Attack-defense semantics of argumentation

Beishui LIAO [a,1], Leendert VAN DER TORRE [b]
[a] *Zhejiang University*
[b] *University of Luxembourg*

Abstract. This is the appendix for all the proofs in the paper "Attack-defense semantics of argumentation".

A. Proofs

Theorem 1. *If D is a stable attack-defense extension, then it is complete and preferred, but not necessarily vice versa.*

Proof. Since for all $z^x_y \in T \setminus D$, $y \in \mathbf{defendee}(D)$, $D \cup \{z^x_y\}$ is not admissible. So, D is complete. Assume that D is not maximal. Let $D' \subset D$ be a preferred attack-defense extension. For each $z^x_y \in D' \setminus D$, $D \cup \{z^x_y\}$ is not admissible. So, D' is not admissible. Contradiction. For the counter example, consider $T = \{a^a_a\}$. The emptyset is both a preferred attack-defense extension and a complete attack-defense extension, but not a stable attack-defense extension. $\square$

Theorem 2. *Let D be a set of attack-defenses. If* $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$*, then D is a complete attack-defense extension iff $D = F_T(D)$. The grounded attack-defense extension is the least complete attack-defense extension of F (w.r.t. set inclusion).*

Proof. Since $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$ and $D = F_T(D)$, D is admissible. Since no attack-defense in $T \setminus D$ is successful w.r.t. D, D is complete. Since F_T is monotonic, there is a unique least fixed point that is the least complete attack-defense extension of F. $\square$

Theorem 3. *Let $T \subseteq \mathscr{D}^U$ be an attack-defense framework. For all $D \in \Sigma(T)$, $x, u \in$* $\mathbf{argument}(T)$*, if z^x_y, $w^u_v \in D$, $z^w_{y'} \in T$, then $z^w_{y'} \in D$.*

Proof. We need to verify this property under complete attack-defense semantics. Since $w^u_v \in D$, $w \in \mathbf{defendee}(D)$. Since $z^x_y \in D$, for each $y' \neq y$ that attacks z, $\exists z^{x'}_{y'} \in D$. Thus, $z^w_{y'}$ is successful w.r.t. D. Since D is a complete attack-defense extension, $z^w_{y'} \in D$. $\square$

Theorem 4. *For all $D \in \Sigma(T)$, if $z^x_y \in D$ and $x \neq \top$ then there exists $x^u_{y'} \in T$ s.t. $x^u_{y'} \in D$.*

Proof. According to Definition 3, since $z^x_y \in D$, it holds that defender $x \in \mathbf{defendee}(D)$. So, there exist $u, y' \in \mathbf{argument}(T)$ such that $x^u_{y'} \in D$. $\square$

[1] Corresponding Author: Beishui Liao.

The third property is about the incompleteness of an attack-defense framework.

Theorem 5. *For all $z_y^x, v_y^u \in T$, it is not necessary that $v_y^x \in T$.*

Proof. By definition, this property obviously holds. □

Theorem 7. *Defenses $z_y^y, z_z^x \in T$ are unsatisfiable. Furthermore, if $z_y^y \in T$, then u_v^y is unsatisfiable under semantics $\Sigma \in \{CO, PR, GR, ST\}$.*

Proof. First, z_y^y means that y self-attacks and it attacks z. If z_y^y is in an admissible set of attack-defenses, then y is a defendee. According to Definition 4, $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) \neq \emptyset$. Contradiction. Second, obviously, z_z^x is not satisfiable. Third, assume that u_v^y is in some attack-defense admissible set D. Then, according to Theorm 4, there exists $y_w^x \in T$ such that $y_w^x \in D$. If $w = y$, then this contradicts $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$. Otherwise, according to the definition of complete attack-defense extension, z_y^y is also in D. This also contradicts $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$. □

Theorem 8. *If there exist $x_z^y, y_x^z, z_y^x \in T$, then x_z^y, y_x^z and z_y^x are unsatisfiable under semantics $\Sigma \in \{CO, PR, GR, ST\}$.*

Proof. Assume that x_z^y is in some admissible set $D \subseteq T$. Then, y_x^z is also in D. As a result, $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) \supseteq \{x\}$, contradicting $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$. □

