On the Different Types of Collective Attacks in Abstract Argumentation: Equivalence Results for SETAFs

Wolfgang Dvořák Anna Rapberger Stefan Woltran

Institute of Logic and Computation, TU Wien, Austria
{dvorak,arapberg,woltran}@dbai.tuwien.ac.at

Abstract

Argumentation frameworks with collective attacks are a prominent extension of Dung's abstract argumentation frameworks, where an attack can be drawn from a set of arguments to another argument. These frameworks are often abbreviated as SETAFs. Although SETAFs have received increasing interest recently, a thorough study on the actual behavior of collective attacks has not been carried out yet. In particular, the richer attack-structure SETAFs provide, can lead to different forms of redundant attacks, i.e., attacks that are subsumed by attacks involving less arguments. Also the notion of strong equivalence, which is fundamental in nonmonotonic formalisms to characterize equivalent replacements, has not been investigated for SETAFs so far. In this paper, we first provide a classification of different types of collective attacks, and analyze for which semantics they can be proven redundant. We do so for eleven well established abstract argumentation semantics. We then study how strong equivalence between SETAFs can be decided with respect to the considered semantics and also consider variants of strong equivalence. Our results show that removing redundant attacks in a suitable way provides direct means to characterize strong equivalence by syntactical equivalence of so-called kernels, thus generalizing well-known results on strong equivalence between Dung AFs.

1 Introduction

Abstract argumentation frameworks (AFs) as introduced by Dung [8] are a core formalism in formal argumentation. A popular line of research investigates extensions of Dung AFs that allow for a richer syntax (see, e.g., [7]). In this work we consider SETAFs as introduced by Nielsen and Parsons [17] which generalize the binary attacks in Dung AFs to collective attacks such that a set of arguments B attacks another argument a but no proper subset of B attacks a. As discussed in [17], there are several scenarios where arguments interact and can constitute an attack on another argument only if these arguments are jointly taken into account. Representing such a situation in Dung AFs often requires additional artificial arguments to "encode" the conjunction of arguments.

SETAFs have received increasing interest in the last years. For instance, semi-stable, stage, ideal, and eager semantics have been adapted to SETAFs in [11, 14]; also ranking semantics have been defined for SETAFs recently [23]. Translations between SETAFs and other abstract argumentation formalisms are studied in [19] and the expressiveness of SETAFs is investigated in [10]. Yun et al. [22] observed that for particular instantiations, SETAFs provide a more convenient target formalism than Dung AFs.

Figure 1 shows an example SETAF with three arguments a, b, and c. Each argument is jointly attacked by the two remaining arguments; in other words, the set $\{a, b\}$ attacks c, $\{a, c\}$ attacks b, and $\{b, c\}$ attacks a. Having only these three attacks indicates that a alone (and likewise b alone) is too weak to attack c, etc. The crucial step towards the definitions of semantics is to fix the notion of conflict for SETAFs. In our example, $S = \{a, b\}$ is a conflict-free set since b—although being attacked by $\{a, c\}$—is not attacked by a alone (or by $\{a, b\}$), and likewise b is not attacked by a or $\{a, b\}$. Set $\{a, b, c\}$, on the other hand, is conflicting in this example. The definition of the actual semantics is then quite straight forward. For instance, stable semantics are defined according to the standard definition by Dung, i.e., are given by conflict-free sets that attack all remaining arguments. In our example, we have that the conflict-free sets

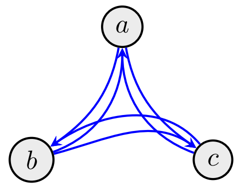


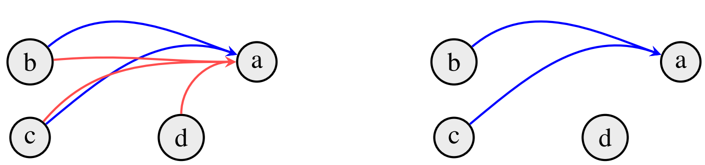
Figure 1: An example SETAF with three collective attacks.

$\{a, b\}$, $\{a, c\}$, and $\{b, c\}$ satisfy this requirement and these three sets are indeed the stable extensions of this example SETAF.

The richer syntactical structure of SETAFs leads to situations that are not possible in Dung AFs, namely that a set S attacks an argument a, but also a proper subset $S' \subset S$ attacks a. Such a situation is depicted in Figure 2a, where we have attacks $(\{b, c, d\}, a)$ and $(\{b, c\}, a)$. As has been already observed by Polberg [19], here the attack $(\{b, c, d\}, a)$ can be safely removed (see Figure 2b), no matter how the remaining framework looks like. Intuitively, for each set of arguments, removing the attack $(\{b, c, d\}, a)$ does not change the arguments attacked by that set. Such attacks (S, a), for which there exists another attack (S', a) with $S' \subset S$ are named *redundant* attacks (cf. [19]). However, this is not the only effect that can be observed in the context of SETAFs and it is one main aim of the paper to investigate further situations where attacks can be safely removed, that is, the extensions remain unchanged. In particular, the concept of self-attacks appears to be more subtle than in Dung AFs. Naturally, an attack (S, a) is considered to be self-attacking if $a \in S$. As it turns out, under certain conditions, self-attacks can be transformed to proper attacks which, in turn, allow for identification of further redundant attacks. Figure 3 shows such an example. Here, we have a self-attack $(\{a, b, c\}, a)$ which—due to the presence of attack $(\{c\}, b)$—can be *reduced* to the proper attack $(\{b, c\}, a)$. Note that after this transformation the attack $(\{b, c, d\}, a)$ becomes redundant, while it was not redundant in F. Without going into too much detail at this point, the faithfulness of the replacement of $(\{a, b, c\}, a)$ by $(\{b, c\}, a)$ is based on the observations that (a) in both F and G, in order to defend a, one needs to attack b or c, and that (b) the conflict-free sets of F and G coincide and attack the same arguments.

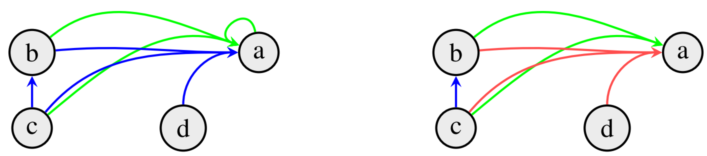
As we will show in the paper both transformations discussed so far are applicable to all established semantics for SETAFs, i.e., conflict-free, naive, admissible, stable, preferred, complete, grounded, stage, semi-stable, ideal, and eager semantics. Further transformations can be obtained for specific semantics. We illustrate this on the concept of *active* attacks and stable semantics. An attack (S, a) is considered active in a SETAF F if S is conflict-free in F. These are those attacks which are potentially "used" by stable semantics in order to attack an argument. Inactive attacks on the other hand lack this capability. For instance, the SETAF G in Figure 4a shows an inactive attack $(\{b, c\}, a)$. As this is the only attack on a, each stable extension S of G with $a \notin S$ would require $b, c \in S$ which, however, is ruled out by the attack $(\{c\}, b)$. Hence, $(\{b, c\}, a)$ is not relevant for stable semantics in this example. Summarizing, for stable semantics one can simplify the SETAF F from Figure 3 to the AF G' from Figure 4. Hence, understanding the different concepts of redundancy allows for substantial simplification of SETAFs which might be useful in practical applications.

In all above examples, the considered transformations can be also carried out within a larger SETAF by just



(a) SETAF F. (b) SETAF G.

Figure 2: Example of a redundant attack (in SETAF F, colored in red) and the framework which arises after removing the redundant attack (SETAF G).



(a) SETAF F. (b) SETAF G.

Figure 3: Example of a reducible attack (in SETAF F, colored in green) and the framework which arises after replacement (SETAF G, attack in red becomes redundant).

(a) SETAF G. (b) SETAF G'.

Figure 4: Example of an inactive attack (in SETAF F, colored in red) and the framework which arises after removal of that attack (SETAF G).

inspecting the involved arguments and their attacks. In other words, redundancy concerns simplification that can identified locally without evaluating the entire framework. The notion of strong equivalence makes this intuition precise and is recognized as a central concept in nonmonotonic reasoning [16, 21, 20]. In terms of AFs, strong equivalence (with respect to a semantics σ) between two frameworks F and G holds, if for any further AF H, $\sigma(F \cup H) = \sigma(G \cup H)$. Hence, replacing a subframework F by a strongly equivalent AF G in any context does not alter the extensions, making strong equivalence the key towards simplifying a part of an argumentation framework without looking at the rest of the framework. For instance, SETAF F from Figure 3 is strongly equivalent to G' from Figure 4 with respect to stable semantics.

For Dung AFs, strong equivalence and variants thereof have been extensively studied in the literature [18, 1, 5, 3, 6]. The main results reveal that strong equivalence can be decided by syntactic identity of so-called kernels of the AFs to be compared. In these kernels, depending on the actual semantics, certain attacks need to be removed. As we will show in the paper, strong equivalence can be similarly characterized for SETAFs but additional treatment is needed to generalize results for AFs to the richer attack structure SETAFs provide. Concepts as discussed as above together with further notions of removable attacks will be required in this endeavor.

Organization and Main Results. The structure of the paper and its main contributions can be outlined as follows. After having provided the necessary background in Section 2, we thoroughly investigate different notions of attacks in Section 3.

- We start by showing adequacy of the concept of redundant attacks for all eleven semantics considered (conflict-free, naive, admissible, stable, preferred, complete, grounded, stage, semi-stable, ideal, and eager) and strengthen the concept for conflict-free and naive semantics.

- In Section 3.1, we provide four more classes of proper attacks that are removable under certain semantics; the main idea is to strengthen the notion of inactive attacks.

- In Section 3.2, we study the already sketched notion of reducible self-attacks and provide two further classes of self-attacks that are removable under certain semantics.

- We relate the introduced classes of attacks in Section 3.3 and then proceed by defining a normal form for SETAFs in Section 3.4 that relies on the interplay between reducible and redundant attacks.

We then turn to strong equivalence in Section 4 where we first introduce strong equivalence and related notions of equivalence. The main results of this section are as follows:

- We first show in Section 4.1 that the normal form translation from Section 3.4 preserves strong equivalence for all semantics under consideration.
- We introduce four different notions of kernels and show that the transformation of a SETAF into its kernel form preserves strong equivalence for particular semantics.
- We then use our kernels to provide exact characterizations for strong equivalence, i.e., we show that two SETAFs are strongly equivalent if and only if their kernels coincide; these results also apply to the related notions of equivalence we consider.
- We clarify the relations between the different kernels and relate strong equivalence with respect to different semantics; as we will show, all relations from Dung AFs carry over to SETAFs.

We conclude the paper with a summary and outlook for future work.

A preliminary version of this paper has been presented at the 42nd German Conference on AI (KI 2019) [13]. That paper was limited to strong equivalence and also did not consider the full range of semantics. In the present paper, certain notions have been refined; these differences are discussed at the relevant occasions.

2 Preliminaries

Throughout the paper, we assume a countably infinite domain $\mathfrak{A}$ of possible arguments.

Definition 1. A SETAF is a pair $F = (A, R)$ where $A \subseteq \mathfrak{A}$ is finite, and $R \subseteq (2^A \setminus \{\emptyset\}) \times A$ is the attack relation. An attack (S, a) is called proper if $a \notin S$, otherwise (S, a) is called a self-attack. SETAFs (A, R), where for all $(S, a) \in R$ it holds that $|S| = 1$, amount to (standard Dung) AFs. In that case, we usually write (a, b) to denote the set-attack $(\{a\}, b)$. Finally, for a SETAF $F = (A, R)$, we use A_F and R_F to identify its arguments A and respectively its attack relation R.

Combining AFs and similar manipulations are defined as follows.

Definition 2. Given SETAFs F, G we define the union of F and G as $F \cup G = (A_F \cup A_G, R_F \cup R_G)$. We also use, for SETAF F and a relation $R \subseteq A_F \times A_F$, $F \cup R$ as a shorthand for $(A_F, R_F \cup R)$ and $F \setminus R$ as a shorthand for $(A_F, R_F \setminus R)$.

We also make use of the following notational conventions.

Definition 3. Given a SETAF $F = (A, R)$, we write $S \mapsto_R b$ if there is a set $S' \subseteq S$ with $(S', b) \in R$. Moreover, we write $S' \mapsto_R S$ if $S' \mapsto_R b$ for some $b \in S$. We drop subscript R in $\mapsto_R$ if there is no ambiguity. For $S \subseteq A$, we use S^+_F to denote the set $\{b \mid S \mapsto_R b\}$ and define the *range* of S (with respect to F), denoted $S^\oplus_F$, as the set $S \cup S^+_F$.

The notions of conflict and defense naturally generalize to SETAFs.

Definition 4. Given a SETAF $F = (A, R)$, a set $S \subseteq A$ is *conflicting* in F if $S \mapsto_R a$ for some $a \in S$. A set $S \subseteq A$ is *conflict-free* in F, if S is not conflicting in F, i.e., if $S' \cup \{a\} \not\subseteq S$ for each $(S', a) \in R$. *cf*(F) denotes the set of all conflict-free sets in F.

Definition 5. Given a SETAF $F = (A, R)$, an argument $a \in A$ is *defended* (in F) by a set $S \subseteq A$ if for each $B \subseteq A$, such that $B \mapsto_R a$, also $S \mapsto_R B$. A set T of arguments is defended (in F) by S if each $a \in T$ is defended by S (in F).

The semantics we study in this work are the naive, admissible, stable, preferred, complete, grounded, stage, semi-stable, ideal, and eager semantics, which we will abbreviate by *naive*, *adm*, *stb*, *pref*, *com*, *grd*, *stage*, *sem*, *ideal*, and *eager* respectively [17, 11, 14].

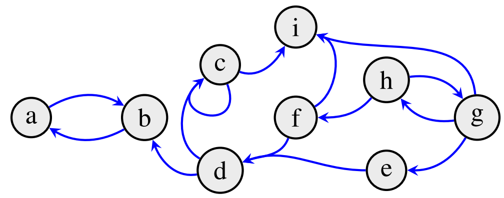


Figure 5: SETAF $F = (A, R)$ from Example 1.

Definition 6. Given a SETAF $F = (A, R)$ and a conflict-free set $S \in \mathit{cf}(F)$. Then,

- $S \in \mathit{naive}(F)$, if there is no $T \in \mathit{cf}(F)$ with $T \supset S$;
- $S \in \mathit{adm}(F)$, if S defends all its arguments $s \in S$ in F;
- $S \in \mathit{stb}(F)$, if $S \mapsto a$ for all $a \in A \setminus S$;
- $S \in \mathit{pref}(F)$, if $S \in \mathit{adm}(F)$ and there is no $T \in \mathit{adm}(F)$ s.t. $T \supset S$;
- $S \in \mathit{com}(F)$, if $S \in \mathit{adm}(F)$ and $a \in S$ for all $a \in A$ defended by S;
- $S \in \mathit{grd}(F)$, if $S = \bigcap_{T \in \mathit{com}(F)} T$;
- $S \in \mathit{stage}(F)$, if $\nexists T \in \mathit{cf}(F)$ with $T_F^{\oplus} \supset S_F^{\oplus}$;
- $S \in \mathit{sem}(F)$, if $S \in \mathit{adm}(F)$ and $\nexists T \in \mathit{adm}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$;
- $S \in \mathit{ideal}(F)$, if $S \in \mathit{adm}(F)$, $S \subseteq \bigcap_{E \in \mathit{pref}(F)} E$ and $\nexists T \in \mathit{adm}(F)$ s.t. $T \subseteq \bigcap_{E \in \mathit{pref}(F)} E$ and $T \supset S$;
- $S \in \mathit{eager}(F)$, if $S \in \mathit{adm}(F)$, $S \subseteq \bigcap_{E \in \mathit{sem}(F)} E$ and $\nexists T \in \mathit{adm}(F)$ s.t. $T \subseteq \bigcap_{E \in \mathit{sem}(F)} E$ and $T \supset S$.

Notice that in [14] the definitions of preferred, semi-stable, ideal, and eager semantics are based on complete extensions instead of admissible sets. However, as shown in Appendix A their definitions are equivalent to the definitions provided above.

We consider the following example to illustrate the semantics.

Example 1. Consider the SETAF $F = (A, R)$ from Figure 5 which contains the proper set-attacks $(\{f, g\}, i)$, $(\{e, f\}, d)$ and the self-attack $(\{c, d\}, c)$. An example for a conflict-free set is given by $E_1 = \{a, c, f\}$ since there are no conflicts among the arguments. Observe that the set is not naive since there is a proper conflict-free set of arguments extending E_1, e.g., the set $E_2 = \{a, c, e, f\}$ which is in fact a naive extension. It can be furthermore checked that E_2 is stage by verifying that $(E_2)_F^{\oplus} = \{a, b, c, d, e, f, i\}$ is maximal among all $T_F^{\oplus}$ with $T \in \mathit{cf}(F)$. Observe that neither E_1 nor E_2 are admissible since they do not defend themselves (e.g., f is not defended against $(\{h\}, f)$).

We give an example for an admissible set: Let $E_3 = \{a, d, g\}$, then E_3 is admissible since it is conflict-free and defends itself: d is defended by g against $\{e, f\}$ and both a and g defend itself against the attack from b and h, respectively. Observe that E_3 is not complete since g additionally defends f against h. The complete extensions are given by $\mathit{com}(F) = \{\emptyset, \{a\}, \{a, d, f, g\}, \{a, d, e, h\}\}$, thus the grounded extension is empty. Moreover, $\mathit{pref}(F) = \{\{a, d, e, h\}, \{a, d, f, g\}\}$. The set $E_4 = \{a, d, f, g\}$ is the unique semi-stable extension of F. Observe that $E_4^{\oplus} = \{a, b, d, e, f, g, h\} \neq A$; it can be checked that F has no stable extensions. Moreover, $\mathit{eager}(F) = \mathit{sem}(F)$ since $\bigcap_{E \in \mathit{sem}(F)} E = E_4$. We compute the ideal extension: Since $\bigcap_{E \in \mathit{pref}(F)} E = \{a, d\}$ we conclude that $\{a\}$ is ideal since it is the maximal admissible set contained in the intersection of the preferred extensions.

The relationship between the semantics has been clarified in [17, 11, 14] and matches with the relations between the semantics for Dung AFs, i.e., for any SETAF F:

$$\mathit{stb}(F) \subseteq \mathit{sem}(F) \subseteq \mathit{pref}(F) \subseteq \mathit{com}(F) \subseteq \mathit{adm}(F) \subseteq \mathit{cf}(F); \tag{1}$$

$$\mathit{stb}(F) \subseteq \mathit{stage}(F) \subseteq \mathit{naive}(F) \subseteq \mathit{cf}(F). \tag{2}$$

Moreover, the grounded extension is the unique minimal complete extension for any SETAF F. Also the ideal and the eager (for finite SETAFs) extension has been shown to be unique [14]. Furthermore, the grounded extension of a SETAF F can be alternatively defined as the least fixed point of the function $\Gamma_F : 2^A \to 2^A$, $\Gamma_F(T) := \{a \in A \mid a \text{ is defended by } T \text{ in } F\}$. The following property also carries over from Dung AFs: For any SETAF F, if $\mathit{stb}(F) \neq \emptyset$ then $\mathit{stb}(F) = \mathit{sem}(F) = \mathit{stage}(F)$.

