at time t.

$$(w \succ s)@t \text{ iff } (w,s) \in \succ \text{ or } w \in R^{t'}[P],\ s \in R^t[\neg P] \text{ where } t' < t$$

A common public knowledge base holds the common knowledge, which is a theory in defeasible logic. The sets of agreed common knowledge construct the theories $T_1, T_2, \ldots, T_n$ respectively as the undefeated defeasible rules from the previously adjacent step are strengthened into strict rules and the defeated defeasible rules are removed. Thus, if P is the conclusion of a defeasible rule of the adjacent previous step t_{i-1}, regardless of its origin, the agreed common knowledge is created by strengthening the status of rules from defeasible to strict in the adjacent next step t_{i+1} if:

$$\exists r \in R_d^{t'}[P]\ t' < t\ \forall t'' : t' < t'' < t\ R_{sd}^{t''}[\neg P] = \emptyset \text{ and } \forall a \in A(r) : +\Delta a@t$$

The proof procedures of the defeasible logic are applied to the critical literal at each time step, thus determining the burden of proof and the outcome of the argumentation game. The first theory $T_1 = (\{\}, R_d^1, \succ)$ is created from the arguments (ARG_1) presented by the first player, and the second theory $T_2 = (\{\}, R_d^2, \succ)$ is created through modifications of T_1 by the arguments (ARG_2) presented by player 2. The transition rules from the first theory to the second theory were devised as follows:

1. If $r \in R_d^1$ and $\forall s \in ARG_2$, $\neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^2$.
2. All rules of ARG_2 are added to T_2 as defeasible rule. Here we assume that ARG_2 is valid and that a valid argument, by the above defined precedence relations, is stronger than any contradictory argument of the previous step.

At time n, $n > 2$ theory T_n is created through modification of T_{n-1} by arguments (ARG_n) of the player who has to play at that step. The rules for transition from T_{n-1} to T_n are

1. If $r \in R_s^{n-1}$ then $r \in R_s^n$.
2. If $r \in R_d^{n-2}$ and $\forall s \in ARG_{n-1}, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_s^n$; otherwise $r \notin R^n$. Here we should note that the player will not oppose its previous argument. Thus, all unchallenged rules of time $n-2$ are upgraded as strict rules at time n.
3. If $r \in R_d^{n-1}$ and $\forall s \in ARG_{n-1}, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^n$. Unchallenged defeasible rules of time $n-1$ are added to T_n as defeasible rules.
4. All rules of ARG_{n-1} are added to T_n as defeasible rules. Here we assume that ARG_{n-1} is valid and that a valid argument, by the above defined precedence relations, is stronger than any contradictory argument of previous step.

The winning criteria for a basic game are devised as an agent to be winning if the claim q is definitely proven $+\Delta q$ at any time step. If an agent at any step of the game proves $+\partial A$ the burden of production as well as persuasion of $-\partial A$, or $-\Delta A$ or $+\Delta\neg A$ or $+\partial\neg A$ are placed on the other party.

Using a symmetric protocol, a dispute between equally strong parties could be modelled as an argumentation game and resolved accordingly. However, in many situations ethical, moral or other reasons (cf., e.g., criminal litigation[4]) advocate for special concerns to be taken on behalf of one of the parties. To accommodate such settings, asymmetric protocols are required.

5 Dialogue Games in Defeasible Logic – An Asymmetric Protocol

Here we present an asymmetric model for adversarial argumentation games between two parties: the *Prosecutor* and the *Defendant*. As in the symmetric protocol, we parse the dialogue into defeasible rules utilizing time of the dialogue as the time of the rule. Each agent has at its disposition a private knowledge base consisting of a finite set of defeasible arguments in support of their claim (the critical literal). Initiated by the prosecutor, the parties take turns in presenting their arguments. At each time step the proof procedures are applied to the critical literal. The outcome of an argumentation game is determined by the final stage of the game, whilst the intermediate stages are illustrating the situation for the current situation. For common sense reasons, as an argument put forward cannot be revoked from impacting the argumentation, we do not allow for backtracking.