Theorem 9. *For any attack-defense framework T, and for $\Sigma \in \{CO, PR, GR\}$, $\Sigma(T) = \Sigma(T^-)$.*

Proof. Under complete attack-defense semantics, let $D \in CO(T)$. Since for all $z_y^x \in u(T)$, z_y^x is unsatifiable, $z_y^x \notin D$. According to Definitions 4 and 5, D is a complete attack-defense extension of T^-. On the other hand, let $D \in CO(T^-)$. Since any attack-defense in $u(T)$ is unsatisfiable, adding it to T^- does not affect the evaluation of other attack-defenses. Therefore, D is a complete attack-defense extension of T. Similarly, under preferred and grounded attack-defense semantics, this property also holds. □

Theorem 10. *For any attack-defense frameworks T and T', for any $B \subseteq$ $\mathbf{argument}(T) \cap \mathbf{argument}(T')$, if $T \equiv_d^\Sigma T'$ then $T \equiv_{r,B}^\Sigma T'$, but not vice versa.*

Proof. Since $T \equiv_d^\Sigma T'$, $\Sigma(T) = \Sigma(T')$. For all $D \in \Sigma(T)$, for all $z \in B$, $root_\Sigma(z, T) = root_\Sigma(z, T')$. As a result, $T \equiv_{r,B}^\Sigma T'$.

For the converse direction, consider T_3 and T_3' in Examples 3 and 4. Under preferred attack-defense semantics, $PR(T_3) \neq PR(T_3')$, i.e., it is not the case that $T_3 \equiv_d^{PR} T_3'$. Meanwhile, let $B = \{a, b\}$. As described in Example 4, we have $T_3 \equiv_{r,B}^{PR} T_3'$. □

Theorem 11. *Let $\mathrm{d}(\mathscr{F})$ be the attack-defense framework of an AF $\mathscr{F} = (\mathscr{A}, \rightarrow)$. For all $z_y^x, v_y^u \in \mathrm{d}(\mathscr{F})$, it holds that $v_y^x \in \mathrm{d}(\mathscr{F})$.*

Proof. Since $z_y^x, v_y^u \in \mathrm{d}(\mathscr{F})$, it holds that $x \rightarrow y$, $y \rightarrow z$, $u \rightarrow y$ and $y \rightarrow v$. Given that $x \rightarrow y$ and $y \rightarrow v$, we have $v_y^x \in \mathrm{d}(\mathscr{F})$. □

Theorem 12. *For all $D \in \Sigma(\mathrm{d}(\mathscr{F}))$, $\mathbf{defendee}(D) \in \sigma(\mathscr{F})$, where $\Sigma \in \{\mathrm{CO}, \mathrm{PR}, \mathrm{GR}, \mathrm{ST}\}$ and $\sigma \in \{\mathrm{co}, \mathrm{pr}, \mathrm{gr}, \mathrm{st}\}$.*

Proof. Let D be a complete attack-defense extension of $\mathrm{d}(\mathscr{F})$. According to Definition 4, $\mathbf{defendee}(D) \cap \mathbf{attacker}(D) = \emptyset$. Hence, $\mathbf{defendee}(D)$ is conflict-free. According to Definition 3, for all $z \in \mathbf{defendee}(D)$, z is an initial argument, or every attacker of z is attacked by an argument in $\mathbf{defendee}(D)$. So, $\mathbf{defendee}(D)$ is an admissible set of arguments. According to Definition 5, each successful attack-defense w.r.t. D is in D. So, each argument that is defended by $\mathbf{defendee}(D)$ is in $\mathbf{defendee}(D)$. Therefore, $\mathbf{defendee}(D)$ is a complete extension of $\mathscr{F}$. When D is a preferred attack-defense extension, it is easy to see that $\mathbf{defendee}(D)$ is a maximal complete extension (w.r.t. set inclusion) and therefore is a preferred extension of $\mathscr{F}$. Similarly, this properpy holds under grounded attack-defense semantics. Finally, when D is a stable attack-defense extension, for all $z_y^x \in \mathrm{d}(\mathscr{F}) \setminus D$, $y \in \mathbf{defendee}(D)$. This means that for each argument $z \in \mathscr{A} \setminus \mathbf{defendee}(D)$, z is attacked by an argument y in $\mathbf{defendee}(D)$. Hence, $\mathbf{defendee}(D)$ is a stable extension of $\mathscr{F}$. □