Finally, we provide some relations concerning equivalent frameworks. The correctness of the statements follows in the same vein as the corresponding results of Dung AFs, see e.g., [18].[1]

Proposition 1. *Let F, G be SETAFs and $\sigma \in \{\mathit{adm}, \mathit{com}\}$. We have*

1. *$\sigma(F) = \sigma(G)$ implies $\mathit{pref}(F) = \mathit{pref}(G)$;*
2. *$\sigma(F) = \sigma(G)$ implies $\mathit{ideal}(F) = \mathit{ideal}(G)$;*
3. *$\sigma(F) = \sigma(G)$ and $\mathit{sem}(F) = \mathit{sem}(G)$ implies $\mathit{eager}(F) = \mathit{eager}(G)$;*
4. *$\mathit{com}(F) = \mathit{com}(G)$ implies $\mathit{grd}(F) = \mathit{grd}(G)$;*
5. *$\mathit{cf}(F) = \mathit{cf}(G)$ if and only if $\mathit{naive}(F) = \mathit{naive}(G)$.*

Proof. For (1) recall that preferred extensions can be characterized as the $\subset$-maximal admissible sets, complete extensions respectively. That is the preferred extensions are fully determined by the admissible sets, complete extensions respectively. (2) Notice that ideal extension can be characterized either in terms of complete and preferred extensions or by admissible sets and preferred extensions. In the light of (1) the ideal extension is thus fully determined by the admissible sets, complete extensions respectively. (3) The eager extension can be characterized either in terms of complete and semi-stable extensions or by admissible sets and semi-stable extensions. Thus the semi-stable extensions and the admissible set, complete extensions respectively, together fully determine the eager extension. (4) By definition the grounded extension is the $\subset$-minimal complete extension and thus fully determined by the complete extensions. (5) As the naive extensions are the $\subset$-maximal conflict-free sets the naive extensions are fully determined by the conflict-free sets. On the other hand the conflict-free sets can be characterized as all the subsets of naive extensions and are thus also fully determined by the naive extensions. □

3 Notions of Removable Attacks

In this section we discuss different types of attacks and investigate which of them are removable with respect to a particular semantics. In general, attacks in SETAFs can be partitioned into *proper attacks*, i.e., attacks (S, a) where $a \notin S$ and *self-attacks*, i.e., attacks (S, a) where $a \in S$. In Sections 3.1 and 3.2, we will separately investigate subclasses of proper attacks and self-attacks, respectively; moreover, we will relate a subclass of proper attacks with a particular type of self-attacks by showing that they can be transformed into each other without a potential loss of information. After consolidating those different notions in Section 3.3, we make use of our findings and discuss normal forms of SETAFs in Section 3.4.

We shall start with a general notion of redundancy in SETAFs which has been already observed by Polberg in [19].

Definition 7. Let $F = (A, R)$ be a SETAF. An attack $(S, a) \in R$ is called *redundant* in F if there exists $(S', a) \in R$ with $S' \subset S$.

As noticed in [19] one can remove redundant attacks from a SETAF without changing its semantics. We extend that result to semi-stable, stage, eager and ideal semantics.

Lemma 1. *Let $F = (A, R)$ be a SETAF, let (S, a) be redundant in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for all semantics σ under consideration.*

[1] $\mathit{com}(F) = \mathit{com}(G)$ and $\mathit{sem}(F) = \mathit{sem}(G)$ implies $\mathit{eager}(F) = \mathit{eager}(G)$ was not mentioned in [18] but is due to results in [9].

Proof. Let (S, a) be redundant in $F = (A, R_F)$ and $G = (A, R \setminus \{(S, a)\}) = (A, R_G)$. We show that (1) each set of arguments $T \subseteq A$ attacks the same arguments in both F and G, i.e., $T^+_{R_F} = T^+_{R_G}$; and (2) every set of arguments $T \subseteq A$ defends the same set of arguments in F and in G. Recall that F and G differ only regarding the presence of the attack (S, a), thus (1) holds in the case $S \not\subseteq T$. In case $S \subseteq T$, all we need to show is that T attacks a in G. By redundancy of (S, a), there is a subset $S' \subset S$ such that $(S', a) \in R_F$, thus also $(S', a) \in R_G$, consequently, a is attacked by T in G. To show (2), observe that (1) implies that T defends the same arguments $b \neq a$ in both F and G, since every such b is attacked by the same sets of arguments in both F and G. Moreover, if T defends a in F, then T defends a in G by (1) and since $R_F \supseteq R_G$. In the case that T defends a in G there is $c \in S' \subset S$ such that T attacks c, thus the presence of the attack $S' \subset S$ ensures that a is defended against S in F.

We have shown that each set of arguments $T \subseteq A$ attacks and defends the same arguments in both SETAFs F and G. Since every semantics under consideration builds upon those central notions of attack and defense, this result implies that both F and G are indistinguishable with respect to every of those semantics, i.e., $\sigma(F) = \sigma(G)$ for every semantics σ under consideration. □

In the case of conflict-free-based semantics, an even stronger notion of redundancy applies. Observe that for conflict-free sets, the particular target of an attack does not yield additional information, i.e., two attacks $(S \cup \{b\}, a)$, $(S \cup \{a\}, b)$ are interchangeable. Therefore, redundancy can also be strengthened in this sense, as we show next.

Definition 8. Let $F = (A, R)$ be a SETAF. An attack $(S, a) \in R$ is called *cf-redundant* in F if there exists $(S', b) \in R$ with $S' \cup \{b\} \subset S \cup \{a\}$.

Next we show that removing cf-redundant attacks does not change conflict-free and naive semantics.

Lemma 2. *Let $F = (A, R)$ be a SETAF, let (S, a) be cf-redundant in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive\}$.*

Proof. Clearly, $S \cup \{a\} \not\subseteq T$ for all $T \in cf(F)$. Moreover, $S \cup \{a\} \not\subseteq T$ for all $T \in cf(G)$: Since (S, a) is cf-redundant, there is an attack $(S', b) \in R$ with $S' \cup \{b\} \subset S \cup \{a\}$, i.e., $S \cup \{a\}$ is not conflict-free in G. It follows that $cf(F) = cf(G)$ and thus also $naive(F) = naive(G)$. □

Note that cf-redundancy from Definition 8 applies to simple attacks (a, b) in Dung AFs in the presence of self-attacks (a, a) or (b, b). On the other hand, the notion of redundancy from Definition 7 does not apply to any attack in a Dung AF and is thus genuine to SETAFs, which already indicates a richer attack structure of SETAFs compared to Dung AFs. In the following sections we present a detailed analysis of different types of attacks in SETAFs. We start by proper attacks and identify semantics-dependent removable attacks before discussing self-attacks and in particular their relation to proper attacks as already briefly discussed in the introduction.

Remark 1. In the preliminary version [13], we distinguish between *active* and *inactive* attacks: An attack is said to be inactive if and only if either $a \notin S$ and S contains a conflict or $a \in S$ and a proper subset of S contains a conflict; active attacks are those attacks which are not inactive. In contrast to the present version, the separation of proper attacks and self-attacks has thus only been considered implicitly by case distinction in the definition of inactive attacks. In the present paper, we consider the separation into proper attacks and self-attacks explicitly in order to accomplish a more fine-grained analysis of different types of attacks in SETAFs. We will therefore introduce a different naming scheme: The concept of active and inactive attacks will be adopted only for proper attacks, while the self-attacking pendants will be called critical and target-conflicting/reducible self-attacks, respectively.

3.1 Proper Attacks

Recall that proper attacks are attacks (S, a) where $a \notin S$. In what follows, we will occasionally just use the term attack when we refer to proper attacks if there is no ambiguity. We start by distinguishing between active and inactive such attacks.

Definition 9. Let $F = (A, R)$ be a SETAF and let $(S, a) \in R$ be a proper attack. (S, a) is called *inactive* in F if S is conflicting in F, i.e., if there exist $S' \subseteq S$ and $b \in S$ such that $(S', b) \in R$. (S, a) is said to be *active* in F if S is not conflicting in F.

(a) SETAF F.
(b) SETAF G.

Figure 6: Example of an inactive attack (in SETAF F, colored in red) and the framework which arises after removing the inactive attack (SETAF G).

An example of an inactive attack is given by Figure 6a. There the attack $(\{b, c, d\}, a)$ is inactive because of the attack $(\{b, c\}, d)$.

Intuitively, active attacks are those which cannot be removed. Just consider as an example $r = (\{a\}, b)$. Clearly, such an attack plays a crucial role in evaluating a SETAF which contains this attack. In fact, it is evident that for a SETAF $(\{a, b\}, \{r\})$ removing r from the attacks results in a change of the extensions for all considered semantics.

For inactive attacks, their actual impact depends on the semantics at hand. Observe that each inactive attack (S, a) is cf-redundant in a SETAF (A, R): given that there is $(S', b) \in R$ with $S' \subseteq S$ and $b \in S$, we have that $S' \cup \{b\} \subset S \cup \{a\}$. It follows that each inactive attack can be removed without affecting the outcome of conflict-free and naive semantics, cf. Lemma 2. The next lemma shows that removing inactive attacks is also possible when it comes to stable and stage semantics.

Lemma 3. *Let $F = (A, R)$ be a SETAF, let (S, a) be inactive in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive, stb, stage\}$.*

Proof. By Lemma 2, we have $cf(F) = cf(G)$ since inactive attacks are cf-redundant. Moreover, $S \cup \{a\} \nsubseteq T$ for every $T \in cf(F)$, thus also $T_F^+ = T_G^+$ for every $T \in cf(F)$. The statement follows by definition of stable and stage semantics. □

Let us illustrate the concept of inactive attacks for stable semantics in terms of a Dung AF $F = (A, R)$: an attack $(a, b) \in R$ is inactive if $(a, a) \in R$. Removing (a, b) from the attacks in R has no effect since (i) the conflicts remain the same (as we have already seen when discussing cf-redundancy); (ii) if b needs to be attacked by a stable extension, this cannot happen due to attack (a, b) (since a will never part of a stable extension due to its conflict).

However, for the remaining semantics, deletion of inactive attacks is problematic. The intuition is that such an attack (S, a) might still be crucial to keep a undefended. In other words, deletion of (S, a) might turn a from an undefended to a defended argument under some extension S; this obviously holds no matter whether S is conflicting or not.

This issue is illustrated in Figure 6 which depicts the SETAF F containing the inactive attack $(\{b, c, d\}, a)$ and the SETAF G which arises after removing the inactive attack. Both SETAFs F and G agree on their conflict-free, stable, naive, and stage extensions but differ on their admissible extensions: observe that $\{a\} \notin adm(F)$ but $\{a\} \in adm(G)$, since the argument a is no longer attacked in the SETAF G.

In what follows we thus strengthen the notion of inactive attacks. First consider inactive attacks (S, a) where there is a self-attack on a that only contains arguments from $S \cup \{a\}$. In the presence of such a self-attack the attack (S, a) has no effect on whether a is defended or not.

Definition 10. Let $F = (A, R)$ be a SETAF. A proper attack $(S, a) \in R$ is called *target self-attacking inactive* (tsa-inactive) in F if (S, a) is inactive in F and there exists $S' \subset S$ such that $(S' \cup \{a\}, a) \in R$.

An example of a tsa-inactive attack is given in Figure 7a. The inactive attack $(\{b, c, d\}, a)$ is tsa-inactive because of the attack $(\{d, a\}, a)$. Given that attack $(\{d, a\}, a)$ we have that a can only be accepted if d is attacked by the extension and thus the attack $(\{b, c, d\}, a)$ is not relevant for the acceptance of a.

In above definition, we decided to use $S' \subset S$ instead of $S' \subseteq S$ since for the case $S' = S$, a proper attack (S, a) would already be redundant in the sense of Definition 7 since $S \subset S \cup \{a\}$. Also note that in the setting of Dung AFs, an attack (a, b) is tsa-inactive in an AF if both a and b are self-attacking in that AF.

(a) SETAF F. (b) SETAF G.

Figure 7: Example of a tsa-inactive attack (in SETAF F, colored in red) and the framework which arises after removing the tsa-inactive attack (SETAF G).

(a) SETAF F. (b) SETAF G.

Figure 8: Example of a ba-inactive attack (in SETAF F, colored in red) and the framework which arises after removing the ba-inactive attack (SETAF G).

By definition tsa-inactive attacks are inactive and can be therefore removed with respect to conflict-free, naive, stable, and stage semantics (cf. Lemma 3). The next result extends this observation to all remaining semantics.

Lemma 4. *Let $F = (A, R)$ be a SETAF, let (S, a) be tsa-inactive in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for all semantics σ under consideration.*

Proof. We write $R_F = R$ and $R_G = R \setminus \{(S, a)\}$ to emphasize the affiliations. For $\sigma \in \{cf, naive, stb, stage\}$, the statement follows since each tsa-inactive attack is inactive by definition. We show that every conflict-free set $T \in cf(F)$ defends the same arguments in both F and in G. Recall that (1) $cf(F) = cf(G)$ and (2) $T_F^+ = T_G^+$ for each $T \in cf(F)$ as shown in the proof of Lemma 3.

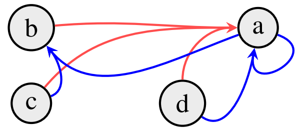
We show that (3) for each conflict-free set $T \subseteq A$, for each argument $b \in A$, T defends b in F iff T defends b in G: The statement follows for every argument $b \neq a$ since T attacks the same arguments in both F and G by (2) and since b is attacked by the same sets of arguments in both F and G. In case $b = a$, observe that if T defends a in F then a is also defended by T in G. Consider the case that T defends a in G. We show that a is defended against S in F, i.e., we show that a is defended against the attack (S, a). Recall that by tsa-inactivity of (S, a), there exists a set $S' \subset S$ such that $(S' \cup \{a\}, a) \in R_F \cap R_G$. We show that T does not attack a in F: Towards a contradiction, assume that T attacks a in F, i.e., there is a set $T' \subseteq T$ such that $(T', a) \in R_F$. Since the case $T' = S$ contradicts the conflict-freeness of T by inactivity of (S, a), we assume that $T' \neq S$. But then $(T', a) \in R_G$. It follows that T is not conflict-free in G since T defends a against T' by assumption, contradiction. Thus we conclude (1) $S' \neq \emptyset$ and (2) T attacks an argument $c \in S'$. It follows that T attacks S in F since $c \in S' \subset S$. That is, T defends b also against the attack (S, a).

$\sigma(F) = \sigma(G)$ for $\sigma \in \{adm, com\}$ follows directly from (1)–(3); moreover, $T_{R_F}^{\oplus} = T_{R_G}^{\oplus}$ for every $T \in cf(F)$ and thus also $sem(F) = sem(G)$. Consequently, also $\sigma(F) = \sigma(G)$ for $\sigma \in \{pref, ideal, grd, eager\}$ by Proposition 1. □

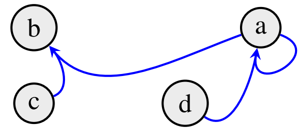
Another approach to strengthen the concept of inactive attacks (S, a) is to make a defending itself against S. This is captured by the following definition.

Definition 11. Let $F = (A, R)$ be a SETAF. A proper attack $(S, a) \in R$ is called *back-attacking inactive* (ba-inactive) in F if (S, a) is inactive in F and there exists $b \in S$ such that $(\{a\}, b) \in R$.

In the setting of Dung AFs, an attack (a, b) is ba-inactive in a Dung AF iff (a, a) and (b, a) are also present in that AF. An example of a ba-inactive attack in the setting of SETAFs is given in Figure 8a. Here, the attack $(\{b, c, d\}, a)$ is ba-inactive because the attacked argument a defends itself by attacking the argument d in the source set.



(a) SETAF F.



(b) SETAF G.

Figure 9: Example of a grd-irrelevant attack (in SETAF F, colored in red) and the framework which arises after removing the grd-irrelevant attack (SETAF G).

By definition, ba-inactive attacks are inactive and can be therefore removed with respect to conflict-free, naive, stable, and stage semantics (cf. Lemma 3). The next result gives further semantics that allow for removal of ba-inactive attacks. As it turns out, ba-inactive attacks can be safely removed under all semantics except complete and grounded.

Lemma 5. *Let $F = (A, R)$ be a SETAF, let (S, a) be ba-inactive in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive, stb, stage, adm, pref, sem, ideal, eager\}$.*

Proof. We write $R_F = R$ and $R_G = R \setminus \{(S, a)\}$ to emphasize the affiliations. For $\sigma \in \{cf, naive, stb, stage\}$, the statement follows by Lemma 3. Recall that (1) $cf(F) = cf(G)$ and (2) $T^+_F = T^+_G$ for each $T \in cf(F)$ by inactivity of (S, a).

We will show that (3) $adm(F) = adm(G)$. Since R_G is a proper subset of R_F we have that each conflict-free set T which defends itself in F also defends itself in G, consequently $adm(F) \subseteq adm(G)$. Let us therefore consider an admissible set $T \in adm(G)$. We show that T defends itself in F by arguing that T is either not attacked or defended against (S, a) in F. Indeed, in the case $a \in T$, then a defends itself against S since, for some $b \in S$, $(\{a\}, b) \in R_G \subset R_F$ by ba-inactivity of (S, a). Observe that in the case $a \notin T$, the set T has the same attacker in both F and G and therefore $adm(F) = adm(G)$.

By (2) and (3) we conclude that $sem(F) = sem(G)$. For the remaining semantics, the statement follows by (3) and by Proposition 1, i.e., $\sigma(F) = \sigma(G)$ for $\sigma \in \{pref, ideal, eager\}$. □

As already mentioned, ba-inactivity is too weak to work for complete and grounded semantics as illustrated in Figure 8. In the SETAF F, the complete extensions are given by the sets $\{b, c\}$ and $\{a, b, c\}$ and thus $grd(F) = \{\{b, c\}\}$, whereas in G the argument a is not attacked by any set and therefore $com(G) = grd(G) = \{\{a, b, c\}\}$. Removing ba-inactive attacks does therefore not preserve defense, i.e., for a set of arguments $T \subseteq A$, the removal of a ba-inactive attack potentially enlarges the set of arguments which are defended by T.

Our final result in this subsection gives one more notion of proper attacks that turns out to be removable with respect to grounded semantics.

Definition 12. Let $F = (A, R)$ be a SETAF. A proper attack $(S, a) \in R$ is called *grd-irrelevant* in F if there exists $S' \subset S$ such that $(S' \cup \{a\}, a) \in R$ and there exist $S'' \subseteq S \cup \{a\}$, $b \in S$ such that $(S'', b) \in R$.

An example of a grd-irrelevant attack is given in Figure 9a. The attack $(\{b, c, d\}, a)$ is grd-irrelevant as a is attacked by $\{d, a\}$; moreover, $\{a, c\}$ attacks the argument b in the source set. As in the case of tsa-attacks, the attack $(\{d, a\}, a)$ ensures that $(\{b, c, d\}, a)$ is not relevant for the acceptance of a. The attack $(\{a, c\}, b)$ then ensures that $(\{b, c, d\}, a)$ cannot be used to attack a from a grounded extension.

Note that a grd-irrelevant is not necessarily inactive (this holds already for the setting of Dung AFs: here, an attack (a, b) is grd-irrelevant in a Dung AF iff (b, b) and (b, a) are also present in that AF); also the attack $(\{b, c, d\}, a)$ in the SETAF F in Figure 9a is in fact active. It is inactive, if S'' in above definition does not contain a, and in this case, the definition matches the one of tsa-inactive attacks from Definition 10. Hence, each tsa-inactive attack is also grd-irrelevant, but not vice versa. However, each grd-irrelevant attack is cf-redundant since $S' \cup \{a\} \subset S \cup \{a\}$, with S' as in above definition. It follows that each grd-irrelevant attack can be removed without affecting the outcome of conflict-free and naive semantics, cf. Lemma 2. The next result adds grounded semantics to this observation.

Lemma 6. *Let $F = (A, R)$ be a SETAF, let (S, a) be grd-irrelevant in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive, grd\}$.*

Proof. Let $R_F = R$ and $R_G = R \setminus \{(S, a)\}$. By Lemma 2, the statement holds for $\sigma \in \{cf, naive\}$. Recall that the grounded extension of a SETAF F can alternatively be defined as the least fixed point of Γ_F. We will prove by induction that for all $i \geq 1$, $\Gamma^i_F(\emptyset) = \Gamma^i_G(\emptyset)$.