The winning criteria for a basic game are devised as an agent to be winning if the claim q is definitely proven $+\Delta q$ at any time step. However, analogously to the burden of persuasion, which imposes a requirement of providing a justified (i.e., strongly defeating) argument for the issue on which the burden rests (based on rebutting defeat) [21], we require of the prosecutor a strong defeat of any argument (including the critical literal) presented by the defendant. Thus, if the prosecutor at any step of the game

[4] "Homo praesumitur bonus donec probetur malus" lat: Innocent until proven guilty. The adoption of this presumption of innocence in many national statutes results in that the defendant of a criminal litigation only is required to at most produce an exception to the accusation.

proves $+\partial A$, the prosecutor still holds the burden to produce proof of $+\Delta A$ in order to win. For the defendant only a burden of production of an exception $+\partial\neg A$, (being subsumed by $-\partial A$, $-\Delta A$ or $+\Delta\neg A$) is imposed. If the defendant at any step of the game proves the exception $+\partial\neg A$, the burden of persuasion placed on the prosecutor necessitates the proof of $+\Delta A$ (including $-\partial\neg A$) in order for the prosecutor to win.

In the symmetric protocol, regardless of its origin, time brings strengthening of undefeated rules from defeasible to strict. Here we require that the strengthening of rules originating from the prosecutor only occurs when the rule could be derived from arguments already put forward by the defendant. In other cases, undefeated rules from the previously adjacent step presented by the prosecutor remain as defeasible rules in the common knowledge base. Defeated rules are removed at each step. As we do not allow the prosecutor to repeat arguments and the arguments put forward have to strongly defeat any arguments put forward by the defendant, the game will terminate.

In the symmetric protocol at each step any agent in turn to move can present an argument if its strength is stronger than contradictory defeasible rules of the previous steps. In our asymmetrical distribution of the burden of proof, the defendant is allowed to present an argument that merely weakly defeats the argument of the prosecutor of the previous steps. As consequence the defendant could remain with the same argument for fulfilling his burden of production of an assumption $+\partial\neg A$ as response to $+\partial A$.

The strength of an argument is determined by either previously known superiority relationships or validity of that rule. Adhering to the above presented syntax, we write $y_i \in R_x^{t_j} | x \in \{d,s,sd\}$. Here y is a rule identifier with the subscripts $i \in \{p,d\}$ where p means that the origin of the rule is the prosecutor and d means that the origin of the rule is the defendant. In the following, unless needed, the indexing is left out for readability reasons.

If $\forall r \in R_s^{t_i}[q]$ and $\forall s \in R_d^{t_i}[\neg q]$, then $(r \succ s)@t_i$

We consider that the rule strength of a strict rule is greater than the rule strength of a defeasible rule.

$(w_d \succ s_p)@t$ iff $(w_d, s_p) \in \succ$ or $s_p \in R^{t'}[P]$ and $w_d \in R^{t''}[\neg P]$, where $t' < t'' < t$

For defeasible rules presented by *the defendant* we simply assume that if at time t_2 we have a *valid* rule $w_d \in R^{t_2}$ which contradicts a defeasible rule $s_p \in R^{t_1}$ of time t_1 and $t_2 > t_1$ then the strength of w_d is greater than s_p. This fits well with the burden of persuasion being placed on the prosecutor. We utilize defeasible logic to determine strength of a new rule presented by the players.

$(s_p \succ w_d)@t$ iff $(s_p, w_d) \in \succ$ or $w_d \in R^{t'}[\neg P]$ and $s_p \in R^{t''}[P]$ and $\forall a \in A(s)$:
$+\Delta a@t$, where $t' < t'' < t$
else
$(w_d \succ s_p)@t$ iff $(w_d, s_p) \in \succ$ or $w_d \in R^{t'}[\neg P]$ and $s_p \in R^{t''}[P]$ and $\exists a \in A(s)$:
$-\Delta a@t$, where $t' < t'' < t$

As the prosecutor holds the burden of persuasion, we assume that unless the rule priority is set, that only if at time t_2 we have a *valid* rule $s_p \in R^{t_2}$ which contradicts a defeasible rule $w_d \in R^{t_1}$ of time t_1 and $t_2 > t_1$ and the rule presented by the prosecutor *strongly* *defeats* the rule of the defendant then the strength of the argument s_p of the prosecutor is *greater* than the argument of the defendant w_d. In all other situations the opposite goes, thus rendering the strength of the argument s_p of the prosecutor *weaker* than the argument of the defendant w_d.