Theorem 13. *For all $E \in \sigma(\mathscr{F})$, let $\mathrm{def}(E) = \{z_y^x \mid z_y^x \in \mathrm{d}(\mathscr{F}) : x, z \in E\} \cup \{z_\perp^\top \mid z_\perp^\top \in \mathrm{d}(\mathscr{F}) : z \in E\}$. Then, $\mathrm{def}(E) \in \Sigma(\mathrm{d}(\mathscr{F}))$.*

Proof. Let E be a complete extension of $\mathscr{F}$. For each $z_y^x \in \mathrm{def}(E)$, since $x, z \in E$ and E is admissible, z_y^x is successful w.r.t. $\mathrm{def}(E)$. $\mathbf{defendee}(\mathrm{def}(E)) \cap \mathbf{attacker}(\mathrm{def}(E)) = \emptyset$. Otherwise, if there is an attacker in $\mathrm{def}(E)$ that is a defendee, then $E = \mathbf{defendee}(\mathrm{def}(E))$ is not conflict-free, contradicting the fact that E is conflict-free. As a result, $\mathrm{def}(E)$ is admissible. Assume that $\mathrm{def}(E)$ is not complete. Then, there exists $z_y^x \in \mathrm{d}(\mathscr{F})$, such that z_y^x is successful w.r.t. $\mathrm{def}(E)$, and $z_y^x \notin \mathrm{def}(E)$. Since $z_y^x \notin \mathrm{def}(E)$, $x \notin E$ or $z \notin E$. If $x \notin E$, then $x \notin \mathbf{defendee}(\mathrm{def}(E))$ and therefore z_y^x is not successful w.r.t. $\mathrm{def}(E)$. Contradiction. Alternatively, if $z \notin E$, then since E is a complete extension of $\mathscr{F}$, z is not defended by E. Then, there exists $y' \neq y$, such that y' attacks z and y' is not attacked by any argument in E. It turns out that z_y^x is not successful w.r.t. $\mathrm{def}(E)$. Contradiction. Hence, $\mathrm{def}(E)$ is a complete attack-defense extension of $\mathrm{d}(\mathscr{F})$.

When E is a preferred extension of $\mathscr{F}$, E is maximal and therefore $\mathrm{def}(E)$ is maximal (w.r.t. set inclusion). Hence, $\mathrm{def}(E)$ is a preferred attack-defense extension of $\mathrm{d}(\mathscr{F})$.

Similarly, this properpy holds under grounded semantics.

Finally, when E is a stable extension of $\mathscr{F}$, for all $z \in \mathscr{A} \setminus E$, z is attacked by an argument in E. This means that for each defense $z_y^x \in \mathrm{d}(\mathscr{F}) \setminus \mathrm{def}(E)$, $y \in \mathbf{defendee}(\mathrm{def}(E))$. Therefore, $\mathrm{def}(E)$ is a stable attack-defense extension of $\mathrm{d}(\mathscr{F})$. □

Theorem 14. *Let $\mathscr{F}$ and $\mathscr{G}$ be two AFs. If $\mathrm{d}(\mathscr{F}) \equiv_d^\Sigma \mathrm{d}(\mathscr{G})$, then $\mathscr{F} \equiv^\sigma \mathscr{G}$, where $\Sigma \in \{CO, PR, GR, ST\}$, $\sigma \in \{co, pr, gr, st\}$.*

Proof. If $\mathrm{d}(\mathscr{F}) \equiv_d^\Sigma \mathrm{d}(\mathscr{G})$, then $\Sigma(\mathrm{d}(\mathscr{F})) = \Sigma(\mathrm{d}(\mathscr{G}))$. According to Theorem 12, $\sigma(\mathscr{F}) = \mathbf{defendee}(\Sigma(\mathrm{d}(\mathscr{F}))) = \mathbf{defendee}(\Sigma(\mathrm{d}(\mathscr{G}))) = \sigma(\mathscr{G})$. Since $\sigma(\mathscr{F}) = \sigma(\mathscr{G})$, $\mathscr{F} \equiv^\sigma \mathscr{G}$. □