Observe that $\Gamma_F(\emptyset) = \Gamma_G(\emptyset)$ holds since $(S' \cup \{a\}, a) \in R_G \subset R_F$ by grd-irrelevance of (S, a), and thus the set of unattacked arguments remains unchanged. Now, fix $i > 1$ and suppose $\Gamma^i_F(\emptyset) = \Gamma^i_G(\emptyset)$. Let $T = \Gamma^i_F(\emptyset)$. We will show that $\Gamma^{i+1}_F(\emptyset) = \Gamma^{i+1}_G(\emptyset)$ (equivalently, $\Gamma_F(T) = \Gamma_G(T)$), i.e., T defends the same arguments in both F and G.

In order to do so, we will first prove that (1) $T^+_F = T^+_G$, i.e., T attacks the same arguments in F and in G. If $S \nsubseteq T$, then the statement follows. Let us therefore assume that $S \subseteq T$. By grd-irrelevance of the attack (S, a), there is an argument $b \in S \subseteq T$ such that b is attacked by a set $S'' \subseteq S \cup \{a\}$. Notice that $a \in S''$ holds, otherwise T would be conflicting via the attack (S'', b), contradiction to conflict-freeness of the grounded extension. Moreover, since T defends itself by construction, it follows that T attacks a (otherwise, T attacks some argument $c \in S'' \setminus \{a\} \subseteq T$, contradiction to conflict-freeness), and thus the statement follows.

By (1), T defends the same arguments $b \neq a$ in both F and in G. Moreover, if T defends a in F, then a is also defended by T in G (by (1) and since a has less attackers in G). It remains to show that if a is defended by T in G then T defends a also in F against (S, a). By grd-irrelevance of (S, a), the argument a is attacked by a set $S' \cup \{a\}$ with $S' \subseteq S$. Notice that $S' \neq \emptyset$ (otherwise, a attacks itself, contradiction to conflict-freeness of the grounded extension). Consequently, T attacks $c \in S' \subset S$, hence the statement follows. □

Removing grd-irrelevant attacks is not semantics-preserving for the remaining semantics under consideration. If we compare SETAF F (cf. Figure 9a) and SETAF G (cf. Figure 9b) then it becomes evident that the removal of grd-irrelevant attacks potentially changes the extensions with respect to admissible-based, complete as well as stable semantics. Observe that the set $\{b, c, d\}$ is stable in F since a is attacked but fails to be admissible in G since it is not defended against the attack $(\{a, c\}, b)$.

3.2 Self-Attacks

Recall that self-attacks in SETAFs are of the form (S, a) with $a \in S$. In AFs, self-attacks are in general not removable. Due to the richer structure of SETAFs, several classes of self-attacks arise, and as will see in this section some of them can be safely removed depending on the actually chosen semantics. However, let us start with those self-attacks that cannot be removed.

Definition 13. Let $F = (A, R)$ be a SETAF. A self-attack $(S, a) \in R$ is called *critical* in F if there is no $S' \subset S$, $b \in S$ such that $(S', b) \in R$.

Similar to self-attacks in standard AFs, critical self-attacks cannot be removed from SETAFs. Basically, this is due to the fact that removing such an attack from a SETAF changes its conflict-free sets. Next, we introduce a class of non-critical self-attacks that can be faithfully transformed into a proper attack and thus do not require special considerations since the results from the previous subsection can then be applied instead.

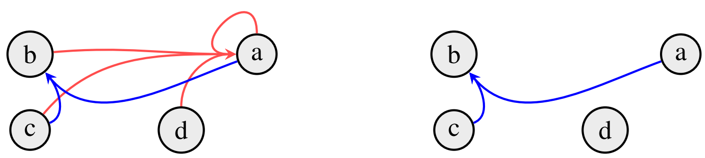
Definition 14. Let $F = (A, R)$ be a SETAF. We call a self-attack $(S, a) \in R$ *reducible* in F if there exist $S' \subseteq S \setminus \{a\}$ and $b \in S$ such that $(S', b) \in R$.

Note that above definition includes redundant attacks only in case $b = a$. In turn, each reducible attack is cf-redundant by definition. Nonetheless, as we show next, any reducible self-attack (S, a) can be safely replaced (under all considered semantics) by the proper attack $(S \setminus \{a\}, a)$. Note that the latter is then by definition an inactive proper attack according to Definition 9.

Lemma 7. *Let $F = (A, R)$ be a SETAF, let (S, a) be reducible in F and let G be the SETAF which arises from F when the reducible attack (S, a) is replaced by $(S \setminus \{a\}, a)$, i.e., $G = (A, (R \setminus \{(S, a)\}) \cup \{(S \setminus \{a\}, a)\})$. Then $\sigma(F) = \sigma(G)$ for all semantics under consideration.*

SETAF F SETAF G

Figure 10: Given that $(\{a, b\}, c)$ is inactive it is equivalent to $(\{a, b, c\}, c)$ for all semantics under our considerations. We call the attack $(\{a, b, c\}, c)$ in F reducible and G the normal form of F.



(a) SETAF F. (b) SETAF G.

Figure 11: Example of a tc self-attack (in SETAF F, colored in red) and the framework which arises after removing the tc self-attack (SETAF G).

Proof. Let us consider a SETAF $F = (A, R)$ with $(S, a) \in R$ being reducible and $G = (A, (R \setminus \{(S, a)\}) \cup \{(S \setminus \{a\}, a)\})$. Since S is conflicting in both F and G we have that $S \not\subseteq T$ for each conflict-free set T in F and in G, consequently (1) $\mathit{cf}(F) = \mathit{cf}(G)$.

Furthermore, (2) for each conflict-free set $T \in \mathit{cf}(F) = \mathit{cf}(G)$ we have $T_F^+ = T_G^+$. Observe that $T_G^+ = T_F^+$ if $S \setminus \{a\} \not\subseteq T$: In that case $S \not\subseteq T$ and $S \setminus \{a\} \not\subseteq T$ and therefore $a \in T_G^+$ iff there is $U \subseteq T$ with $(U, a) \in R$ iff $a \in T_F^+$. We show that $S \setminus \{a\} \not\subseteq T$: By reducibility of (S, a) there exists $S' \subseteq S \setminus \{a\}, b \in S$ such that $(S', b) \in R$. In the case $b = a$, we have that (S, a) is redundant in F and $(S \setminus \{a\}, a)$ is redundant in G, in both cases witnessed by the attack (S', a). Removing both redundant attacks yields the same framework (A, R) and does not change the semantics, thus the statement follows. In the case $b \neq a$, we have that $S \setminus \{a\}$ is conflicting in F and in G and therefore $S \setminus \{a\} \not\subseteq T$ for each conflict-free set T. Consequently, $T_F^+ = T_G^+$.

Moreover, (3) each conflict-free set $T \in \mathit{cf}(F) = \mathit{cf}(G)$ defends the same arguments in F and in G. First note that for each attack $(U, b) \in R$ and for each $b \in A$, we have that T defends an argument b against (U, b) in F iff T defends b against (U, b) in G since $T_F^+ = T_G^+$. We show that a is defended by T against (S, a) in F iff a is defended by T against $(S \setminus \{a\}, a)$ in G. Clearly, if T attacks $c \in S \setminus \{a\}$, then T attacks $c \in S$. Now let T defend a in F against (S, a), then there is $c \in S$, $T' \subseteq T$ such that $(T', c) \in R$. By conflict-freeness of T we have that $c \neq a$ (otherwise it exists $d \in T'$, $T'' \subseteq T$ such that $(T'', d) \in R$), thus $c \in S \setminus \{a\}$ and therefore a is defended by T in G against $(S \setminus \{a\}, a)$.

By (1) it follows that $\mathit{naive}(F) = \mathit{naive}(G)$. Moreover, (1) and (3) implies that admissible and complete sets coincide. Furthermore, $\mathit{sem}(F) = \mathit{sem}(G)$, $\mathit{stage}(F) = \mathit{stage}(G)$, and $\mathit{stb}(F) = \mathit{stb}(G)$, since $T_F^+ = T_G^+$ by (2). As a consequence, $\sigma(F) = \sigma(G)$ for $\sigma \in \{\mathit{pref}, \mathit{ideal}, \mathit{eager}, \mathit{grd}\}$ (cf. Proposition 1). □

Note that the above lemma can also be seen from a different direction in the sense that inactive proper attacks can be turned into non-critical self-attacks. This indicates the special role of inactive attacks in SETAFs.

We proceed by analyzing self-attacks that are neither critical nor reducible. By the above definitions these are those self-attacks (S, a) that have a proper subset S' with $a \in S'$ that corresponds to some attack.

Definition 15. Let $F = (A, R)$ be a SETAF. A self-attack $(S, a) \in R$ is called *target-conflicting* (tc) in F if there exist $S' \subset S$ with $a \in S'$ and $b \in S \setminus \{a\}$ such that $(S', b) \in R$.

An example of a tc self-attack is given in Figure 11a where the source set of the tc self-attack $(\{a, b, c, d\}, a)$ contains a proper subset $\{a, c\}$ attacking the argument b. Notice that the condition $a \in S'$ ensures that tc self-attacks are not reducible. Also observe that tc self-attacks are always cf-redundant, hence, they can be faithfully removed under conflict-free and naive semantics. In addition, this applies also to stable and stage semantics.

Lemma 8. *Let $F = (A, R)$ be a SETAF, let (S, a) be a tc self-attack in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive, stb, stage\}$.*

Proof. As each tc self-attack is cf-redundant, by Lemma 2, we have $cf(F) = cf(G)$. Moreover, for every $T \in cf(F)$ we have $S \not\subseteq T$, consequently $T^+_F = T^+_G$ holds. The statement follows. □

Observe that tc self-attacks cannot be removed for semantics which require self-defense of the extensions, since the removal might turn undefended arguments into defended ones. This behavior can be observed in the example given in Figure 11, where the SETAF G depicts the framework which arises after removing the tc self-attack from F. The set $\{a, d\}$ is admissible in G but not admissible in F because the argument a is not defended against the tc self-attack $(\{a, b, c, d\}, a)$, i.e., the set turns into an admissible extension if the tc self-attack is removed.

For grounded semantics, a further observation can be made: Each tc self-attack (S, a) can be replaced by the proper attack $(S \setminus \{a\}, a)$ without changing the grounded extension. This becomes evident since in both cases, the attack ensures that the argument a is only part of the grounded extension if some argument $b \in S \setminus \{a\}$ is attacked; moreover, neither the attack (S, a) nor the attack $(S \setminus \{a\}, a)$ defends additional arguments: The set S is never part of the grounded extension since it is self-attacking; if $S \setminus \{a\}$ is a subset of the grounded extension, it must be the case that a is already attacked by the grounded extension since there is some $S' \subset S$ such that $S' \cup \{a\}$ attacks some $b \in S \setminus \{a\}$ by definition of tc self-attacks.

Lemma 9. *Let $F = (A, R)$ be a SETAF, let (S, a) be a tc self-attack in F and let G be the SETAF which arises from F by replacing (S, a) by $(S \setminus \{a\}, a)$, i.e., $G = (A, (R \setminus \{(S, a)\}) \cup \{(S \setminus \{a\}, a)\})$. Then $grd(F) = grd(G)$.*

Proof. Let $R_F = R$ and $R_G = R \setminus \{(S, a)\}$. We will prove by induction that $\Gamma^i_F(\emptyset) = \Gamma^i_G(\emptyset)$ for all $i \geq 1$.

Observe that $\Gamma_F(\emptyset) = \Gamma_G(\emptyset)$ because every argument which is attacked in F is also attacked in G (only the source set of the attacks changes potentially).

Now, fix $i > 1$ and assume $\Gamma^i_F(\emptyset) = \Gamma^i_G(\emptyset)$. Let $T = \Gamma^i_F(\emptyset)$. We will show that $\Gamma_F(T) = \Gamma_G(T)$. i.e., T defends the same arguments in both F and G.

We will prove that T attacks a in F iff T attacks a in G. If $S \setminus \{a\} \not\subseteq T$, then the statement follows. Let us therefore assume that $S \setminus \{a\} \subseteq T$. By definition of tc self-attacks, there exist $S' \subset S$ with $a \in S'$ and $b \in S \setminus \{a\}$ such that $(S', b) \in R_F$ (and also $(S', b) \in R_G$). It follows that T attacks a in both F and in G since T defends itself against S'.

Since T attacks a in F iff T attacks a in G, we have $T^+_F = T^+_G$ and thus it follows that T defends the same arguments $b \in A \setminus \{a\}$. We show that a is defended by T in F iff a is defended by T in G. In case a is defended by T in G then T attacks some argument $b \in S \setminus \{a\} \subset S$, therefore T defends a also in F. In case a is defended in F, then T attacks some $b \in S$. Since T attacking a implies that T is conflicting we conclude that $b \in S \setminus \{a\}$, thus the statement follows. □

Similar as in the previous subsection, where we added further properties to inactive attacks in order to become removable, we consider a further restriction to tc self-attacks in order to extend the range of semantics under which such attacks can be removed. The following definition mirrors the concept of back-attacking attacks from Definition 11 in the setting of self-attacks.

Definition 16. Let $F = (A, R)$ be a SETAF. A self-attack $(S, a) \in R$ is called *back-attacking target-conflicting* (batc) self-attack in F if (S, a) is tc self-attacking and there exists $b \in S$ such that $(\{a\}, b) \in R$.

Figure 12a illustrates the concept of batc self-attacks; here, the attacked argument a defends itself against the batc self-attack $(\{a, b, c, d\}, a)$ via the attack $(\{a\}, d)$. We show that batc self-attacks can be removed with respect to all semantics except grounded and complete.

Lemma 10. *Let $F = (A, R)$ be a SETAF, let (S, a) be a batc self-attack in F and let $G = (A, R \setminus \{(S, a)\})$. Then $\sigma(F) = \sigma(G)$ for $\sigma \in \{cf, naive, stb, stage, adm, pref, sem, eager, ideal\}$.*

Proof. We will prove the statement in a similar way like Lemma 5. We write $R_F = R$ and $R_G = R \setminus \{(S, a)\}$ to emphasize the affiliations. For $\sigma \in \{cf, naive, stb, stage\}$, the statement follows by Lemma 8, since every batc self-attack is a tc self-attack. Recall that (1) $cf(F) = cf(G)$ and (2) $T^+_F = T^+_G$ for each $T \in cf(F)$.

(a) SETAF F. (b) SETAF G.

Figure 12: Example of a batc self-attack (in SETAF F, colored in red) and the framework which arises after removing the batc attack (SETAF G).

We will show that (3) $\mathit{adm}(F) = \mathit{adm}(G)$. Since R_G is a proper subset of R_F we have that each conflict-free set T which defends itself in F also defends itself in G, consequently $\mathit{adm}(F) \subseteq \mathit{adm}(G)$. Let us therefore consider an admissible set $T \in \mathit{adm}(G)$. We show that T defends itself in F. In the case $a \in T$, a defends itself against S since, for some $b \in S$, $(\{a\}, b) \in R_G \subset R_F$. In the case $a \notin T$, the set T has the same attacker in both F and G and therefore $\mathit{adm}(F) = \mathit{adm}(G)$.

For the remaining semantics, the statement follows by (3) and by Proposition 1, i.e., $\sigma(F) = \sigma(G)$ for $\sigma \in \{\mathit{pref}, \mathit{sem}, \mathit{ideal}, \mathit{eager}\}$. □

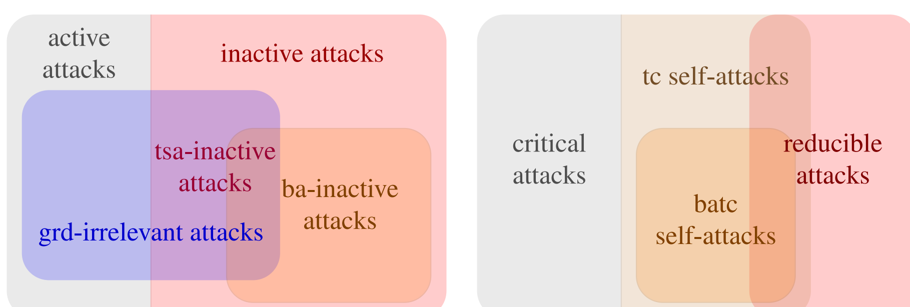
Figure 12 illustrates that batc self-attacks cannot be removed when dealing with grounded and complete semantics. In the Figure we have two SETAFs F, G that only differ by one batc self-attack but we have $\mathit{grd}(F) = \{c\}$ while $\mathit{grd}(G) = \{a, c\}$.

One might ask why we have not adapted the concepts of target-self-attacking (cf. Definition 10) and grd-irrelevant (cf. Definition 12) to self-attacks. The reason is that such adaptations yield simply redundant attacks (cf. Definition 7). Recall that, given a SETAF (A, R), for an attack $(S, a) \in R$ being target-self-attacking (or grd-irrelevant) implies that there exists $S' \subset S$ such that $(S' \cup \{a\}, a) \in R$, too. Clearly the latter attack is redundant in (A, R) in case $a \in S$.

3.3 Summary

So far, we have identified several types of attacks which are removable under different semantics. We have obtained several rules for removing proper attacks and self-attacks; in addition, we have shown that certain self-attacks can be transformed into proper inactive attacks and vice versa.

The distinction of *proper attacks* and *self-attacks* yields a partition of attacks in SETAFs which can be further distinguished into several sub-classes. Figure 13 depicts the (sub-)set-relation of classes of proper attacks (Figure 13a) and self-attacks (Figure 13b). Proper attacks can be further partitioned into *active* and *inactive attacks*, where the latter contains both *ba-inactive* and *tsa-inactive attacks*. Recall that *grd-irrelevant attacks* can be both active or inactive by



(a) Proper attacks. (b) Self-attacks.

Figure 13: Relations between proper attacks (left) and self-attacks (right).

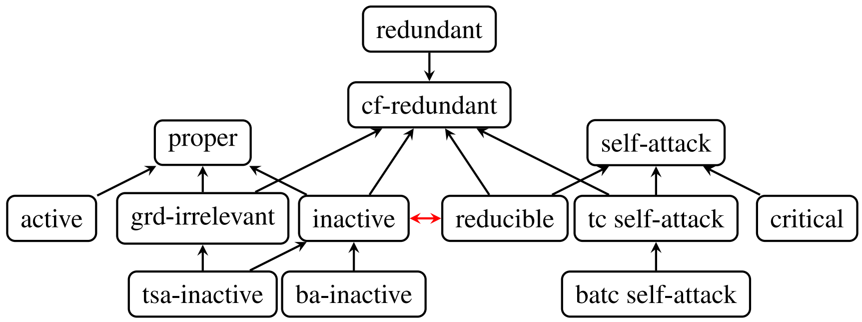


Figure 14: Attacks in SETAFs. A black arrow from node a to b indicates that every a attack is a b attack. The red arrow represents the transformation from reducible attacks to inactive attacks and vice versa.

definition. Furthermore notice that the set of inactive grd-irrelevant attacks contains precisely the tsa-inactive attacks as illustrated in Figure 13a. Self-attacks yield a somewhat different picture. Here, we distinguish between *critical self-attacks* and introduce two (non-disjoint) classes for non-critical attacks, namely *tc self-attacks* and *reducible self-attacks*. The former class also offers a self-attacking analogon of ba-inactive attacks, namely *batc self-attacks*, which is contained in the class of tc self-attacks.

Figure 14 gives a full picture of classes of attacks in SETAFs we considered. Here, a black arrow from node a to b indicates that every a attack is a b attack, e.g., every active attack is a proper attack and every tsa-inactive attack is both grd-irrelevant and inactive. The figure also depicts *redundant* and *cf-redundant attacks* which can be both proper and self-attacking; in fact, every inactive, grd-irrelevant and every non-critical self-attack is cf-redundant as visualized in Figure 14. The red arrow between inactive and reducible attacks relates proper attacks and self-attacks. Recall that every inactive attack can be transformed into a reducible attack and vice versa; in this way, we avoided self-attacking versions of tsa-inactive attacks and ba-inactive attacks which are not tc self-attacking.