In this asymmetric protocol the criteria for strengthening the rule strength of a defeasible rule to a strict rule are devised as follows:

$$\exists r \in R_d^{t'}[P]\ t' < t\ \forall t'' : t' < t'' < t\ R_{sd}^{t''}[\neg P] = \emptyset \text{ and } \forall a \in A(r) : +\Delta a@t$$

If P is the conclusion of a defeasible rule $r \in R_d^{t'}$ of the adjacent previous step t' and the rule was presented by the *defendant* then we can upgrade the rule status from defeasible to strict in the next time step t if no counterarguments are presented by the prosecutor at time t''. For $t^{Even} < t^{Odd} < t$

$\exists r \in R_d^{t^{Odd}}[P], R_{sd}^{t^{Even''}}[\neg P] = \emptyset$ and $\forall a \in A(r) : +\Delta a@t$ and
1) $\exists r \in R_{ds}^{t^{Even}}[P]$ and $\forall a \in A(r) : +\Delta a@t$, or
2) $+\partial[P]@t from R_{ds}^{t^{Even}}$

However, if P is the conclusion of a defeasible rule $r \in R_d^{t^{Odd}}$ of the adjacent previous step t^{Odd} and the rule was presented by the *prosecutor* then we can upgrade the rule status from defeasible to strict in next step only in the case of no counterarguments being presented by the defendant at the adjacently following time t^{Even} *and* the defeasible (or strict) rule r has been put forward by the defendant *or* the conclusion P follows defeasibly from the defeasible or strict rules $R_{ds}^{t^{Even}}$ presented by the defendant at the adjacently following time t^{Even}.

An argumentation game is initiated at time 1 by the prosecutor putting forward arguments (ARG_1) from its private knowledge into the common knowledge base to prove its claim (critical literal) A. As the parties take turns in presenting their arguments, at time 2 the defendant agent responds to the accusations. We allow arguments in the form of valid defeasible rules being as strong or stronger than at least some rules of theory T_1. The common sets of argument construct the theories $T_1, T_2, T_3, \ldots, T_n$ respectively where the subscripts indicate the time at which the common sets of argument are constructed. As all arguments in the private knowledge base of the agents are defeasible, in the first two theories the common set of arguments consists only of defeasible rules from both time 1 and time 2, according to the following transition rules:

Let the first theory $T_1 = (\{\}, R_d^1, \succ)$ be created from arguments (ARG_1), the operative plea of prosecutor, and the second theory $T_2 = (\{\}, R_d^2, \succ)$ be created through modifications of T_1 by arguments (ARG_2) from the defendant. Now the transition rules from the first theory T_1 to the second theory T_2 are as follows:

1. If $r \in R_d^1$ and $\forall s \in ARG_2$, $\neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^2$.
2. All rules of (ARG_2) are added to T_2 as defeasible rules. Under the assumption of (ARG_2) be valid and that, by the above defined precedence relations, any valid argument from the defendant is stronger than its contradictory argument (from the prosecutor) of the adjacent previous step. As all unchallenged rules of the prosecutor are added to T_2 as defeasible rules T_2 now consists of all unchallenged rules of the prosecutor and all arguments (ARG_2) of the defendant.

At time $n = 2m+1$ $(m > 0)$, theory T_n is created through modification of T_{n-1} by arguments (ARG_n) of the prosecutor. Accounting for the heterogeneity of the parties we capture the asymmetrical burden of proof by the following rules for transition from theories T_{n-1} to T_n:

1. If $r \in R_d^1$ and $\forall s \in ARG_{n-1}, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, and 1) $r \in R_d^{n-1}$ or 2) $R_d^{n-1} \vdash C(r)$, then $r \in R_s^n$.
2. If $r \in R_d^{n-2}$ and $\forall s \in ARG_{n-1}$, $\neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^n$. Here we should note that, in contrast to the symmetric protocol, even though the defendant has not actively challenged these arguments and the prosecutor will not oppose its previous argument by the rules of the game, we find it to be a too strong presumption to strengthen the rule status of these rules to strict rules. Thus, unless the argument is acknowledged by the defendant (see transition rule 1.), all unchallenged rules of time $n-2$ of the prosecutor remain as defeasible rules at time n.
3. If $r \in R_d^{n-1}$ and $\forall s \in ARG_n, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^n$. All unchallenged defeasible rules of time $n-1$ (originating from the defendant) are added as defeasible rules at time n.
4. If $r \in R_d^{n-1}$ and $\forall s \in ARG_n, \neg C(s) = C(r) \wedge \neg C(s) \notin A(r)$, and $r \succeq s$ then $r \in R_d^3$. For removal is required that all rules of the defendant have to be strongly defeated by the prosecutor. Thus, also the defeasible rules of time $n-1$ (originating from the defendant) of equal or stronger strength are added as defeasible rules at time n.
5. If $r \in ARG_n$ and $\forall s \in R_d^{n-1}$, $\neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^n$. All unchallenged rules of (ARG_n) are added to T_n as defeasible rules.
6. If $r \in ARG_n$ and $\forall s \in R_d^{n-1}$, $\neg C(s) = C(r) \wedge \neg C(s) \notin A(r)$ and $r \succeq s$, then $r \in R_d^n$. All rules of (ARG_n) that are of higher priority, i.e., strongly defeat the arguments of the defendant are added to T_n as defeasible rules. Here due to the burden of production of the prosecutor, all arguments added are required to either be unchallenged or to strongly defeat all previous arguments of the defendant. This way, by putting forward new arguments, the prosecutor could strengthen its claim.

As a result T_n consists of the unchallenged defeasible rules of T_{n-2} of the prosecutor, the unchallenged defeasible rules and the rules of by T_{n-1} of the defendant that are challenged by (ARG_n) but found equally strong or stronger, and the unchallenged defeasible rules of the prosecutor from (ARG_n).

At time $n = 2m$ $(m > 1)$, theory T_n is created through modification of T_{n-1} by arguments (ARG_n) of the defendant. The transitions from T_{n-1} to T_n are devised as follows:

1. If $r \in R_d^{n-1}$ and $\forall s \in ARG_n, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_s^n$. The defendant will not oppose its previous argument by the rules of the game. Thus, all the unchallenged defeasible rules R_d^{n-1} are upgraded as strict rules, i.e., facts, at time n.
2. If $r \in R_d^{n-1}$ and $\forall s \in ARG_n, \neg C(s) \neq C(r) \wedge \neg C(s) \notin A(r)$, then $r \in R_d^n$. All unchallenged defeasible rules R_d^{n-1} are added as defeasible rules at time n. As already stated, the defendant cannot challenge her own rules presented in R_d^{n-1}.
3. All rules of (ARG_n) are added to T_n as defeasible rules. In contrast to arguments, e.g., (ARG_{n-1}) originating from the prosecutor, as (ARG_n) originates from the defendant we merely require that all rules of (ARG_n) are valid and at least as strong (or stronger) than any of its contradictory arguments presented by the prosecutor.

An Example – Presumption of Innocence The asymmetric model of argumentation game defeasible logic is illustrated by elaboration on the example of [24]. Consider a particular argumentation game between the prosecutor Alice and the defendant Bob. Alice is trying to convict Bob by proving A and Bob is claiming $\neg A$. At each step they maintain a current set of rules (CR_t, where t is the time). Here a rule consists of its name $R'i$ (where $R'i$ indicates that the rule belongs to the current set CR_t as opposed to the rules present in the private knowledge bases of the parties denoted by R_i), its antecedent $A(r)$, which is a finite set of literals, an arrow, its consequent $C(r)$, which is a literal and $@t_x|x \in \{1,2,3,\dots,n\}$ denotes the time of the rule, which is updated at each step. Alice initiates the game at time t_1 by presenting her first move as

$$R1 : \emptyset \Rightarrow B \qquad R2 : B \Rightarrow A$$

This will generate two defeasible rules as $R'1(\emptyset \Rightarrow B)@t_1$, $R'2(B \Rightarrow A)@t_1$. Thus at time t_1 CR_{t_1}=[$R'1$, $R'2$] and we can prove $+\partial A@t_1$. Now at next time point, Bob gets its chance to disprove A. At time t_2, Bob presents the following argument,