Lemma 1. *It holds that $CO(\mathrm{d}(\mathscr{F})) = CO(\mathrm{d}(\mathscr{F}^{\mathrm{ck}}))$.*

Proof. Since for every attack-defense that is related to a self-attacking argument is unsatsifiable, it is clear that $CO(\mathrm{d}(\mathscr{F})) = CO(\mathrm{d}(\mathscr{F}^{\mathrm{ck}}))$. □

Theorem 15. *Let $\mathscr{F}$ and $\mathscr{G}$ be two AFs. If $\mathscr{F} \equiv_s^{\mathrm{co}} \mathscr{G}$, then $\mathrm{d}(\mathscr{F}) \equiv_d^{CO} \mathrm{d}(\mathscr{G})$.*

Proof. Obvious. □

Theorem 16. *Let $\mathscr{F} = (\mathscr{A}_1, \rightarrow_1)$ and $\mathscr{H} = (\mathscr{A}_2, \rightarrow_2)$ be two AFs. If* $\mathrm{d}(\mathscr{F}) \equiv_r^{\Sigma} \mathrm{d}(\mathscr{H})$*, then $\mathscr{F} \equiv^{\sigma} \mathscr{H}$, where* $\Sigma \in \{\mathrm{CO}, \mathrm{PR}, \mathrm{GR}, \mathrm{ST}\}$ *and* $\sigma \in \{\mathrm{co}, \mathrm{pr}, \mathrm{gr}, \mathrm{st}\}$.

Proof. According to Definition 14, under complete semantics, the number of extensions of $\mathrm{co}(\mathscr{F})$ is equal to the number of $root_{CO}(z, \mathrm{d}(\mathscr{F}))$, where $z \in \mathscr{A}_1$. Since $root_{CO}(z, \mathrm{d}(\mathscr{F})) = root_{CO}(z, \mathrm{d}(\mathscr{H}))$, $\mathscr{A}_1 = \mathscr{A}_2$.

Let $root_{CO}(z, \mathrm{d}(\mathscr{F})) = root_{CO}(z, \mathrm{d}(\mathscr{H})) = \{R_1, \ldots, R_n\}$. Let $\mathrm{co}(\mathscr{F}) = \{E_1, \ldots, E_n\}$ be the set of extensions of $\mathscr{F}$, where $n \geq 1$.

For all $\alpha \in \mathscr{A}_1$, for all R_i, $i = 1, \ldots, n$, we have $\alpha \in E_i$ iff $R_i \neq \{\}$, in that in terms of Definition 14, when $R_i \neq \{\}$, there is a reason to accept α.

On the other hand, let $\mathrm{co}(\mathscr{H}) = \{S_1, \ldots, S_n\}$ be the set of extensions of $\mathscr{H}$. For all $\alpha \in \mathscr{A}_2 = \mathscr{A}_1$, for all R_i, $i = 1, \ldots, n$, for the same reason, we have $\alpha \in S_i$ iff $R_i \neq \{\}$. So, it holds that $E_i = S_i$ for $i = 1, \ldots, n$, and hence $\mathrm{co}(\mathscr{F}) = \mathrm{co}(\mathscr{H})$, i.e., $\mathscr{F} \equiv^{\mathrm{co}} \mathscr{H}$.

Simlarly, this property holds under preferred semantics and grounded semantics, respectively.

Finally, under stable semantics, if $ST(\mathrm{d}(\mathscr{F})) = ST(\mathrm{d}(\mathscr{H})) = \emptyset$, then $st(\mathscr{F}) = st(\mathscr{H}) = \emptyset$. Hence, $\mathscr{F} \equiv^{st} \mathscr{H}$. Otherwise, $st(\mathscr{F}) = st(\mathscr{H}) \neq \emptyset$. In this case, as verified under complete semantics, it holds that $\mathscr{F} \equiv^{st} \mathscr{H}$. $\square$