Table 1 summarizes the removal rules for each semantics which have been obtained in the current section. While redundant and tsa-inactive attacks are removable with respect to every semantics, the remaining types of attacks can be removed only for a subset of semantics without changing the extensions, e.g., grd-irrelevant attacks can be removed with respect to grounded, naive, and conflict-free semantics. Observe that for the classes of proper attacks, active attacks, self-attacks, and critical attacks, no removal rules are known, thus they are not included in the table.

Table 1: Removable attacks for each semantics.

	com	*grd*	*adm*	*pref*	*sem*	*ideal*	*eager*	*stb*	*stage*	*naive*	*cf*
redundant	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tsa-inactive	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ba-inactive			✓	✓	✓	✓	✓	✓	✓	✓	✓
batc self-attack			✓	✓	✓	✓	✓	✓	✓	✓	✓
inactive								✓	✓	✓	✓
reducible								✓	✓	✓	✓
tc self-attack								✓	✓	✓	✓
cf-redundant										✓	✓
grd-irrelevant		✓								✓	✓

3.4 Normal Forms of SETAFs

Based on the notion of redundant attacks, Polberg [19] introduced the concept of the minimal form of SETAF. The minimal form of a SETAF F is obtained by deleting all redundant attacks from F.

Definition 17. Let $F = (A, R)$ be a SETAF. The *minimal form* F_{min} of F is defined as $F_{min} = (A, R_{min})$ with $R_{min} = R \setminus \{(S, a) \mid (S, a)$ redundant in $F\}$.

As already observed in [19] and indicated by Lemma 1 a SETAF is equivalent to its minimal form.

Proposition 2. *For each SETAF F we have $\sigma(F) = \sigma(F_{min})$ for all semantics σ under consideration.*

Proof. First notice that if (S, a) and (S', b) are both redundant attacks in $F = (A, R)$ then (S, a) is also redundant in $(A, R \setminus \{(S', b)\})$. That is, we can iteratively remove redundant attacks and by Lemma 1 we obtain that $\sigma(F) = \sigma(F_{min})$. □

We next strengthen the minimal form to our normal form of SETAFs by removing further attacks that are irrelevant for the evaluation of the SETAF. In Section 3.2, we have shown that reducible self-attacks can be equivalently represented as inactive attacks and vice versa (cf. Lemma 7). Thus in order to accomplish a uniform representation of SETAFs we will introduce the normal form F_{norm} of a SETAF F where (a) each redundant attack is deleted and (b) each reducible attack is transformed to its inactive pendant. By performing the second step, further redundancies can be obtained: Consider the reducible attack (S, a), then the attack (B, a) with $B \supset S \setminus \{a\}$ is redundant after replacing (S, a) with $(S \setminus \{a\}, a)$. Furthermore also the replacement of inactive attacks with reducible self-attacks can reveal further redundancies: Consider the attacks (S, a), $(S' \cup \{a\}, a)$ where $S' \subset S$ and (S, a) is inactive ((S, a) is in fact a tsa-inactive attack), then the reducible attack $(S \cup \{a\}, a)$ is redundant in this scenario.

In order to eliminate every redundant attack that can be obtained by transforming reducible to inactive attacks and vice versa, we define the normal form as follows: Starting from the minimal form we construct a normal form by first removing all inactive attacks that can be made redundant by transforming it into the equivalent reducible attack and then replacing all reducible attacks by the corresponding inactive attack.

Definition 18. Let $F = (A, R)$ be a SETAF and F_{min} its minimal form. For the *normal form* $F_{norm} = (A, R_{norm})$ of F we first compute

$$R' = R_{min} \setminus \{(S, a) \mid (S, a) \text{ inactive}, \exists (S', a) : S' \setminus a \subset S\}$$

and then replace each reducible attack $(S, a) \in R'$ by the equivalent inactive attack $(S \setminus \{a\}, a) \in R'$ to obtain the attack relation R_{norm} of the normal form.

Remark 2. In our conference version [13], the normal form of a SETAF $F = (A, R)$ is defined as the minimal form of $(A, R \cup (\{S \setminus \{a\}, a) \mid (S, a)$ is reducible in $F\})$. Thus in contrast to the present version, the attacks which get redundant by replacing inactive attacks with reducible attacks (i.e., tsa-inactive attacks) are not removed. The present version therefore strengthens the definition of the normal form in [13] by identifying and removing additional redundancies.

We next show that the normal form is indeed equivalent to the original SETAF (under all semantics we consider). The key observation is that we delete exactly the tsa-inactive attacks. For our proof we need the following technical lemmas. The first stating that a tsa-inactive attack remains tsa-inactive when deleting other tsa-inactive attacks.

Lemma 11. *For SETAF $F = (A, R)$ with different attacks $(S, a), (S', b) \in R$ being tsa-inactive in F, the attack (S, a) is tsa-inactive in $G = (A, R \setminus \{(S', b)\})$.*

Proof. Let $R_F = R$ and $R_G = R \setminus \{(S', b)\}$. As (S, a) is tsa-inactive there are attacks $(T, c), (U, a) \in R_F$ such that $T \cup \{c\} \subset S$ and $a \in U$. By the latter we have that (U, a) cannot be tsa-inactive and thus $(U, a) \in R_G$. If $(T, c) \neq (S', b)$ then $(T, c) \in R_G$ and thus (S, a) is tsa-inactive in G as well. Otherwise, as (S', b) is tsa-inactive there is $(T', c') \in R_F$ such that $T' \cup \{c'\} \subset S' \subset S$. Then we have $(T', c') \in R_G$ and thus again (S, a) is tsa-inactive in G. □

We next give a similar lemma for reducible attacks.

Lemma 12. *For SETAF $F = (A, R)$ with two different attacks $(S, a), (S', b) \in R$ being reducible in F, the attack (S, a) is reducible in $G = (A, (R \cup \{(S \setminus \{b\}, b)\}) \setminus \{(S', b)\})$.*

Proof. Let $R_F = R$ and $R_G = (R \cup \{(S \setminus \{b\}, b)\}) \setminus \{(S', b)\}$. As (S, a) is reducible there is an attack $(T, c) \in R_F$ such that $T \cup \{c\} \subseteq S \setminus \{a\}$. If $(T, c) \neq (S', b)$ then $(T, c) \in R_G$ and thus (S, a) is reducible in G as well. Otherwise, as (S', b) is reducible there is $(T', c') \in R_F$ such that $T' \cup \{c'\} \subset S' \subset S \setminus \{a\}$. Then we have $(T', c') \in R_G$ and thus again (S, a) is reducible in G. ☐

We are now ready to show the equivalence between a SETAF and its normal form.

Proposition 3. *For each SETAF F we have $\sigma(F) = \sigma(F_{norm})$ for all semantics σ under consideration.*

Proof. Let $F = (A, R)$ and R' be as in Definition 18. We already know that $\sigma(F) = \sigma(F_{min})$. In a first step we show that $\sigma(F_{min}) = \sigma(A, R')$. Notice that the set $\{(S, a) \mid (S, a) \text{ inactive}, \exists (S', a) : S' \setminus a \subset S\}$ contains exactly the tsa-inactive attacks of F. Thus by Lemma 4 and Lemma 11 we can remove them one by one and end up with (A, R') and $\sigma(F_{min}) = \sigma(A, R')$. Now by Lemma 7 and Lemma 12 we can replace reducible self-attacks by the equivalent proper attacks one-by-one and obtain F_{norm} and $\sigma(F) = \sigma(F_{norm})$. ☐

By the above we can always obtain a simplified SETAF that has neither redundant, reducible, nor tsa-inactive attacks.

Corollary 1. *For each SETAF $F = (A, R)$ we have that its normal form $\sigma(F_{norm})$ has neither redundant, reducible, nor tsa-inactive attacks.*

Proof. Towards a contradiction assume that there would be such an attack (S, a). First, assume $(S, a) \in R_{norm}$ is redundant and thus there is $(S', a) \in R_{norm}$ with $S' \subset S$. By definition only one of them can be in the minimal form F_{min} and thus either $(S \cup \{a\}, a)$ is reducible in F_{min} or $(S' \cup \{a\}, a)$ is reducible in F_{min}. The former attack would be redundant in F_{min}; a contradiction. We thus have $(S, a) \in R_{min}$ and $(S' \cup \{a\}, a) \in R_{min}$ and as the later is reducible both are inactive. That is (S, a) would have been removed in the normal form, a contradiction to our initial assumption.

Second, assume $(S, a) \in R_{norm}$ is reducible and thus there is $(S', b) \in R_{norm}$ with $S' \cup \{b\} \subset S \setminus \{a\}$. By construction either $(S', b) \in R'$ or $(S' \cup \{b\}, b) \in R'$. In both cases (S, a) is reducible in R' and thus would have been removed, a contradiction.

Finally, assume $(S, a) \in R_{norm}$ that is tsa-inactive in F_{norm}. That is there is an attack $(S' \cup \{a\}, a) \in R_{norm}$ with $S' \subseteq S$ and an attack $(T, b) \in R_{norm}$ with $T \cup \{b\} \subseteq S$. The former attack is also in F_{min} and thus $(S \cup \{a\}, a) \notin R_{min}$ Together with the attack $(T, b) \in R_{norm}$ we can infer that (S, a) is inactive in F_{min}. That is (S, a) is tsa-inactive in F_{min} and thus would have been removed, a contradiction. ☐

4 Characterizations of Strong Equivalence

In this section we investigate when attacks can be safely removed from a SETAF without affecting the semantics no matter how the SETAF is extended later on. That is we investigate the notion of strong equivalence and variants thereof in the context of SETAFs, generalizing previous work for AFs [1, 18]. We first show that the normal form translation preserves strong equivalence for all semantics under consideration. Then we introduce different notions of kernels and show that the transformation of a SETAF into its kernel form preserves strong equivalence for particular semantics. Moreover, these kernels provide exact characterizations for strong equivalence, i.e., two SETAFs are strongly equivalent if and only if their kernels are in a certain correspondence. Finally, we clarify the relations between the different kernels and relate strong equivalence with respect to different semantics. Notable all relations known from Dung AFs carry over to SETAFs.

We define the notion of strong equivalence for SETAFs along the lines of [18].

Definition 19. Two SETAFs F and G are *strongly equivalent* to each other with respect to a semantics σ, in symbols $F \equiv_s^\sigma G$, iff for each SETAF H, $\sigma(F \cup H) = \sigma(G \cup H)$ holds.

By definition, we have that $F \equiv_s^\sigma G$ implies $\sigma(F) = \sigma(G)$, i.e., standard equivalence between F and G with respect to σ. However, no matter which of the considered semantics we choose for σ, the converse direction does not hold in general (cf. [18]).

We consider two weakenings for Definition 19 by restricting the potential context SETAF H. First, we let H to be only an AF instead of a SETAF. We consider this an interesting restriction in the sense of whether an AF is sufficient to reveal the potential difference between the compared SETAFs in terms of strong equivalence. Another weakening has first been proposed in [1] under the name *normal expansion equivalence*. Here the framework H is not allowed to add attacks between "existing" arguments (in F or G), and thus better reflects that in dynamic scenarios new arguments may be proposed but the relation between given arguments remains unchanged.

Definition 20. Let F and G be SETAFs and σ be a semantics. Moreover, let $B = A_F \cup A_G$. We write

- $F \equiv_n^\sigma G$, if for each SETAF H with $R_H \cap (2^B \times B) = \emptyset$, $\sigma(F \cup H) = \sigma(G \cup H)$;
- $F \equiv_{sd}^\sigma G$, if for each AF H, $\sigma(F \cup H) = \sigma(G \cup H)$;
- $F \equiv_{nd}^\sigma G$, if for each AF H with $R_H \cap (B \times B) = \emptyset$, $\sigma(F \cup H) = \sigma(G \cup H)$.

Note that for any SETAFs F, G, (i) $F \equiv_s^\sigma G$ implies $F \equiv_n^\sigma G$; (ii) $F \equiv_s^\sigma G$ implies $F \equiv_{sd}^\sigma G$; (iii) $F \equiv_n^\sigma G$ implies $F \equiv_{nd}^\sigma G$; and (iv) $F \equiv_{sd}^\sigma G$ implies $F \equiv_{nd}^\sigma G$. We will implicitly make use of these implications in the following subsections.

4.1 Normal Form and Strong Equivalence

We will show that for each SETAF F, its minimal form and its normal form are strongly equivalent to F with respect to all semantics under consideration. The following lemma is crucial as it states that an attack that falls into a class of attacks in F that can be safely removed from F is in the same class of attacks in $F \cup H$.

Lemma 13. *Let F be a SETAF, $(S, a) \in R_F$ and let $p \in \{$redundant, cf-redundant, grd-irrelevant, inactive, tsa-inactive, ba-inactive, reducible, tc, batc$\}$. If (S, a) is p in F then (S, a) is p in $F \cup H$ for every SETAF H.*

Proof. Consider an arbitrary SETAF H. Observe that, for every property p under consideration, (S, a) satisfies the property p *if and only if* there exists a set of attacks $R' \subseteq R_F$ satisfying specific requirements: In the case of redundant attacks, $R' = \{(S', a)\}$ with $S' \subset S$; for cf-redundant attacks, $R' = \{(S', b)\}$ with $S' \cup \{b\} \subset S \cup \{a\}$; an attack is inactive iff there is $R' = \{(S', b)\}$ with $S' \subseteq S$, $b \in S$; tsa-inactive attacks strengthen this definition by requiring $R' = \{(S', b), (S'' \cup \{a\}, a)\}$ with $S', S'' \subseteq S$, $b \in S$; ba-inactive attacks require $R' = \{(S', b), (\{a\}, c)\}$ with $S' \subseteq S$, $b, c \in S$; in the case of grd-irrelevant attacks, $R' = \{(S' \cup \{a\}, a), (S'', b)\}$ with $S' \subset S$, $b \in S$, $S'' \subseteq S \cup \{a\}$; for reducible attacks, $R' = \{(S', b)\}$ with $S' \subset S \setminus \{a\}$, $b \in S$; for tc self-attacks, $R' = \{(S', b)\}$ with $S' \subset S$, $a \in S'$, $b \in S$; batc self-attacks require the existence of a set $R' = \{(S', b), (\{a\}, c\}$ with $S' \subset S$, $a \in S'$, $b, c \in S$.

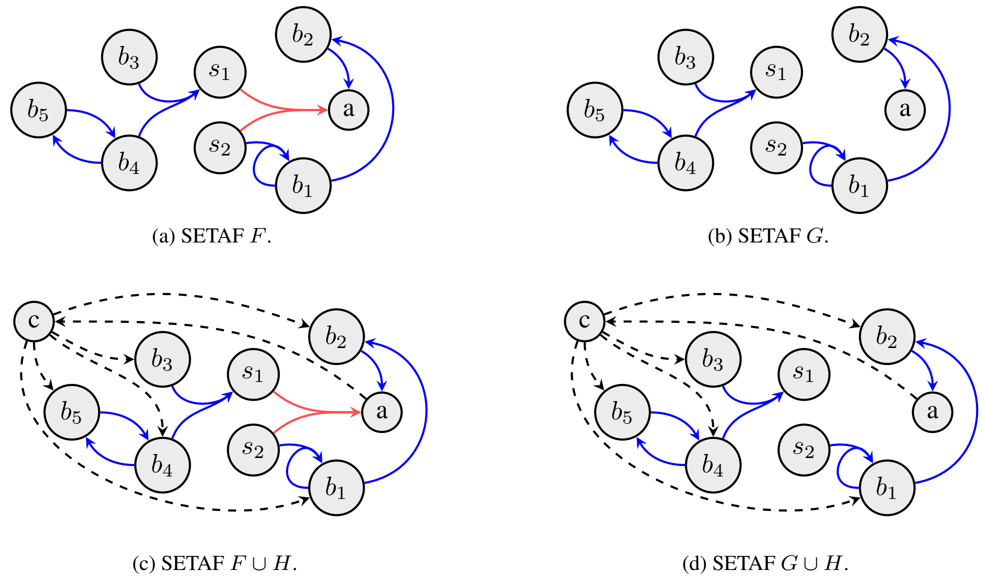
It is crucial that neither of the properties p under consideration is defined based on the absence of an attack. This is in contrast to proper and critical attacks which require that no conflict on a (proper) subset of the source set exists. Thus the property p is preserved in any expansion of F, i.e., if the attack (S, a) is p in F it remains p in $F \cup H$ since $R' \subseteq R_F \subseteq R_{F \cup H}$. □

With the above lemma at hand we can show that both the minimal form F_{min} and the normal form F_{norm} are strongly equivalent to the original SETAF F.

Proposition 4. *For a SETAF F and its minimal form F_{min}, we have $F \equiv_\times^\sigma F_{min}$, for $\times \in \{s, sd, n, nd\}$ and all semantics σ under consideration.*

Proof. We first show $F \equiv_s^\sigma F_{min}$. Let R be the set of redundant attacks in F and consider an arbitrary SETAF H. By Lemma 13, the attacks in R are also redundant in the SETAF $F \cup H$ and thus, by Lemma 1 and Proposition 2, we can iteratively remove them and obtain $\sigma(F \cup H) = \sigma((F \cup H) \setminus R) = \sigma((F \setminus R) \cup H) = \sigma(F_{min} \cup H)$. Now as $\sigma(F \cup H) = \sigma(F_{min} \cup H)$ for each SETAF H we can conclude that $F \equiv_s^\sigma F_{min}$. It follows that $F \equiv_{sd}^\sigma F_{min}$, $F \equiv_n^\sigma F_{min}$ and $F \equiv_{nd}^\sigma F_{min}$. □

Proposition 5. *For a SETAF F and its normal form F_{norm}, we have $F \equiv_\times^\sigma F_{norm}$, for $\times \in \{s, sd, n, nd\}$ and all semantics σ under consideration.*



(a) SETAF F. (b) SETAF G.

(c) SETAF $F \cup H$. (d) SETAF $G \cup H$.

Figure 15: Example of two SETAFs F and G which disagree on their kernels (witnessed by the attack $(\{s_1, s_2\}, a)$ colored in red) but $com(F) = com(G)$ and the extended SETAFs $F \cup H$, $G \cup H$ where $com(F \cup H) \neq com(G \cup H)$ holds.

Proof. We first show $F \equiv_s^\sigma F_{norm}$. Let R_{rdt} R_{rdb}, R_{tsa} denote the sets of redundant, reducible and, respectively, tsa-inactive attacks in F and consider an arbitrary SETAF H. By Lemma 13, every attack in R_{rdt} R_{rdb}, R_{tsa} remains redundant, reducible and, respectively, tsa-inactive in the SETAF $F \cup H$. By Lemma 1, Lemma 4, and Lemma 11, redundant and tsa-inactive attacks can be iteratively removed, thus we obtain $\sigma(F \cup H) = \sigma(((F \cup H) \setminus R_{drt}) \setminus R_{tsa}) = \sigma((F \cup H) \setminus (R_{drt} \cup R_{tsa}))$. By Lemma 7 and Lemma 12, we can replace every attack $(S, a) \in R_{rdb} \setminus (R_{rdt} \cup R_{tsa})$ with $(S \setminus \{a\}, a)$ and obtain $\sigma(F \cup H) = \sigma(((F \cup H) \setminus (R_{drt} \cup R_{tsa} \cup R_{rdb})) \cup \{(S \setminus \{a\}, a) \mid (S, a) \in R_{rdb} \setminus (R_{rdt} \cup R_{tsa})\}) = \sigma(F_{norm} \cup H)$. Now as $\sigma(F \cup H) = \sigma(F_{norm} \cup H)$ for each SETAF H we obtain that $F \equiv_s^\sigma F_{norm}$. It follows that $F \equiv_{sd}^\sigma F_{norm}$, $F \equiv_n^\sigma F_{norm}$ and $F \equiv_{nd}^\sigma F_{norm}$. $\square$

4.2 Complete Semantics

In this and the forthcoming sections we show that strong equivalence with respect to a semantics σ can be decided by comparing particular subframeworks, the so-called σ-kernels of two SETAFs. All types of kernels are built from the arguments of the original SETAF but the attack relation undergoes certain changes.