$$R3 : \emptyset \Rightarrow D \qquad R4 : D \Rightarrow \neg A$$

This will generate two new defeasible rules as $R'3(\emptyset \Rightarrow D)@t_2$ $R'4(D \Rightarrow \neg A)@t_2$. Now, Bob only attacks $R'2$ presented by Alice at previous step by $R'4$ and $R'2$ is removed from CR_{t_2}. Note that as $t_2 > t_1$, the strength of $R'4$ is greater than $R'2$ according to the strength determination rule for the defendant. At time t_2, $R'1$ remains unchallenged but as this rule is not utilized even as a premise in the reasoning of Bob it does not commit Bob to this rule, (leaving Bob the possibility to dismiss this rule by contesting it at a later time or Alice to present evidence to strengthen this rule by the strength determination rule). Thus, the rule remain in CR_{t_2} as $R'1(\emptyset \Rightarrow B)@t_2$. This is in contrast to the symmetric protocol where $R'1$ as unchallenged is changed to a strict rule $R'1(\emptyset \rightarrow B)@t_2$ (a fact) regardless of its origin. Note that we change the time stamp of the rule from t_1 to t_2 to indicate that it is member of CR_{t_2} at time t_2. Thus, CR_{t_2} =[$R'1$, $R'3$, $R'4$]. The proof at time t_2 is $+\partial\neg A$ (which implies that we also have $-\Delta\ A$ as the latter rule $R'4$ is stronger than $R'2$ in accordance to the first strength determination rule.

Next at time t_3, in order to defeat the arguments presented by Bob, Alice presents the following arguments:

$$R5 : B \Rightarrow \neg D \qquad R6 : \emptyset \Rightarrow E \qquad R7 : E \Rightarrow A$$

So the translated defeasible rules are $R'5(B \Rightarrow \neg D)@t_3$, $R'6(\emptyset \Rightarrow E)@t_3$, $R'7(E \Rightarrow A)@t_3$. Now the CR_{t_3} is [$R'1$, $R'3$, $R'4$, $R'6$] as the rule $R'4$ is stronger than $R'7$ and the rule $R'3$ is stronger than $R'5$ according to the transition rule as neither the argument $R'7$ nor $R'5$ can strongly defeat $R'4$ or $R'3$ respectively and thus they are removed. So the proof at this time point remain $+\partial\neg A$. If Alice cannot present any additional arguments strongly defeating $R'4$ in the next step the rule $R'4$ is strengthened into a strict rule resulting in the proof of $+\Delta\neg A$. Thus, in contrast to the symmetric protocol example where Alice wins the game as she were able to upgrade the defeasible rules supporting the proof of $+\partial A$ by use of the rule priority assumption in 5, she would need the rule to be strongly defeated as e.g by addition of a rule $R'i : E \Rightarrow \neg D$ to prevent the rule $R'4$ from being strengthened into a strict rule at time t_4. As this is not the case, Bob wins the game at time t_4 and is acquitted from the criminal charge of A.

Another Example – Beyond Reasonable Doubt Consider a second argumentation game between the prosecutor Alice and the defendant Bob. Alice is still trying to convict Bob by proving A and Bob is claiming $\neg A$. Again Alice initiates the game at time t_1 by presenting her first move as

$$R1 : \emptyset \Rightarrow B \qquad R2 : F \Rightarrow A$$

This will generate two defeasible rules as $R'1(\emptyset \Rightarrow B)@t_1$, $R'2(F \Rightarrow A)@t_1$. Thus at time t_1 CR_{t_1}=[$R'1$, $R'2$] and we have proof $+\partial A@t_1$. Now at next time point, Bob gets its chance to disprove A. At time $t_2(t_2 > t_1)$, Bob presents the following argument,

$$R3 : E \Rightarrow D \qquad R4 : B, D \Rightarrow \neg A$$

This will generate two new defeasible rules as $R'3(E \Rightarrow D)@t_2$, $R'4(B, D \Rightarrow \neg A)@t_2$. Now, Bob only attacks $R'2$ presented by Alice at previous step by $R'4$ and $R'2$ is removed from CR. As $t_2 > t_1$, the strength of $R'4$ is greater than $R'2$ according to the strength determination rule for the defendant. Thus, CR_{t_2} =[$R'1$, $R'3$, $R'4$]. The proof at time t_2 is $-\partial\ A$, which includes that we also have $-\Delta\ A$ as the latter rule $R'4$ is stronger than $R'2$.

Next at time t_3, Alice presents the following arguments,

$$R5 : \neg E \Rightarrow \neg D \qquad R6 : \emptyset \Rightarrow \neg E \qquad R7 : B \Rightarrow A$$

The translated defeasible rules are $R'5(\neg E \Rightarrow \neg D)@t_3$, $R'6(\emptyset \Rightarrow \neg E)@t_3$, $R'7(B \Rightarrow A)@t_3$. Now CR_{t_3} =[$R'1$, $R'4$, $R'5$, $R'6$, $R'7$] as $R'3$ is strongly defeated by $R'5$ and $R'6$ and thus, it is removed. At time t_3, $R'1$ remains unchallenged and as it is utilized as a premise in the reasoning of Bob and thus commits Bob to this rule (which is justified as Bob could not be allowed to rely on not actively presented inconsistencies), presented by the prosecutor Alice it is strengthened into a strict rule (i.e., a fact) as $R'1(\emptyset \rightarrow B)@t_3$. So the proof at this time point is $+\partial A$ as the rule is stronger according to the transition rules. If Bob does not present valid arguments in the next step Alice wins the game as from Bob's argumentation her claim is corroborated.

6 Related Work

In this paper we augment the model of dialogue games in defeasible logic of [24]. The work is based on ALIS [21], and we are inspired by [13] as we have separated the knowledge of the players into (1) private knowledge and (2) public knowledge. The common public knowledge forms the common set of arguments, which is a theory in defeasible logic. As such our model in contrast to [21] provides a closer approximation of argumentation games for the agent setting, as the agent could choose at what time and which parts of its arguments (i.e., private knowledge) be disclosed to its opponent.

Substantial work have been done on argumentation games in the AI and Law-field. [3] presents an early specification and implementation of an argumentation game based on the Toulmin argument-schema without a specified underlying logic. [7] presented The Pleadings Game as a normative formalization and fully implemented computational model, using conditional entailment. The goal of the model was to identify issues in the argumentation rather than as in our case elaborating on the status of the main claim. The dialectic proof procedures presented by [6] focus on minimizing the culprit of argumentation. The proof procedures are expressed as metalogic programs. Our use of defeasible logic establishes a difference as to the syntactic limitations as the approach in [6] is built on assumptions that are atomic, whereas in our framework the arguments are expressible as rules of propositional defeasible logic, not being directly applicable in the abstract argumentation framework underlying [5]. DiaLaw [15] is a two player game, in which both players make argument moves. The model combines exchange of statements and exchange of arguments, dealing with rhetorical as well as psychological issues of argumentation. However, the main focus for the two players is to convince each other rather than defeating an adversarial as in our case. The abstract argumentation systems of [25,26] study arguments as the object of defeat. The results however are more related to stable semantics than sceptical as in the defeasible logic utilized in our framework and the study is devised as meta games for changing the rules of argumentation games.

Modelling argumentation games in defeasible logic has been addressed by [13,12,14]. [14], in contrast to our work, focuses on persuasion dialogues in a cooperative setting. It includes in the process cognitive states of agents such as knowledge and beliefs, and presents some protocols for some types of dialogues (*e.g.* information seeking, explanation, persuasion). The main reasoning mechanism is based on basic defeasible logic (cf. Section 3), while ignoring recent development in extensions of defeasible logic with modal and epistemic operators for representing the cognitive states of agents [9,10]. [13] provides an extension of defeasible logic to include the step of the dialogue in a way very similar to what we have presented in the paper. A main difference is that the resulting mechanism just defines a metaprogram for an alternative computational algorithm for ambiguity propagating defeasible logic while the logic presented here is ambiguity blocking. In [12], the authors focus on rule sceptic characterizations of arguments and propose the use of sequences of defeasible (meta) theories, while using meta-reasoning (meta-rules or high level rules) to assess the strength of rules for the theories at lower levels.

7 Conclusion

In this paper Defeasible Logic is used to capture an asymmetric protocol for argumentation games. We have shown that our model provides for a closer approximation of argumentation games for heterogeneous agent settings. The agent characteristics or the agents' relative importance in fulfilling the overall goal of the system could be captured, while all the same the agent is allowed to argue its case in the best way it knows, *e.g.* choosing at what time any subset of its arguments (*i.e.* private knowledge) be disclosed to its adversary.