For complete semantics, the kernel coincides with the normal form we have introduced in Section 4.1.

Definition 21. For a SETAF $F = (A, R)$, we define the *complete kernel* $F^{ck} = (A, R^{ck})$ of F as $F^{ck} = F_{norm}$ (i.e., $R^{ck} = R_{F_{norm}}$).

We will show that two SETAFs are strongly equivalent with respect to complete semantics under all considered variants of strong equivalence if and only if their complete kernels coincide. Observe that one direction is immediate by results from Section 4.1 where we have shown that each SETAF is strongly equivalent to its normal form (cf. Proposition 5); it follows that two SETAFs are strongly equivalent with respect to complete semantics if their kernels are syntactically equivalent. It remains to show that we can construct a counter example in case that the kernels of two SETAFs F and G disagree on. Figure 15 exemplifies such a case together with the construction of a SETAF H acting as counter example: We consider two SETAFs $F = (A, R_F)$ and $G = (A, R_G)$ which are depicted in Figures 15a and 15b where $F^{ck} \neq G^{ck}$ (the attack $(\{s_1, s_2\}, a)$ is present in F^{ck} but not in G^{ck}). Observe that $com(F) = com(G) = \{\{b_3, s_2\}, \{b_3, b_5, s_1, s_2\}, \{b_3, b_4, s_2\}\}$. We extend both F and G by a framework $H = (A \cup \{c\}, \{(a, c)\} \cup \{(c, b) \mid$

$b \in A \setminus (S \cup \{a\})\})$, which serves as a witness for $F \not\equiv_s^{com} G$: Observe that the newly introduced argument c attacks every argument $b \notin S \cup \{a\}$ and is attacked by a, thus $\{c, s_1, s_2\}$ defends itself in $F \cup H$ but not in $G \cup H$ as c is not defended in $G \cup H$ against a. We have $com(F \cup H) = \{\{s_2\}, \{c, s_1, s_2\}\} \neq \{\{s_2\}\} = com(G \cup H)$ as $\{c, s_1, s_2\}$ is not complete in $G \cup H$.

The main part of the proof of the following theorem consists of similar considerations depending on different cases two kernels can disagree on. In the example, we considered the case where the arguments of the kernels coincide and the attack $(\{s_1, s_2\}, a) \in F^{ck} \setminus G^{ck}$ is active (cf. case (1) of the proof).

Theorem 1. *For any two SETAFs F, G, the following are equivalent: (a) $F \equiv_s^{com} G$; (b) $F \equiv_n^{com} G$; (c) $F \equiv_{sd}^{com} G$; (d) $F \equiv_{nd}^{com} G$; (e) $F^{ck} = G^{ck}$.*

Proof. By definition, (a) implies (b) and (c). Likewise, (b) implies (d) and (c) implies (d). It remains to show (1) $F^{ck} = G^{ck}$ implies $F \equiv_s^{com} G$ and (2) $F^{ck} \neq G^{ck}$ implies $F \not\equiv_{nd}^{com} G$. Let $F^{ck} = (A_F^{ck}, R_F^{ck})$ and $G^{ck} = (A_G^{ck}, R_G^{ck})$.

(1) By Proposition 5, $F \equiv_s^{com} F^{ck}$ and $G \equiv_s^{com} G^{ck}$. Now as $F^{ck} = G^{ck}$, by the transitivity of $\equiv_s^{com}$, we obtain $F \equiv_s^{com} G$.

(2) Now, suppose that $F^{ck} \neq G^{ck}$, we show $F \not\equiv_{nd}^{com} G$. In case $com(F^{ck}) \neq com(G^{ck})$, we are done: By Proposition 5, $F \equiv_s^{com} F^{ck}$ and $G \equiv_s^{com} G^{ck}$, therefore we get that $com(F) \neq com(G)$. Consequently, $F \not\equiv_{nd}^{com} G$. Thus we assume $com(F^{ck}) = com(G^{ck})$.

First, we consider the case $A_F^{ck} \neq A_G^{ck}$. This implies $A_F \neq A_G$. Without loss of generality, let $a \in A_F \setminus A_G$. Notice that $a \notin S$ for all $S \in com(F^{ck})$ since $com(F^{ck}) = com(G^{ck})$ by assumption. Let $H = (\{a\}, \emptyset)$. Then $a \in S$ for all $S \in com(G \cup H)$, as a is not attacked at all in $G \cup H$. On the other hand, we have $F \cup H = F$ and thus $com(F \cup H) = com(F) = com(F^{ck})$. Consequently, $F \not\equiv_{nd}^{com} G$.

Now, suppose $A_F^{ck} = A_G^{ck}$ (and therefore, $A_F = A_G$) and $R_F^{ck} \neq R_G^{ck}$. We show $F^{ck} \not\equiv_{nd}^{com} G^{ck}$: Without loss of generality, there exists $(S, a) \in R_F^{ck} \setminus R_G^{ck}$ such that there is no $(S', b) \in R_G^{ck} \setminus R_F^{ck}$ with $|S' \cup \{b\}| < |S \cup \{a\}|$ (otherwise exchange the roles of F and G); we refer to this below as minimality assumption. The complete kernel does neither contain reducible, redundant nor tsa-inactive attacks. We consider therefore the following cases: (1) (S, a) is active in F^{ck}, (2) (S, a) is a critical self-attack in F^{ck}, (3) (S, a) is inactive in F^{ck} and (4) (S, a) is a tc self-attack in F^{ck}.

For fresh arguments c, d, we define

$$H_1 = (A_F \cup \{c\}, \{(a, c)\} \cup \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\}),$$
$$H_2 = (A_F \cup \{c, d\}, \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(d, d), (d, b) \mid b \in S \setminus \{a\}\}).$$

1. Let (S, a) be active in F^{ck}. We show that $com(F^{ck} \cup H_1) \neq com(G^{ck} \cup H_1)$ by generalizing the construction presented in Figure 15: First observe that S is not conflicting in F. It follows that $E = S \cup \{c\}$ is stable in $F^{ck} \cup H_1$: E is conflict-free and attacks all arguments in $A_F \setminus E$ by construction. Thus $E \in com(F^{ck} \cup H_1)$. Observe that E is not admissible (and thus also not complete) in $G^{ck} \cup H_1$, since $(S', a) \notin R_G^{ck}$ for all $S' \subseteq S$, but $(a, c) \in R_G^{ck} \cup R_{H_1}$.

2. Let $(S, a), a \in S$ be critical in F^{ck}, we show that $com(F^{ck} \cup H_2) \neq com(G^{ck} \cup H_2)$. We have $E = \{a, c\} \notin com(F^{ck} \cup H_2)$, since a is not defended against S in F^{ck}. However, $E \in com(G^{ck} \cup H_2)$: E is conflict-free by construction; moreover, E defends itself in G^{ck}: By construction all arguments in $A_G^{ck} \setminus S$ are attacked by c, thus a is defended against each attack (T, a) with $T \setminus S \neq \emptyset$. Now assume there is $(T, a) \in R_G^{ck}$ such that $T \subset S$ (note that $T \neq S$ by assumption $(S, a) \notin R_G^{ck}$). In case $T = S \setminus \{a\}$, i.e., if $(S \setminus \{a\}, a)$ is active in G consider case (1). Else $|T \cup \{a\}| < |S \cup \{a\}|$ and therefore $(T, a) \in R_F^{ck}$ by assumption, which makes (S, a) redundant in F, contradiction. Furthermore, E contains all arguments it defends, since c attacks all arguments in $A_F \setminus S$ and E does not defend any argument $b \in S \setminus \{a\}$ against d.

3. Let (S, a) be inactive in F^{ck}, we show that $com(F^{ck} \cup H_2) \neq com(G^{ck} \cup H_2)$. $\{c\}$ is complete in $F^{ck} \cup H_2$: c is not attacked at all, arguments in $A_F \setminus (S \cup \{a\})$ are not defended against c, arguments in S are not defended against d and a is not defended against S. On the other hand, $\{c\}$ is not complete in $G^{ck} \cup H_2$ since c defends a. Indeed, if there is an attack $(T, a) \in R_G^{ck}$ such that $T \subseteq S \cup \{a\}$, then either $|T \cup \{a\}| < |S \cup \{a\}|$ and thus $(T, a) \in R_F^{ck}$, contradiction to F^{ck} being in minimal form since (S, a) would be redundant in F; or

$|T \cup \{a\}| = |S \cup \{a\}|$, that is, $T = S$, contradiction to $(S, a) \notin R_G^{ck}$, or $T = S \cup \{a\}$, which is in contradiction to F^{ck} being in normal form since (T, a) would be reducible in G.

4. Let (S, a) be a tc self-attack in F^{ck}, we show that $\mathit{com}(F^{ck} \cup H_2) \neq \mathit{com}(G^{ck} \cup H_2)$. $\{c\}$ is complete in $F^{ck} \cup H_2$: c is not attacked at all, arguments in $A_F \setminus S$ are not defended against c, arguments in $S \setminus \{a\}$ are not defended against d and a is not defended against S. On the other hand, $\{c\}$ is not complete in $G^{ck} \cup H_2$ since c defends a. Towards a contradiction assume that c does not defend a. Then there is an attack $(T, a) \in R_G^{ck}$ such that $T \subseteq S$. In case $|T \cup \{a\}| < |S \cup \{a\}|$ we obtain $(T, a) \in R_F^{ck}$ by minimality assumption, contradiction to F^{ck} being in minimal form. In case $|T \cup \{a\}| = |S \cup \{a\}|$, either $T = S \setminus \{a\}$, then (T, a) is active in G^{ck} and we consider case (1); or $T = S$, which is in contradiction to $(S, a) \notin R_G^{ck}$.

In all cases, we found a witness (H_1 or H_2) showing that $F^{ck} \not\equiv_{nd}^{com} G^{ck}$. By $F \equiv_{nd}^{com} F^{ck}$ and $G \equiv_{nd}^{com} G^{ck}$ we conclude $F \not\equiv_{nd}^{com} G$. □

The above result generalizes the corresponding result for complete semantics in AFs [18]. The complete kernel of an AF $F = (A, R)$ has been defined as $F^{ck} = (A, R^{ck})$ with $R^{ck} = R \setminus \{(a, b) \mid a \neq b, (a, a), (b, b) \in R\}$. Observe that, for AFs, the normal form collapses precisely to the complete kernel defined in [18] since neither redundant nor reducible attacks exist in AFs (self-attacking sets are simply of the form (a, a) for some argument a); moreover, tsa-inactive attacks are of the form (a, b) with $a \neq b$ where both a and b are self-attacking. Thus the complete kernel for SETAFs is indeed a proper extension of the complete kernel for AFs.

4.3 Admissible-Based Semantics

For admissible-based semantics, the kernel differs from the complete kernel in the way that additional attacks can be removed. The admissible kernel F^{ak} is constructed by first computing the normal form F_{norm} before removing ba-inactive attacks and batc self-attacks.

Definition 22. For a SETAF $F = (A, R)$ in normal form, we define the *admissible kernel* of F as $F^{ak} = (A, R^{ak})$ with

$$R^{ak} = R \setminus \{(S, a) \in R \mid (S, a) \text{ is a ba-inactive attack or a batc self-attack in } F\}.$$

The admissible kernel of an arbitrary SETAF F is the admissible kernel of the normal form of F.

We will show that every SETAF F is strongly equivalent to its admissible kernel F^{ak} with respect to admissible, preferred, semi-stable, ideal, and eager semantics. To do so, we need the following technical lemma, which states that both ba-inactive attacks and batc self-attacks can be safely removed.

Lemma 14. *Let $F = (A, R)$ in normal form with two different attacks $(S, a), (S', b) \in R$ being either ba-inactive attacks or batc self-attacks (independently of each other). Then the attack (S, a) remains a ba-inactive attack (a batc self-attack) in $G = (A, R \setminus \{(S', b)\})$.*

Proof. Let $R_G = R \setminus \{(S', b)\}$. Since (S, a) is ba-inactive or batc self-attacking, there are attacks (T, c), $(\{a\}, d)$ with $T \subseteq S$, $c, d \in S$. In case (S, a) is a batc self-attack we furthermore have that $T \subset S$, $a \in T$ and $c \in S \setminus \{a\}$.

If $(S', b) \neq (T, c)$ and $(S', b) \neq (\{a\}, d)$, we are done since in this case, $(T, c), (\{a\}, d) \in R_G$. Moreover, $(S', b) \neq (\{a\}, d)$ since the attack $(\{a\}, d)$ is neither a ba-inactive attack nor a batc self-attack: In both cases this would imply that $(\{a\}, a) \in R$ and thus (S, a) is tsa-inactive, contradiction to our assumption that F is in normal form. Thus $(\{a\}, d) \in R_G$. Now, let $(S', b) = (T, c)$ and consider the cases (1) (S, a) is a ba-inactive attack; (2) (S, a) is a batc self-attack.

1. Let (S, a) be a ba-inactive attack.

   First consider the case (S', b) is ba-inactive. Then there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \subseteq S$, $b' \in S'$ by inactivity of (S', b). Consequently, $(S'', b'), (\{a\}, d) \in R_G$, $b' \in S$ and thus (S, a) is ba-inactive in G.

   In case (S', b) is a batc self-attack, then (S', b) is tc and thus there is an attack $(S'', b') \in R$ with $S'' \subset S' \subseteq S$, $b \in S''$ and $b' \in S' \setminus \{b\}$. Since $(S'', b'), (\{a\}, d) \in R_G$, $b' \in S$, we have shown that (S, a) is ba-inactive in G.

2. Let (S, a) be a batc self-attack.

   First consider the case (S', b) is ba-inactive. Then there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \subseteq S$, $b' \in S'$ by inactivity of (S', b). Notice that $b' \neq a$, otherwise (S, a) is redundant, contradiction to F being in normal form. Moreover, $a \in S''$, otherwise (S, a) is reducible, contradiction to F being in normal form. Consequently, $(S'', b'), (\{a\}, d) \in R_G$, $a \in S''$ and $b' \in S \setminus \{a\}$ and thus (S, a) is a batc self-attack in G.

   In case (S', b) is a batc self-attack, then (S', b) is tc and thus there is an attack $(S'', b') \in R$ with $S'' \subset S' \subseteq S$, $b \in S''$ and $b' \in S' \setminus \{b\}$. We can conclude that $b' \neq a$ and $a \in S''$ by the assumption F being in normal form. Consequently, $S'' \subset S$, $a \in S''$, $b' \in S \setminus \{a\}$ and $(S'', b'), (\{a\}, d) \in R_G$, therefore we have shown that (S, a) is a batc self-attack in G. □

With the above lemmas we next show that a SETAF is strongly equivalent to its admissible kernel for $\sigma \in \{adm, pref, sem, ideal, eager\}$.

Proposition 6. *For every SETAF F and its admissible kernel F^{ak}, $F \equiv^\sigma_s F^{ak}$ for $\sigma \in \{adm, pref, sem, ideal, eager\}$.*

Proof. Let $\sigma \in \{adm, pref, sem, ideal, eager\}$ and consider an arbitrary SETAF H. By Proposition 5, $F \equiv^\sigma_s F_{norm}$, i.e., $\sigma(F \cup H) = \sigma(F_{norm} \cup H)$. By Lemma 13, every ba-inactive attack (every batc self-attack) remains ba-inactive (batc self-attacking) in $F \cup H$. By Lemma 5, Lemma 10, and Lemma 14, ba-inactive attacks and batc self-attacks can be iteratively removed in the SETAF $F_{norm} \cup H$ without changing the σ-extensions and thus we obtain $\sigma(F \cup H) = \sigma(F^{ak} \cup H)$ and therefore $F \equiv^\sigma_s F^{ak}$. □

As a consequence of Proposition 6 we get that two SETAFs F and G which agree on their admissible kernel are strongly equivalent with respect to the admissible-based semantics under consideration. The following theorem states that also the other direction holds, that is, strong equivalence and all considered variants thereof with respect to admissible, preferred, semi-stable, ideal, and eager semantics can be characterized via the admissible kernel.

Theorem 2. *For any two SETAFs F, G and for $\sigma \in \{adm, pref, sem\, ideal, eager\}$, the following are equivalent: (a) $F \equiv^\sigma_s G$; (b) $F \equiv^\sigma_n G$; (c) $F \equiv^\sigma_{sd} G$; (d) $F \equiv^\sigma_{nd} G$; (e) $F^{ak} = G^{ak}$.*

Proof. Let $\sigma \in \{adm, pref, sem, ideal, eager\}$. By definition, (a) implies (b) and (c). Likewise, (b) implies (d) and (c) implies (d). It remains to show (1) $F^{ak} = G^{ak}$ implies $F \equiv^\sigma_s G$ and (2) $F^{ak} \neq G^{ak}$ implies $F \not\equiv^\sigma_{nd} G$. Let $F^{ak} = (A^{ak}_F, R^{ak}_F)$ and $G^{ak} = (A^{ak}_G, R^{ak}_G)$.

(1) By Proposition 6, $F \equiv^\sigma_s F^{ak}$ and $G \equiv^\sigma_s G^{ak}$. Now as $F^{ak} = G^{ak}$, by the transitivity of $\equiv^\sigma_s$, we obtain $F \equiv^\sigma_s G$.

(2) Now, suppose that $F^{ak} \neq G^{ak}$. We show $F \not\equiv^\sigma_{nd} G$. In case $\sigma(F^{ak}) \neq \sigma(G^{ak})$ we are done: By Proposition 6, $F \equiv^\sigma_s F^{ak}$ and $G \equiv^\sigma_s G^{ak}$, therefore we get that $\sigma(F) \neq \sigma(G)$. Consequently, $F \not\equiv^\sigma_{nd} G$. Thus we assume $\sigma(F^{ak}) = \sigma(G^{ak})$.

First, we consider the case $A^{ak}_F \neq A^{ak}_G$. This implies $A_F \neq A_G$. Without loss of generality, let $a \in A_F \setminus A_G$. Notice that $a \notin S$ for all $S \in \sigma(F^{ak})$ since $\sigma(F^{ak}) = \sigma(G^{ak})$ by assumption. Let $H = (\{a\}, \emptyset)$. Then $\{a\} \in adm(G \cup H)$. Moreover, $a \in S$ for all $S \in \sigma(G \cup H)$ for $\sigma \in \{pref, sem\}$: First note that $a \in S$ for all $S \in pref(G \cup H)$ since a is not attacked in $G \cup H$. But then $a \in S$ for all $S \in sem(G \cup H)$ since $sem(G \cup H) \subseteq pref(G \cup H)$. Moreover, $\{a\} \in adm(G \cup H)$ and $a \in \bigcap_{S \in \tau(G \cup H)} S$ for $\tau \in \{pref, sem\}$, thus a is contained in the ideal and in the eager extension of $G \cup H$. On the other hand, we have $\sigma(F \cup H) = \sigma(F)$, thus $F \not\equiv^\sigma_{nd} G$ for $\sigma \in \{adm, pref, sem, ideal, eager\}$.

Now, suppose $A^{ak}_F = A^{ak}_G$ (and therefore, $A_F = A_G$) and $R^{ak}_F \neq R^{ak}_G$. We show $F^{ak} \not\equiv^\sigma_{nd} G^{ak}$: Without loss of generality, there exists $(S, a) \in R^{ak}_F \setminus R^{ak}_G$ such that there is no $(S', b) \in R^{ak}_G \setminus R^{ak}_F$ with $|S' \cup \{b\}| < |S \cup \{a\}|$ (otherwise exchange the roles of F and G).

By definition, the admissible kernel neither possesses redundant, tsa-inactive, reducible, ba-inactive attacks nor batc self-attacks. We distinguish therefore the following cases: (1) (S, a) is active in F^{ak}, (2) (S, a) is critical in F^{ak}, (3) (S, a) is inactive in F^{ak} and (4) (S, a) is a tc self-attack in F^{ak}.

For fresh arguments c, d we define

$$H_1 = (A_F \cup \{c, d\}, \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(a, c), (c, d), (d, d)\} \cup \{(d, b) \mid b \in A_F\}),$$
$$H_2 = (A_F \cup \{c, d\}, \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(d, d), (d, b) \mid b \in S \setminus \{a\}\}).$$

1. Let (S,a) be active in F^{ak}. We show that $\sigma(F^{ak} \cup H_1) \neq \sigma(G^{ak} \cup H_1)$. Observe that $E = S \cup \{c\}$ is the unique non-empty admissible set in $F^{ak} \cup H_1$: E is conflict-free and attacks all arguments in $A_F \setminus E$. Every non-empty admissible set must contain c since there is no other argument attacking d (and d attacks every argument $b \in A_F$). Moreover, S is contained in every non-empty admissible set since c is attacked by a and there is no other set $T \subseteq A_F$ such that both T is not attacked by c and $(T,a) \in R_F^{ak}$ holds. It follows that E is the unique non-empty admissible set and thus the unique preferred, semi-stable, ideal and eager extension. Observe that E is not admissible (thus also neither preferred, semi-stable, ideal nor eager) in $G^{ak} \cup H_1$, since $(S',a) \notin R_G^{ak}$ for all $S' \subseteq S$, but $(a,c) \in R_G^{ak} \cup R_{H_1}$. Therefore $\sigma(F^{ak} \cup H_1) \neq \sigma(G^{ak} \cup H_1)$.

2. Let (S,a) be critical in F^{ak}, we show that $\sigma(F^{ak} \cup H_2) \neq \sigma(G^{ak} \cup H_2)$. We have $E = \{a,c\} \notin \mathit{adm}(F^{ak} \cup H_2)$, since a is not defended against S in F^{ak}. However, $E \in \mathit{adm}(G^{ak} \cup H_2)$. E is conflict-free by construction (note that $(\{a\},a) \notin R_F^{ak}$ otherwise (S,a) is redundant in F, and thus $(\{a\},a) \notin R_G^{ak}$ by minimality assumption). It remains to show that E defends itself in G^{ak}. By construction all arguments outside of $S \cup \{c\}$ are attacked, thus a is defended against each attack (T,a) with $T \cup (A_G \setminus S) \neq \emptyset$. Now assume there is $(T,a) \in R_G^{ak}$ such that $T \subset S$ (note that $T \neq S$ by assumption $(S,a) \notin R_G^{ak}$). In case $T = S \setminus \{a\}$, i.e., if $(S \setminus \{a\}, a)$ is active in G^{ak} then consider case (1). In case $T \subset S \setminus \{a\}$, $|T \cup \{a\}| < |S \cup \{a\}|$ and therefore $(T,a) \in R_F^{ak}$ by minimality assumption, contradiction to (S,a) being critical. Moreover, E is the unique preferred extension in $G^{ak} \cup H$. Note that $c \in T$ for all $T \in \mathit{pref}(G^{ak} \cup H)$ as c is not attacked in $G^{ak} \cup H$. All arguments beside c and a are either attacked by c or d and thus cannot be contained in a admissible set as they cannot be defended. Thus also $E \in \mathit{sem}(G^{ak} \cup H)$ and $E = \sigma(G^{ak} \cup H)$ for $\sigma \in \{\mathit{ideal}, \mathit{eager}\}$. Therefore $\sigma(F^{ak} \cup H) \neq \sigma(G^{ak} \cup H)$ for $\sigma \in \{\mathit{adm}, \mathit{pref}, \mathit{sem}, \mathit{ideal}, \mathit{eager}\}$.

3. Let (S,a) be inactive in F, we show that $\sigma(F^{ak} \cup H_2) \neq \sigma(G^{ak} \cup H_2)$. $E = \{a,c\}$ is not admissible in $F^{ak} \cup H_2$ as the argument a is not defended against S. On the other hand, E is admissible in $G^{ak} \cup H_2$. E is conflict-free: $(a,c),(c,a) \notin R_G^{ak}$ by construction. Now assume $(a,a) \in R_G^{ak}$. But then $|\{a\}| < |S \cup \{a\}|$ and therefore $(a,a) \in R_F^{ak}$, contradiction to $(S,a) \in R_F^{ak}$. We show that E defends itself: Towards a contradiction assume that a is not defended by E. Then there is an attack $(T,a) \in R_G^{ak}$ such that $T \subseteq S \cup \{a\}$ and, without loss of generality, T is minimal. In case $|T \cup \{a\}| < |S \cup \{a\}|$ we have that $(T,a) \in R_F^{ak}$ by minimality assumption. But then (S,a) is redundant (in case $a \notin T$) or tsa-inactive (in case $a \in T$) which contradicts $(S,a) \in R_F^{ak}$. In case $|T \cup \{a\}| = |S \cup \{a\}|$ either $T = S$ or $T = S \cup \{a\}$ holds. The former is in contradiction with the assumption $(S,a) \notin R_G^{ak}$, the latter contradicts G^{ak} being in normal form since in that case, (T,a) is reducible in G. Moreover, as argued in case (2), E is then also the unique preferred extension in $G^{ak} \cup H$. Thus also $E \in \mathit{sem}(G^{ak} \cup H)$ and $E = \sigma(G^{ak} \cup H)$ for $\sigma \in \{\mathit{ideal}, \mathit{eager}\}$. Therefore $\sigma(F^{ak} \cup H) \neq \sigma(G^{ak} \cup H)$ for $\sigma \in \{\mathit{adm}, \mathit{pref}, \mathit{sem}, \mathit{ideal}, \mathit{eager}\}$.

4. Let (S,a) be a tc self-attack in F, we show that $\sigma(F^{ak} \cup H_2) \neq \sigma(G^{ak} \cup H_2)$. $E = \{a,c\}$ is not admissible in $F^{ak} \cup H_2$ as the argument a is not defended against S. On the other hand, E is admissible in $G^{ak} \cup H_2$. We show that c defends a: Towards a contradiction assume that there is an attack $(T,a) \in R_G^{ak}$ such that $T \subseteq S$ and, without loss of generality, T is minimal. In case $|T \cup \{a\}| < |S \cup \{a\}|$ we have that $(T,a) \in R_F^{ak}$ by minimality assumption. But then (S,a) is redundant, contradiction to $(S,a) \in R_F^{ak}$. In case $|T \cup \{a\}| = |S \cup \{a\}|$ we have either $T = S$ or $T = S \setminus \{a\}$. The former contradicts the assumption $(S,a) \notin R_G^{ak}$. In case $T = S \setminus \{a\}$, we have (T,a) is active in G^{ak}: If (T,a) is inactive, there is an attack $(U,c) \in R_G^{ak}$, $U \subseteq T$, $c \in T$ and thus $(U,c) \in R_F^{ak}$ by minimality assumption, therefore (S,a) is reducible, contradiction to F^{ak} being in normal form. Thus we can consider case (1) with switched roles for F and G. Moreover, as in case (2) & (3), E is the unique preferred extension in $G \cup H$. Thus also $E \in \mathit{sem}(G \cup H)$ and $E = \sigma(G \cup H)$ for $\sigma \in \{\mathit{ideal}, \mathit{eager}\}$. Therefore $\sigma(F \cup H) \neq \sigma(G \cup H)$ for $\sigma \in \{\mathit{adm}, \mathit{pref}, \mathit{sem}, \mathit{ideal}, \mathit{eager}\}$.

In all cases, we found a witness H showing that $F^{ak} \not\equiv_{nd}^{\sigma} G^{ak}$ for $\sigma \in \{\mathit{adm}, \mathit{pref}, \mathit{sem}, \mathit{eager}, \mathit{ideal}\}$. Using $F \equiv_{nd}^{\sigma} F^{ak}$ and $G \equiv_{nd}^{\sigma} G^{ak}$ we conclude $F \not\equiv_{nd}^{\sigma} G$. □

The above theorem generalizes the corresponding characterization of strong equivalence for the admissible-based semantics in AFs [18]. We next relate the admissible kernel for SETAF to the admissible kernel of AFs. In [18], the admissible kernel of an AF $F = (A,R)$ has been defined as $F^{ak} = (A, R^{ak})$ with $R^{ak} = R \setminus \{(a,b) \mid a \neq$

$b, (a, a) \in R, \{(b, a), (b, b)\} \cap R \neq \emptyset\}$. Recall that the construction of the normal form of an AF removes the tsa-inactive attacks which are, in AFs, of the form (a, b) with $a \neq b$ where both a and b are self-attacking. The admissible kernel for SETAFs therefore indeed extends the admissible kernel for AFs since batc self-attacks do not exist in AFs (self-attacks (a, a) cannot contain additional conflicts) and ba-inactive attacks are of the form (a, b), $a \neq b$, where a is self-attacking and b attacks a.

4.4 Grounded Semantics

In this section we define the grounded kernel of a SETAF and show that all the considered variants of strong equivalence with respect to grounded semantics can be characterized via this kernel. To this end, recall that by Lemma 9, every tc self-attack (S, a) can be replaced with its proper version $(S \setminus \{a\}, a)$ without changing the grounded extension. Thus in order to ensure syntactical equivalence of the kernels, we define the grounded kernel in three steps: First, we compute the normal form; second, we replace every tc self-attack with its proper variant; and finally, we remove every grd-irrelevant attack.

Definition 23. For a SETAF $F = (A, R)$ in normal form, let $F' = (A, R')$ with

$$R' = (R \setminus \{(S, a) \mid (S, a) \text{ is a tc self-attack}\}) \cup \{(S \setminus \{a\}, a) \mid (S, a) \text{ is a tc self-attack}\},$$

i.e., we replace every tc self-attack (S, a) with its proper variant $(S \setminus \{a\}, a)$. We define the *grounded kernel* of F as $F^{gk} = (A, R^{gk})$ with

$$R^{gk} = R' \setminus \{(S, a) \in R \mid (S, a) \text{ is grd-irrelevant in } F\}.$$

The grounded kernel of an arbitrary SETAF F is the grounded kernel of the normal form of F.

Observe that the grounded kernel contains only critical self-attacks, since every reducible or tc self-attack (S, a) is transformed into its proper variant $(S \setminus \{a\}, a)$. The next lemma ensures that every redundancy is removed.

Lemma 15. *For every SETAF F, the grounded kernel F^{gk} does not contain redundant attacks.*

Proof. In case there is a tc self-attack (S, a) in F such that $(S \setminus \{a\}, a)$ is redundant in F', i.e., if there is an attack (S', a) with $S' \subset S \setminus \{a\}$, then (S, a) is redundant in F and gets deleted when computing F_{norm}. In case there is a tc self-attack (S, a) and an attack (S', a) with $S' \supseteq S \setminus \{a\}$ in F (i.e., (S', a) gets redundant in F'), then (S', a) is grd-irrelevant and thus gets removed in F^{gk}. □

We will prove that two SETAFs are strongly equivalent with respect to grounded semantics iff their grounded kernels coincide. To do so, we will first prove the following technical lemmas which show that both tc self-attacks as well as grd-irrelevant attacks can be iteratively processed without affecting the remaining tc self-attacks or grd-irrelevant attacks, respectively.

Lemma 16. *For a SETAF $F = (A, R)$ in normal form with two different tc self-attacks $(S, a), (S', b) \in R$, the attack (S, a) a tc self-attack in $G = (A, R \setminus \{(S', b)\} \cup \{(S' \setminus \{b\}, b))$.*

Proof. Let $R_G = R \setminus \{(S', b)\} \cup \{(S' \setminus \{b\}, b)$. As (S, a) is a tc self-attack in F there is an attack $(T, c) \in R$ with $T \subseteq S$, $a \in T$ and $c \in S \setminus \{a\}$. If $(T, c) \neq (S', b)$ then $(T, c) \in R_G$, thus (S, a) is a tc self-attack in G. In case $(T, c) = (S', b)$ there is $(T \setminus \{c\}, c) \in R_G$ and thus (S, a) remains a tc self-attack in G. □

Lemma 17. *For a SETAF $F = (A, R)$ with two different grd-irrelevant attacks (S, a), $(S', b) \in R$, the attack (S, a) is grd-irrelevant in $G = (A, R \setminus \{(S', b)\})$.*

Proof. Let $R_G = R \setminus \{(S', b)\}$. Since (S, a) is grd-irrelevant, there are attacks $(T \cup \{a\}, a)$, (T', c) with $T \subset S$, $T' \subseteq S \cup \{a\}$, $c \in S$. The attack $(T \cup \{a\}, a)$ is not proper and therefore not grd-irrelevant by definition. In case $(S', b) = (T', c)$, there is an attack $(U, d) \in R$ with $U \subseteq T' \cup \{c\} \subseteq S \cup \{a\}$, $d \in T'$. Then $(U, d) \in R_G$ and thus (S, a) is grd-irrelevant in G. □

We show that the grounded kernel F^{gk} and the SETAF F are strongly equivalent under grounded semantics.

Proposition 7. *For every SETAF F, $F \equiv_s^{grd} F^{gk}$.*

Proof. Consider an arbitrary SETAF H. By Proposition 5, $F \equiv_s^{grd} F_{norm}$, i.e., $grd(F \cup H) = grd(F_{norm} \cup H)$. By Lemma 13, every tc self-attack remains a tc self-attack in $F \cup H$, similarly, every grd-irrelevant attack remains grd-irrelevant in $F \cup H$. By Lemma 9 and Lemma 16, each tc self-attack (S, a) can be iteratively replaced by its proper variant $(S \setminus \{a\}, a)$ in the SETAF $F_{norm} \cup H$ without changing the grounded extension; moreover, by Lemma 6 and Lemma 17, grd-irrelevant attacks can be iteratively removed in the SETAF $F_{norm} \cup H$ without changing the grounded extension and thus we obtain $grd(F \cup H) = grd(F^{gk} \cup H)$ and therefore $F \equiv_s^{grd} F^{gk}$. □

We arrive at the desired result: two SETAFs are strongly equivalent with respect to grounded semantics exactly if their grounded kernels coincide.

Theorem 3. *For any two SETAFs F, G, the following are equivalent: (a) $F \equiv_s^{grd} G$; (b) $F \equiv_n^{grd} G$; (c) $F \equiv_{sd}^{grd} G$; (d) $F \equiv_{nd}^{grd} G$; (e) $F^{gk} = G^{gk}$.*

Proof. By definition, (a) implies (b) and (c). Likewise, (b) implies (d) and (c) implies (d). It remains to show (1) $F^{gk} = G^{gk}$ implies $F \equiv_s^{grd} G$ and (2) $F^{gk} \neq G^{gk}$ implies $F \not\equiv_{nd}^{grd} G$. We write $F^{gk} = (A_F^{gk}, R_F^{gk})$ and $G^{gk} = (A_G^{gk}, R_G^{gk})$.

(1) By Proposition 7, $F \equiv_s^{grd} F^{gk}$ and $G \equiv_s^{grd} G^{gk}$. Now as $F^{gk} = G^{gk}$, by the transitivity of $\equiv_s^{grd}$, we obtain $F \equiv_s^{grd} G$.

(2) Now, suppose that $F^{gk} \neq G^{gk}$. We show $F \not\equiv_{nd}^{grd} G$. In case $grd(F^{gk}) \neq grd(G^{gk})$ we are done: By Proposition 7, $grd(F^{gk}) = grd(F)$ and $grd(G^{gk}) = grd(G)$, i.e., $grd(F) \neq grd(G)$; consequently $F \not\equiv_{nd}^{grd} G$. Thus we assume $grd(F^{gk}) = grd(G^{gk})$.

First, we consider the case $A_F^{gk} \neq A_G^{gk}$. This implies $A_F \neq A_G$. Without loss of generality, let $a \in A_F \setminus A_G$ and let $S = grd(F)$. Notice that $a \notin S$ since $grd(F) = grd(G)$ by assumption. Let $H = (\{a\}, \emptyset)$. Then $a \in grd(G \cup H)$ since a is not attacked in $G \cup H$. On the other hand, we have $grd(F \cup H) = grd(F)$, thus $F \not\equiv_{nd}^{grd} G$.

Now, suppose $A_F^{gk} = A_G^{gk}$ (and therefore, $A_F = A_G$) and $R_F^{gk} \neq R_G^{gk}$. We will show $F^{gk} \not\equiv_{nd}^{grd} G^{gk}$ ($F \not\equiv_{nd}^{grd} G$ follows since $F \equiv_{nd}^{grd} F^{gk}$ and $G \equiv_{nd}^{grd} G^{gk}$). Without loss of generality, there exists $(S, a) \in R_F^{gk} \setminus R_G^{gk}$ such that there is no $(S', b) \in R_G^{gk} \setminus R_F^{gk}$ with $|S' \cup \{b\}| < |S \cup \{a\}|$ (otherwise exchange the roles of F and G). We distinguish the following cases: (1) (S, a) is active in F and $\nexists (S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$, $b \in S$, (2) (S, a) is not active in F or there exists $(S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$.

For fresh arguments c, d, we define

$$H_1 = (A_F \cup \{c, d\}, \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(a, d)\}),$$
$$H_2 = (A_F \cup \{c, d\}, \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(d, d), (d, b) \mid b \in S \setminus \{a\}\}).$$

1. Let (S, a) be active in F and there is no $(S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$ and $b \in S$. In this case, we consider the AF H_1 and show that $grd(F^{gk} \cup H_1) \neq grd(G^{gk} \cup H_1)$. Then, in both F^{gk} and G^{gk}, (i) S is conflict-free and (ii) c defends S. Notice that (i) is immediate by definition of active attacks and by minimality of (S, a), i.e., there are no smaller attacks in both F^{gk} and G^{gk}. Furthermore, (ii) follows since c attacks every argument in $A_F \setminus (S \cup \{a\})$ and since F^{gk} contains no redundant attacks and, moreover, there is no attack $(S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$ and $b \in S$ by assumption. Thus the grounded extension in F^{gk} is given by $S \cup \{c, d\}$ since S defends d against a in F^{gk}. On the other side. an attack (S', a) with $S' \subset S$ in G^{gk} would be in contradiction to the minimality assumption on (S, a) and thus $grd(G^{gk}) = S \cup \{c\}$ since d is not defended against a in G^{gk}.

2. We consider now the case where either (S, a) is not active in F (i.e., (S, a) is either critical or inactive) or (S, a) is active in F and there exists $(S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$. We will first show that in this case, (i) there exists no attack $(S'', a) \neq (S, a)$, $S'' \subseteq S \cup \{a\}$ in both F^{gk} and G^{gk}. In case (S, a) is critical in F, the statement is immediate by definition and by minimality of (S, a). In case (S, a) is inactive, the statement holds for F^{gk}, otherwise, (S, a) is either redundant (if $S'' \subseteq S$) or grd-irrelevant (if $a \in S''$), contradiction to the construction of the grounded kernel. For G^{gk}, the statement follows since in case $S'' \subset S \cup \{a\}$, the existence of such an attack contradicts the minimality of (S, a) and in case $S'' = S \cup \{a\}$, the attack is reducible, contradiction to the

construction of G^{gk}. In case (S,a) is active and $\exists(S' \cup \{a\}, b) \in R_F^{gk}$ with $S' \subset S$, $b \in S$, the statement holds for F^{gk}, otherwise, (S,a) is either redundant (if $S'' \subseteq S$) or grd-irrelevant (if $a \in S''$), contradiction to the construction of the grounded kernel. For G^{gk}, the statement follows since in case $S'' \subset S \cup \{a\}$, the existence of such an attack contradicts the minimality of (S,a) and in case $S'' = S \cup \{a\}$, the attack is a tc self-attack, contradiction to the construction of G^{gk}.