References

1. G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–287, 2001.

2. G. Antoniou, D. Billington, G. Governatori, M. J. Maher, and A. Rock. A family of defeasible reasoning logics and its implementation. In *Proc. ECAI 2000*: 459–463. 2000.
3. T. J. M. Bench-Capon. Specification and Implementation of Toulmin Dialogue Game. In *Proc. JURIX 98.* Nijmegen, 5-20, GNI, 1984.
4. D. Billington. Defeasible logic is stable. *Journal of Logic and Computation*, 3:370–400, 1993.
5. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence*, 77:321–357, 1995.
6. P. M. Dung, R. A. Kowalski and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation, *Artificial Intelligence*, 170: 114-159, 2006.
7. T. Gordon. The Pleadings Game: An artificial Intelligence Model of Procedural Justice *Journal of Artificial Intelligence and Law*, 2(4), Springer Verlag, the Netherlands, 1993.
8. G. Governatori, M. J. Maher, D. Billington, and G. Antoniou. Argumentation semantics for defeasible logics. *Journal of Logic and Computation*, 14(5):675–702, 2004.
9. G. Governatori and A. Rotolo. Defeasible logic: Agency, intention and obligation. In *Deontic Logic in Computer Science*, LNAI 3065: 114–128. Springer, 2004.
10. G. Governatori, A. Rotolo, and V. Padmanabhan. The cost of social agents. In *Proc. AAMAS 2006*: 513–520. ACM Press, 2006.
11. B. N. Grosof. Representing e-commerce rules via situated courteous logic programs in RuleML. *Electronic Commerce Research and Applications*, 3(1):2–20, 2004.
12. J. Eriksson Lundström, A. Hamfelt and J. Fischer Nilsson. A Rule-Sceptic Characterization of Acceptable Legal Arguments In *Proc. ICAIL 2007*: 283–284. ACM Press, 2007.
13. A. Hamfelt, J. Eriksson Lundström, and J. Fischer Nilsson. A metalogic formalization of legal argumentation as game trees with defeasible reasoning. In *Proc. ICAIL 2005*: 250–251. ACM Press, 2005.
14. I. A. Letia and R. Varic. Defeasible protocols in persuasion dialogues. In *Proc. WE-IAT'06*, IEEE, 2006.
15. A. R. Lodder. DiaLaw: On Legal Justification and Dialogical Models of Argumentation. *Artificial Intelligence and Law*, 8(2-3): 265-276(12), 2000.
16. M. J. Maher, A. Rock, G. Antoniou, D. Billington, and T. Miller. Efficient defeasible reasoning systems. *International Journal of Artificial Intelligence Tools*, 10(4):483–501, 2001.
17. D. Nute. Defeasible logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, 3: 353–395. Oxford University Press, 1994.
18. H. Prakken. Modelling defeasibility in law:logic or procedure? In *Fundamenta informaticae*, 253–271, 2001.
19. H. Prakken and G. Sartor. Rules about rules: Assessing conflicting arguments in legal reasoning. In *Artificial Intelligence and Law*: 331–368, 1996.
20. H. Prakken and G. Sartor. Presumptions and burdens of proof. In *Proc. Jurix 2006*: 21–30, IOS Press, 2006.
21. H. Prakken and G. Sartor. Formalising arguments about the burden of persuation. In *Proc. ICAIL 2007*: 97–106. ACM Press, 2007.
22. H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127: 187–219, 2001.
23. B. Roth, R. Riveret, A. Rotolo and G. Governatori. Strategic Argumentation: A Game Theoretical Investigation. In *Proc. ICAIL 2007*: 81–90. ACM Press, 2007.
24. S. Thakur, G. Governatori, V. Padmanabhan Nair and J. Eriksson Lundström. Dialogue Games in Defeasible Logic. In *Proc. AUSTAI 2007*: LNAI 4830, 497-506. Springer, 2007.
25. G. Vreeswijk. Abstract Argumentation Systems. *Artificial Intelligence*, 90: 225-279, 1997.
26. G. Vreeswijk. Representation of formal dispute with a standing order. *Artificial Intelligence Law* 8(2/3): 205-231, 2000.