We will show that $grd(F^{gk} \cup H_2) \neq grd(G^{gk} \cup H_2)$. Observe that c defends a in G^{gk} against every attack in R_G^{gk} since c attacks every argument $b \in A_F \setminus (S \cup \{a\})$ and since a is not attacked by any $S'' \subseteq S \cup \{a\}$. It follows that $grd(G^{gk}) = \{c, a\}$. On the other hand, c does not defend a in F^{gk} since c does not attack S. Therefore $grd(F^{gk}) = \{a\} \neq \{c, a\} = grd(G^{gk})$.

In both cases, we found a witness for $F \not\equiv_{nd}^{grd} G$. □

The above theorem generalizes the corresponding results for AFs [18]. For AFs, the replacement of tc self-attacks can be omitted since tc self-attacks do not exist in AFs. It follows that for AFs, the following attacks are removed: (a) every attack (a,b) with $a \neq b$ and $(a,a),(b,b) \in R$ (by computing the normal form) and (b) every attack (a,b) with $a \neq b$, $(b,b) \in R$ and $(b,a) \in R$ (by removing grd-irrelevant attacks). This matches the definition of the grounded kernel for AFs in [18] and shows that the grounded kernel for SETAFs extends the grounded kernel for AFs.

4.5 Stable and Stage Semantics

Towards the definition of a stable kernel first notice that for stable and stage semantics, the direction of self-attacks, i.e., the particular argument which is attacked in a self-attacking set does not affect the extensions: A self-attacking set S will be never part of a stable and stage extension since it is conflicting, thus the attack does not contribute to the range. In order to take this into account we consider a weakening of syntactical equivalence where two SETAFs are equivalent if they coincide at their proper attacks and possess the same self-attacking sets.

Definition 24. Two SETAFs $F = (A_F, R_F)$ and $G = (A_G, R_G)$ are equal up to the direction of self-attacks, in symbols $F =_{sa} G$, if $A_F = A_G$ and

$$\begin{aligned} \{(S,a) \mid (S,a) \text{ proper in } F\} &= \{(S,a) \mid (S,a) \text{ proper in } G\}, \text{ and} \\ \{S \cup \{a\} \mid (S,a) \text{ self-attack in } F\} &= \{S \cup \{a\} \mid (S,a) \text{ self-attack in } G\}. \end{aligned}$$

We can now formalize the above observation that the direction of self-attacks is not significant for stable and stage semantics, i.e., for two SETAFs F and G with $F =_{sa} G$, $\sigma(F) = \sigma(G)$ for $\sigma \in \{stb, stage\}$. In fact, it even holds that $F \equiv_s^\sigma G$ for $\sigma \in \{stb, stage\}$ as we show next.

Lemma 18. *For any two SETAFs F, G, if $F =_{sa} G$ then $F \equiv_s^\sigma G$ for $\sigma \in \{stb, stage\}$.*

Proof. Let $\sigma \in \{stb, stage\}$ and consider an arbitrary SETAF H. We first show that $cf(F \cup H) = cf(G \cup H)$: Without loss of generality, let $T \in cf(F \cup H)$ and assume $T \notin cf(G \cup H)$, then there is an attack $(S,a) \in R_G \cup R_H$ such that $S \cup \{a\} \subseteq T$. In case (S,a) is proper, then $(S,a) \in R_F \cup R_H$, contradiction to the conflict-freeness of T, in case (S,a) is a self-attack, then there is a self-attack $(S', a') \in R_F \cup R_H$ such that $S' \cup \{a'\} = S \cup \{a\} \subseteq T$, contradiction to T being conflict-free. Moreover, $T^+_{F \cup H} = T^+_{G \cup H}$ for every $T \in cf(F \cup H)$ since T does not contain self-attacks and $F \cup H$, $G \cup H$ agree on their proper attacks. □

Next we define the stable kernel F^{sk} of a SETAF F which consists of all active and critical attacks in F.

Definition 25. For a SETAF $F = (A,R)$ in minimal form[2], we define the *stable kernel* of F as $F^{sk} = (A, R^{sk})$ with

$$R^{sk} = \{(S,a) \in R \mid (S,a) \text{ is active or critical in } F\}.$$

The stable kernel of an arbitrary SETAF F is the stable kernel of the minimal form of F.

[2]Notice, that one can equivalently use the normal form instead of the minimal form as basis of the stable kernel as the normal form does not remove active or critical attacks.

In the remaining part of the section we will show that two SETAFs F and G are strongly equivalent with respect to stable and stage semantics if and only if F^{sk} and G^{sk} are equal up to the direction of self-attacks. To do so, we will first prove the following technical lemma.

Lemma 19. *Let $F = (A, R)$ in minimal form with two different attacks $(S, a), (S', b) \in R$ which are inactive, reducible or tc self-attacks. Then (S, a) remains inactive, reducible or a tc self-attack in $G = (A, R \setminus \{(S', b)\})$.*

Proof. Let $R_G = R \setminus \{(S', b)\}$. First consider the case (S, a) inactive in F. Since (S, a) is inactive there is an attack (T, c) in F with $T \subseteq S$, $c \in S$, moreover, $c \neq a$ since F is in minimal form. If $(S', b) \neq (T, c)$ then we are done, thus we assume $(S', b) = (T, c)$. In case (S', b) is inactive, there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \subseteq S$, $b' \in S'$; if (S', b) is reducible, there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \setminus \{b\} \subseteq S$, $b' \in S'$; in case (S', b) is a tc self-attack, there is an attack $(S'', b') \in R$ with $S'' \subset S' \subseteq S$, $b \in S''$, $b' \in S' \setminus \{b\}$. In every case S remains conflicting in G and thus (S, a) is inactive in G.

Now, let (S, a) be reducible in F, that is, there is an attack (T, c) in F with $T \subseteq S \setminus \{a\}$, $c \in S$, $c \neq a$. We assume $(S', b) = (T, c)$. Regardless whether (S', b) is inactive, reducible or a tc self-attack, there exists an attack $(S'', b') \in R$ with $S'' \subseteq S' \subseteq S \setminus \{a\}$ and $b \in S'$ and thus (S, a) is reducible in G.

In case (S, a) is a tc self-attack in F, there is an attack (T, c) in F with $T \subset S$, $a \in T$, $c \in S \setminus \{a\}$. If $(S', b) \neq (T, c)$ then we are done, thus we assume $(S', b) = (T, c)$. In case (S', b) is inactive, there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \subset S$, $b' \in S'$; if (S', b) is reducible, there is an attack $(S'', b') \in R$ with $S'' \subseteq S' \setminus \{b\} \subset S$, $b' \in S'$; in case (S', b) is a tc self-attack, there is an attack $(S'', b') \in R$ with $S'' \subset S' \subset S$, $b \in S''$, $b' \in S' \setminus \{b\}$. It follows that (S, a) is either a tc self-attack in G if $a \in S''$ or reducible in G if $a \notin S''$. $\square$

Next we show that each SETAF F is strongly equivalent to its stable kernel F^{sk} with respect to stable and stage semantics.

Proposition 8. *For every SETAF F and its stable kernel F^{sk}, $F \equiv^{\sigma}_{s} F^{sk}$ for $\sigma \in \{stb, stage\}$.*

Proof. Let $\sigma \in \{stb, stage\}$ and consider an arbitrary SETAF H. By Proposition 4, $F \equiv^{\sigma}_{s} F_{min}$, i.e., $\sigma(F \cup H) = \sigma(F_{min} \cup H)$. By Lemma 13, every inactive attack, every reducible attack and every tc self-attack remains inactive, reducible and a tc self-attack, respectively, in $F \cup H$. As a consequence of Lemma 7, Lemma 11, and Lemma 3, a reducible attack (S, a) can be removed without changing the outcome under stable and stage extensions (by first transforming it to its inactive pendant $(S \setminus \{a\}, a)$ which can be removed under stable and stage semantics). Let B denote the set of inactive, reducible attacks, and tc self-attacks in F_{min}. By Lemma 3, Lemma 8, Lemma 19, and the observation above, it follows that every attack in B can be iteratively removed in the SETAF $F_{min} \cup H$ without changing the σ-extensions. Thus we obtain $\sigma(F \cup H) = \sigma(F^{sk} \cup H)$ and therefore $F \equiv^{\sigma}_{s} F^{sk}$. $\square$

The following theorem states that every two SETAFs F and G are strongly equivalent with respect to stable and stage semantics exactly if their kernels are equal up to the direction of self-attacks. Observe that in contrast to the strong equivalence characterizations of the semantics we discussed so far in the previous chapters, we do not assume syntactical equivalence of the kernels.

Theorem 4. *For any two SETAFs F, G and for $\sigma \in \{stb, stage\}$, the following are equivalent: (a) $F \equiv^{\sigma}_{s} G$; (b) $F \equiv^{\sigma}_{n} G$; (c) $F \equiv^{\sigma}_{sd} G$; (d) $F \equiv^{\sigma}_{nd} G$; (e) $F^{sk} =_{sa} G^{sk}$.*

Proof. By definition, (a) implies (b) and (c). Likewise, (b) implies (d) and (c) implies (d). It remains to show (1) $F^{sk} =_{sa} G^{sk}$ implies $F \equiv^{\sigma}_{s} G$ and (2) $F^{sk} \neq_{sa} G^{sk}$ implies $F \not\equiv^{\sigma}_{nd} G$. We write $F^{sk} = (A^{sk}_F, R^{sk}_F)$ and $G^{sk} = (A^{sk}_G, R^{sk}_G)$.

(1) By Proposition 8, $F \equiv^{\sigma}_{s} F^{sk}$ and $G \equiv^{\sigma}_{s} G^{sk}$. Now as $F^{sk} =_{sa} G^{sk}$ by Lemma 18 we have $F^{sk} \equiv^{\sigma}_{s} G^{sk}$. By the transitivity of $\equiv^{\sigma}_{s}$, we obtain $F \equiv^{\sigma}_{s} G$.

(2) First, we consider the case $A^{sk}_F \neq A^{sk}_G$. This implies $A_F \neq A_G$. Without loss of generality, let $a \in A_F \setminus A_G$. We use $B = (A_F \cup A_G) \setminus \{a\}$, and c as a fresh argument. Consider $H = \big(B \cup \{c\}, \{(c, b) \mid b \in B\}\big)$. Note that H is conform with the definition of $\equiv^{\sigma}_{nd}$, i.e., it is a simple AF not changing the relation between existing arguments. Suppose now, a is contained in some $S \in \sigma(F \cup H)$. Then, we are done since a cannot be contained in any $S' \in \sigma(G \cup H)$, since $a \notin A(G \cup H)$. Otherwise, we extend H to $H' = H \cup (\{a\}, \emptyset)$. Then, $\{a, c\}$ is the unique

stable extension (and thus unique stage extension) of $G \cup H'$. On the other hand, observe that $F \cup H' = F \cup H$, hence by assumption, a is not contained in any $S \in \sigma(F \cup H')$. In both cases, we get $F \not\equiv^\sigma_{nd} G$. Now suppose $A^{sk}_F = A^{sk}_G$. Without loss of generality, we assume that there is either (1) a proper attack $(S, a) \in R^{sk}_F \setminus R^{sk}_G$ or (2) a critical attack $(S, a) \in R^{sk}_F \setminus R^{sk}_G$ such that there is no attack $(S, b) \in R^{sk}_G$ with $b \in S$. We show that in both cases, $F^{sk} \not\equiv^\sigma_{nd} G^{sk}$.

1. Without loss of generality, we can assume that there is no $(S', a) \in R^{sk}_G$ with $S' \subset S$ (otherwise we exchange the roles of F and G as $(S', a) \notin R^{sk}_F$). For fresh arguments c, t, we define $H = (A_F \cup \{c, t\}, R_H)$ with

$$R_H = \{(t, c), (c, t)\} \cup \{(c, b) \mid b \in A_F \setminus (S \cup \{a\})\} \cup \{(t, b) \mid b \in A_F\}.$$

   First, by construction we have that $\{t\} \in \mathit{stb}(F^{sk} \cup H)$ and $\{t\} \in \mathit{stb}(G^{sk} \cup H)$ and thus stable and stage semantics coincide in both $F^{sk} \cup H$ and $G^{sk} \cup H$. Thus we can restrict ourselves to stable semantics. We have $S \cup \{c\} \in \mathit{stb}(F^{sk} \cup H)$, since $S \cup \{c\}$ is conflict-free and attacks all arguments $b \notin S$ either collectively via S or via the newly introduced argument c. However, $S \cup \{c\} \notin \mathit{stb}(G^{sk} \cup H)$ as by the assumption there is no $(S', a) \in R^{sk}_G$ with $S' \subseteq S$ and thus $S \cup \{c\}$ does not attack a.

2. Without loss of generality, we can assume that there is no $(S', b) \in R^{sk}_G$ with $S' \cup \{b\} \subseteq S$ (otherwise we exchange the roles of F and G as $(S', d) \notin R^{sk}_F$ for every $d \in S$). For a fresh argument c, we define

$$H = (A_F \cup \{c\}, \{(c, b) \mid b \in A_F \setminus S\}).$$

   We have $S \cup \{c\} \notin \sigma(F^{sk} \cup H)$ for $\sigma \in \{\mathit{stb}, \mathit{stage}\}$ since (S, a) is a conflict within the set $S \cup \{c\}$. However, for $G^{sk} \cup H$ we have that $S \cup \{c\}$ is conflict-free and attacks all argument outside the set, i.e., $S \cup \{c\} \in \sigma(G^{sk} \cup H)$.

In all cases, we found a witness H showing that $F^{sk} \not\equiv^\sigma_{nd} G^{sk}$ for $\sigma \in \{\mathit{stb}, \mathit{stage}\}$. Using $F \equiv^\sigma_{nd} F^{sk}$ and $G \equiv^\sigma_{nd} G^{sk}$ we conclude $F \not\equiv^\sigma_{nd} G$. □

The above result generalizes the corresponding result for AFs in [18]. For AFs, inactive attacks are of the form (a, b), $a \neq b$, where a is self-attacking; moreover, every self-attack is of the form (a, a) and is therefore critical by definition. Thus the presented definition extends the stable kernel for AFs which has been defined in [18]. Here, the stable kernel of an AF $F = (A, R)$ is defined as $F^{sk} = (A, R^{sk})$ with $R^{sk} = R \setminus \{(a, b) \mid a \neq b, (a, a) \in R\}$. Furthermore notice that for AFs, syntactical equivalence is ensured since self-attacking sets are singletons.

Remark 3. The presented characterization of strong equivalence with respect to stable and stage semantics differs from the preliminary version in [13] with respect to the handling of self-attacks. In [13], we define the stable kernel by using additional self-attacks, i.e., for every self-attack (S, a) we add (S, b) for every $b \in S$. In this way we ensure that two SETAFs F and G are strongly equivalent with respect to stable and stage semantics if and only if their stable kernels are syntactically equivalent. The present characterization weakens syntactical equivalence in order to avoid the potential blow-up of the kernel; moreover, avoiding additional attacks also allows for comparing the kernels in terms of a subset-relation (cf. Section 4.7).

4.6 Conflict-Free and Naive Semantics

Towards the definition of a kernel for conflict-free semantics we need to take into account that the direction of attacks is not significant for conflict-free semantics. We thus further relax the equality relation from Definition 24 in order to compare kernels we shall use for conflict-free and naive semantics.

Definition 26. Two SETAFs $F = (A_F, R_F)$ and $G = (A_G, R_G)$ are equal up to the direction of attacks, in symbols $F =_a G$, if $A_F = A_G$ and $\{S \cup \{a\} \mid (S, a) \in R_F\} = \{S \cup \{a\} \mid (S, a) \in R_G\}$.

We have that SETAFs that are equal up to the direction of attacks are strongly equivalent with respect to conflict-free and naive semantics.

Lemma 20. *For any two SETAFs F, G, if $F =_a G$ then $F \equiv^\sigma_s G$ for $\sigma \in \{\mathit{cf}, \mathit{naive}\}$.*

Proof. Let $\sigma \in \{cf, naive\}$ and consider an arbitrary SETAF H. Let $T \in cf(F \cup H)$ and assume $T \notin cf(G \cup H)$. Then there is an attack $(S, a) \in R(G \cup H)$ such that $S \cup \{a\} \subseteq T$. Since $F =_a G$ there is an attack $(S', a') \in R(F \cup H)$ such that $S' \cup \{a'\} = S \cup \{a\} \subseteq T$, contradiction to T being conflict-free. □

We next introduce the conflict-free kernel for SETAFs.

Definition 27. For a SETAF $F = (A, R)$, we define the *conflict-free kernel* of F as $F^{cfk} = (A, R^{cfk})$ with

$$R^{cfk} = R \setminus \{(S, a) \mid (S, a) \text{ cf-redundant in } F\}.$$

Notice that $R^{cfk} \subseteq R^{sk}$ for every SETAF F, that is, the conflict-free kernel contains a subset of active and critical attacks. Moreover, R^{cfk} does not contain redundant attacks since every redundant attack is cf-redundant.

We are now ready to prove our final characterization result: two SETAFs F and G are strongly equivalent with respect to conflict-free and naive semantics if and only if their conflict-free kernels are equal up to the direction of attacks. To do so, we first prove the following technical lemma.

Lemma 21. *Let $F = (A, R)$ be a SETAF, then for every $(S, a) \in R$, either $(S, a) \in R^{cfk}$ or there is an attack $(T, b) \in R$, $T \cup \{b\} \subset S \cup \{a\}$ such that $(T, b) \in R^{cfk}$.*

Proof. Towards a contradiction let $(S, a) \in R$ be an attack violating the condition of the lemma, i.e., (a) $(S, a) \notin R^{cfk}$ and (b) for every attack $(T, b) \in R$ with $T \cup \{b\} \subset S \cup \{a\}$ we have $(T, b) \notin R^{cfk}$; moreover, let (S, a) be a minimal such attack, i.e., every attack $(T, b) \in R$ with $|S \cup \{a\}| > |T \cup \{b\}|$ satisfies the condition. Observe that (S, a) is cf-redundant otherwise $(S, a) \in R^{cfk}$ by definition. By cf-redundancy of (S, a), there is an attack $(T, b) \in R$ with $T \cup \{b\} \subset S \cup \{a\}$, that is $|T \cup \{b\}| < |S \cup \{a\}|$. By assumption (T, b) satisfies the condition of the lemma, therefore either $(T, b) \in R^{cfk}$ or there is an attack (T', b'), $T' \cup \{b'\} \subset T \cup \{b\} \subset S \cup \{a\}$ such that $(T', b') \in R^{cfk}$, contradiction to the assumption that (S, a) does not satisfy (a) and (b). □

We next show that each SETAF F is strongly equivalent to its conflict-free kernel with respect to conflict-free and naive semantics.

Proposition 9. *For every SETAF F and its conflict-free kernel F^{cfk}, $F \equiv^\sigma_s F^{cfk}$ for $\sigma \in \{cf, naive\}$.*

Proof. Let $\sigma \in \{cf, naive\}$ and consider an arbitrary SETAF H. Since $R(F^{cfk} \cup H) \subseteq R(F \cup H)$ we have $cf(F \cup H) \subseteq cf(F^{ck} \cup H)$. We show $cf(F^{cfk} \cup H) \subseteq cf(F \cup H)$: Let $T \in cf(F^{cfk} \cup H)$ and assume $T \notin cf(F \cup H)$. Then there is an attack $(S, a) \in R(F \cup H)$ such that $S \cup \{a\} \subseteq T$. In case $(S, a) \in R(H)$ we are done, thus we assume that $(S, a) \in R(F)$. Then by Lemma 21, either $(S, a) \in R^{cfk}$ or there exists an attack $(S', b) \in R$, $S' \cup \{b\} \subset S \cup \{a\}$ such that $(S', b) \in R^{cfk}$. In both cases, the set $S \cup \{a\} \subseteq T$ is conflicting in F^{cfk}, contradiction. Thus we obtain $\sigma(F \cup H) = \sigma(F^{cfk} \cup H)$ and therefore $F \equiv^\sigma_s F^{cfk}$. □

We will prove that strong equivalence with respect to conflict-free and naive semantics can be characterized by the conflict-free kernel.

Theorem 5. *For any two SETAFs F, G and for $\sigma \in \{cf, naive\}$, the following are equivalent: (a) $F \equiv^\sigma_s G$; (b) $F \equiv^\sigma_n G$; (c) $F \equiv^\sigma_{sd} G$; (d) $F \equiv^\sigma_{nd} G$; (e) $F^{cfk} =_a G^{cfk}$.*

Proof. Let $\sigma \in \{cf, naive\}$. By definition (a) implies (b) and (c). Likewise, (b) implies (d) and (c) implies (d). It remains to show (1) $F^{cfk} =_a G^{cfk}$ implies $F \equiv^\sigma_s G$ and (2) $F^{cfk} \neq_a G^{cfk}$ implies $F \not\equiv^\sigma_{nd} G$.

(1) By Proposition 9, $F \equiv^\sigma_s F^{cfk}$ and $G \equiv^\sigma_s G^{cfk}$. By $F^{cfk} =_a G^{cfk}$ and Lemma 20 we obtain $F^{cfk} \equiv^\sigma_s G^{cfk}$. By the transitivity of $\equiv^\sigma_s$ we can conclude $F \equiv^\sigma_s G$.

(2) Now, suppose that $F^{cfk} \neq G^{cfk}$. First, we consider the case $A^{cfk}_F \neq A^{cfk}_G$. This implies $A_F \neq A_G$. Without loss of generality, let $a \in A_F \setminus A_G$. We use $B = (A_F \cup A_G) \setminus \{a\}$, and c as a fresh argument. Consider $H = \big(B \cup \{c\}, \{(c, b) \mid b \in B\}\big)$. Note that H is conform with the definition of $\equiv^\sigma_{nd}$, i.e., it is a simple AF not changing the relation between existing arguments. Suppose now, a is contained in some $S \in \sigma(F \cup H)$. Then, we are done since a cannot be contained in any $S' \in \sigma(G \cup H)$, since $a \notin A(G \cup H)$. Otherwise, we extend H to $H' = H \cup (\{a\}, \emptyset)$. Then, $\{a, c\}$ is conflict-free in $G \cup H'$, moreover, it is a maximal conflict-free set (thus a naive extension) because c

$$
\begin{array}{ccccccc}
 & \subsetneq & F^{sk} & \subseteq & F^{ak} & \subseteq & F^{ck} \\
F^{\mathit{cfk}} & & & & & & \\
 & \subseteq & F^{gk} & & & &
\end{array}
$$

Figure 16: Relations between kernels (for every SETAF F).

attacks every $b \in (A_F \cup A_G) \setminus \{a\}$. On the other hand, observe that $F \cup H' = F \cup H$, hence by assumption, a is not contained in any $S \in \sigma(F \cup H')$. In both cases, we get $F \not\equiv^\sigma_{nd} G$.

Now suppose $A^{\mathit{cfk}}_F = A^{\mathit{cfk}}_G$. Without loss of generality, we assume that there is an attack $(S, a) \in R^{\mathit{cfk}}_F \setminus R^{\mathit{cfk}}_G$ such that there is no attack $(S', b) \in R^{sk}_G$ with $S' \cup \{b\} = S \cup \{a\}$, and moreover, (S, a) is a minimal such attack, i.e., there is no $(T, c) \in R^{\mathit{cfk}}_G$ with $T \cup \{c\} \subset S \cup \{a\}$ (otherwise exchange the roles of F and G). Then the set $S \cup \{a\}$ is not conflict-free in F^{cfk} but is conflict-free in G^{cfk} by the minimality and by the definition of the conflict-free kernel, i.e., there is no attack $(S', b) \in R_F$ such that $S' \cup \{b\} \subseteq S \cup \{a\}$. It follows that $F^{\mathit{cfk}} \not\equiv^\sigma_k G^{\mathit{cfk}}$ for $k \in \{s, n, sd, nd\}$ and $\sigma \in \{\mathit{cf}, \mathit{naive}\}$. Using $F \equiv^\sigma_{nd} F^{\mathit{cfk}}$ and $G \equiv^\sigma_{nd} G^{\mathit{cfk}}$ we conclude $F \not\equiv^\sigma_{nd} G$. □

Strong equivalence for AFs with respect to naive semantics has been characterized in [15] without explicitly defining a corresponding kernel. Instead it has been shown that two AFs F, G are strongly equivalent with respect to naive semantics if and only if (a) $\mathit{naive}(F) = \mathit{naive}(G)$ (or, equivalently, $\mathit{cf}(F) = \mathit{cf}(G)$) and (b) the arguments in the compared AFs coincide. Since for conflict-free and naive semantics, the direction of the attacks is not of importance, a corresponding kernel could be defined by either dropping syntactical equivalence or by adding additional attacks such that the resulting framework is symmetric. Moreover, attacks of the form (a, b) with either a or b being self-attacking can be removed since the latter already ensures that the set $\{a, b\}$ is not conflict-free. Observe that such attacks match the definition of being cf-redundant attacks restricted to the AF case. In [15], such a kernel construction has been discussed in the context of strong equivalence with respect to naive semantics restricted to weakly symmetric AFs.

4.7 Relations between Kernels

In the present section, we study the relations between the introduced kernels and the strong equivalence relations of different semantics.

Proposition 10. *Let F be a SETAF. The relations between the kernels F^{cfk}, F^{sk}, F^{ak}, F^{gk} and F^{ck} depicted in Figure 16 hold.*

Proof. The proof uses results on the relations between attacks from Section 3.3. $F^{\mathit{cfk}} \subseteq F^{sk}$ because every attack $(S, a) \in R^{\mathit{cfk}}$ is active or critical; moreover, $F^{sk} \subseteq F^{ak}$ because F^{ak} additionally contains a subset of inactive and tc self-attacks; similarly, $F^{ak} \subseteq F^{ck}$ because F^{ak} strengthens F^{ck} by additionally removing ba-inactive attacks and batc self-attacks. We show that $F^{\mathit{cfk}} \subseteq F^{gk}$: First of all, every grd-irrelevant attack is cf-redundant and is thus removed in F^{cfk} due to redundancy. Moreover, every tc self-attack is cf-redundant, thus the transformation of a tc self-attack (S, a) to the proper variant $(S \setminus \{a\}, a)$ does not affect attacks in the conflict-free kernel. We provide counter examples for the remaining relations:

- $F^\alpha \not\subseteq F^{\mathit{cfk}}$ for $\alpha \in \{sk, ak, ck, gk\}$: Let $F = (\{a, b, c\}, \{(\{a, b\}, c), (\{c\}, a)\})$. Then $F^{sk} = F^{ak} = F^{ck} = F^{gk} = F$ but $F^{\mathit{cfk}} = (\{a, b, c\}, \{(\{c\}, a)\})$.
- $F^\alpha \not\subseteq F^{sk}$ for $\alpha \in \{ak, ck, gk\}$; $F^{gk} \not\subseteq F^\beta$ for $\beta \in \{ak, ck\}$; $F^\gamma \not\subseteq F^{gk}$ for $\gamma \in \{ak, ck, sk\}$: Let $F = (\{a, b\}, \{(\{a, b\}, a), (\{a\}, b)\}$, then $(\{a, b\}, a)$ is a tc self-attack and $F^{ak} = F^{ck} = F$, $F^{sk} = (\{a, b\}, \{(\{a\}, b)\}$ and $F^{gk} = (\{a, b\}, \{(\{b\}, a), (\{a\}, b)\}$.
- $F^{ck} \not\subseteq F^{ak}$: Consider $F = (\{a, b\}, \{(\{b\}, a), (\{a\}, b), (\{b\}, b)\})$, then the attack $(\{b\}, a)$ is ba-inactive and thus $F^{ak} = (\{a, b\}, \{(\{a\}, b), (\{b\}, b)\})$, but $F^{ck} = F$. □

However, concerning the actual relation between the strong equivalence relations with respect to the different semantics more can be said. In fact, we will show that equality of particular kernels implies equality of other forms of kernels. Ultimately, we end up with exactly the same picture that is known for strong equivalence between AFs. We start with one more technical lemma.

Lemma 22. *Let $F = (A, R)$ be a SETAF and let $(S, a) \in R$ be inactive, redundant, reducible or tc self-attacking. Then there exists an attack $(S', b) \in R$ with $S' \subseteq S$ and $b \in S$ that is active or critical in F.*

Proof. Towards a contradiction let $(S, a) \in R$ be an attack violating the condition of the lemma such that all attacks $(T, b) \in R$ with $|S \cup \{a\}| > |T \cup \{b\}|$ satisfy the condition, i.e., (S, a) is minimal in this respect. We distinguish the following cases: (1) (S, a) is inactive, redundant or tc self-attacking; (2) (S, a) is reducible.

1. Let (S, a) be inactive, redundant or tc self-attacking. Then there is an attack (S', b), $S' \cup \{b\} \subseteq S \cup \{a\}$ with $|S' \cup \{b\}| < |S \cup \{a\}|$: In case (S, a) is inactive, there exists $S' \subseteq S$, $b \in S$ such that $(S', b) \in R$ and $|S' \cup \{b\}| < |S \cup \{a\}|$; in case (S, a) is redundant, there exists $S' \subset S$, $(S', a) \in R$ and $|S' \cup \{a\}| < |S \cup \{a\}|$; if (S, a) is a tc self-attack, there exists $S' \subset S$, $a \in S'$, $b \in S \setminus \{a\}$ such that $(S', b) \in R$ and $|S' \cup \{b\}| < |S \cup \{a\}|$. Since (S, a) is a minimal attack violating the condition of the lemma, we obtain that there is an active or critical attack $(S'', c) \in R$ with $S'' \subseteq S'$ and $c \in S'$. It follows that there is an active or critical attack (S'', c) with $S'' \subseteq S$, $c \in S$, contradiction to our initial assumption.

2. Let (S, a) be reducible. Then there exists $S' \subseteq S \setminus \{a\}$, $b \in S$ such that $(S', b) \in R$. In case $|S' \cup \{b\}| < |S \cup \{a\}|$ we obtain a contradiction to our initial assumption since (S, a) is a minimal attack violating the condition of the lemma (cf. case (1)). In case $|S' \cup \{b\}| = |S \cup \{a\}|$ we have $S' = S \setminus \{b\}$, $b \in S$. By assumption (S', b) is neither active nor critical, that is, there is an inactive, redundant, reducible, or tc self-attacking $(S'', c) \in R$ such that $S'' \subseteq S'$ and $c \in S'$. But then $|S'' \cup \{c\}| < |S \cup \{a\}|$ and, by the minimality of (S, a), we obtain that there is an active or critical attack $(T, d) \in R$ with $T \subseteq S''$ and $d \in S''$. Since $S'' \subseteq S$, we obtain that there is an active or critical attack (T, d) with $T \subseteq S''$ and $d \in S''$, contradiction to our initial assumption. □

We next state our results about when a certain equality between kernels implies the equality of another kernel.

Proposition 11. *For any two SETAFs F and G,*

(a) $F^{ck} = G^{ck}$ implies $F^\beta = G^\beta$ for $\beta \in \{ak, sk, cfk, gk\}$;

(b) $F^{ak} = G^{ak}$ implies $F^\beta = G^\beta$ for $\beta \in \{sk, cfk\}$;

(c) $F^\alpha = G^\alpha$ implies $F^{cfk} = G^{cfk}$ for $\alpha \in \{sk, gk\}$;

(d) $F^{sk} =_{sa} G^{sk}$ implies $F^{cfk} =_a G^{cfk}$.

Proof. Consider two SETAFs F and G. We write $F^\kappa = (A_F, R_F^\kappa)$, $G^\kappa = (A_G, R_G^\kappa)$ for $\kappa \in \{ak, sk, cfk, gk, ck\}$.

To show (d) assume $F^{sk} =_{sa} G^{sk}$ and let $(S, a) \in R_F^{cfk}$. We show that there is $(S', a') \in G_F^{cfk}$ with $S' \cup \{a'\} = S \cup \{a\}$. In case (S, a) is proper in F^{sk} we have that $(S, a) \in G^{sk}$. By Lemma 21, either $(S, a) \in R_G^{cfk}$ or there is an attack $(T, b) \in R_G$, $T \cup \{b\} \subset S \cup \{a\}$ such that $(T, b) \in R_G^{cfk}$. If the former case holds, then we are done, so we assume that there is an attack $(T, b) \in R_G^{cfk}$ with $T \cup \{b\} \subset S \cup \{a\}$. Thus $(T, b) \in R_G^{sk}$ since $G^{cfk} \subseteq G^{sk}$ and therefore there is $(T', b') \in R_F^{sk}$ with $T' \cup \{b'\} = T \cup \{b\}$ since $F^{sk} =_{sa} G^{sk}$, contradiction to $(S, a) \in R_F^{cfk}$. Thus $(S, a) \in R_G^{cfk}$ holds. In case (S, a) is a self-attack in F^{sk} we have that there is $(S', a') \in G^{sk}$ with $S' \cup \{a'\} = S \cup \{a\}$. By Lemma 21, either $(S', a') \in R_G^{cfk}$ or there is an attack $(T, b) \in R_G$, $T \cup \{b\} \subset S' \cup \{a'\}$ such that $(T, b) \in R_G^{cfk}$. If the former case holds, then we are done, so we assume that there is an attack $(T, b) \in R_G^{cfk}$ with $T \cup \{b\} \subset S' \cup \{a'\}$. Thus $(T, b) \in R_G^{sk}$ since $G^{cfk} \subseteq G^{sk}$ and therefore there is $(T', b') \in R_F^{sk}$ with $T' \cup \{b'\} = T \cup \{b\}$ since $F^{sk} =_{sa} G^{sk}$, contradiction to $(S, a) \in R_F^{cfk}$. Thus $(S', a') \in R_G^{cfk}$.

To show (c) assume $F^\alpha = G^\alpha$ for $\alpha \in \{sk, gk\}$ and let $(S, a) \in R_F^{cfk}$. We show $(S, a) \in R_G^{cfk}$, the other direction is by symmetry. By Proposition 10, $F^{cfk} \subseteq F^\alpha$, thus $(S, a) \in R_F^\alpha$ and therefore $(S, a) \in R_G^\alpha$. By Lemma 21, either $(S, a) \in R_G^{cfk}$ or there is an attack $(T, b) \in R_G$, $T \cup \{b\} \subset S \cup \{a\}$ such that $(T, b) \in R_G^{cfk}$. If the former case holds, then we are done, so we assume that there is an attack $(T, b) \in R_G^{cfk}$ with $T \cup \{b\} \subset S \cup \{a\}$. But then also $(T, b) \in R_F^\alpha$ since $G^{cfk} \subseteq G^\alpha = F^\alpha$, contradiction to $(S, a) \in R_F^{cfk}$. Thus $(S, a) \in R_G^{cfk}$ holds.

To show (b) assume $F^{ak} = G^{ak}$. We show $F^{sk} \subseteq G^{sk}$, the other direction follows by symmetry; moreover, $F^{cfk} = G^{cfk}$ follows by (c). Let $(S, a) \in R_F^{sk}$, then also $(S, a) \in R_F^{ak}$ and $(S, a) \in R_G^{ak}$. Assume $(S, a) \notin R_G^{sk}$, that is, (S, a) is neither active nor critical in R_G. By Lemma 22, there is $(S', b) \in R_G$ with $S' \subseteq S$ and $b \in S$ that is active or critical in F, thus $(S', b) \in R_G^{sk}$, moreover, $(S', b) \in R_G^{ak} = F^{ak}$ since $G^{sk} \subseteq G^{ak}$ and by assumption $F^{ak} = G^{ak}$. But then $(S, a) \notin F^{sk}$, contradiction.

To show (a) assume $F^{ck} = G^{ck}$. We show (1) $F^{ak} = G^{ak}$ (then $F^{sk} = G^{sk}$ and $F^{cfk} = G^{cfk}$ follows by (b)) and (2) $F^{gk} = G^{gk}$. (1) We first show $F^{ak} \subseteq G^{ak}$, the other direction follows by symmetry. Let $(S, a) \in R_F^{ak}$, then also $(S, a) \in R_G^{ck} = R_F^{ck}$ and assume $(S, a) \notin R_G^{ak}$. Since $(S, a) \in R_G^{ck} \setminus R_G^{ak}$ we conclude (S, a) is either ba-inactive or a batc self-attack, that is, there is an argument $b \in S \setminus \{a\}$ such that $(\{a\}, b) \in R_G$. Clearly, $(\{a\}, b)$ is active and thus $(\{a\}, b) \in G^{ck} = F^{ck}$, which makes (S, a) back-attacking in F, contradiction to $(S, a) \in F^{ak}$. (2) To prove $F^{gk} = G^{gk}$ we show that $F^{ck} = G^{ck}$ implies $F \equiv_s^{grd} G$. Consider an arbitrary SETAF H, then $com(F \cup H) = com(G \cup H)$ since F and G are strongly equivalent with respect to complete semantics by Theorem 1. By Proposition 1, $grd(F \cup H) = grd(G \cup H)$ and thus $F \equiv_s^{grd} G$. By Theorem 3, $F^{gk} = G^{gk}$ follows. □

Above relations can be immediately translated in terms of strong equivalence relations. Together with the fact that some kernels characterize strong equivalence with respect to to different semantics at the same time, we obtain the following overall picture.

Corollary 2. *For any two SETAFs F and G,*

1. *$F \equiv_s^{\sigma} G$ if and only if $F \equiv_s^{\tau} G$, for $\sigma, \tau \in \{adm, pref, sem, ideal, eager\}$;*
2. *$F \equiv_s^{stb} G$ if and only if $F \equiv_s^{stage} G$;*
3. *$F \equiv_s^{cf} G$ if and only if $F \equiv_s^{naive} G$;*
4. *$F \equiv_s^{com} G$ implies $F \equiv_s^{\tau} G$ for $\tau \in \{adm, pref, sem, ideal, eager, stb, stage, cf, naive, grd\}$;*
5. *$F \equiv_s^{\sigma} G$ implies $F \equiv_s^{\tau} G$ for $\sigma \in \{adm, pref, sem, ideal, eager\}$ and $\tau \in \{stb, stage, cf, naive\}$;*
6. *$F \equiv_s^{\sigma} G$ implies $F \equiv_s^{\tau} G$ for $\sigma \in \{stb, stage, grd\}$ and $\tau \in \{cf, naive\}$.*

Proof. Statements (1) to (3) follow from Theorem 2, Theorem 4 and Theorem 5, respectively. By Theorem 1, $F \equiv_s^{com} G$ implies $F^{ck} = G^{ck}$, thus $F^{\beta} = G^{\beta}$ for $\beta \in \{ak, sk, cfk, gk\}$ by Proposition 11 (a), hence statement (4) follows by the Theorems 2, 3, 4 and 5. The statements (5) and (6) follow via similar considerations: By Theorem 2, $F \equiv_s^{\sigma} G$ implies $F^{ak} = G^{ak}$ for $\sigma \in \{adm, pref, sem, ideal, eager\}$ thus $F^{\beta} = G^{\beta}$ for $\beta \in \{sk, cfk\}$ by Proposition 11 (b). Consequently, statement (5) follows, i.e., $F \equiv_s^{\tau} G$ for $\tau \in \{stb, stage, cf, naive\}$ by the Theorems 4 and 5. By Theorem 3, $F \equiv_s^{grd} G$ implies $F^{gk} = G^{gk}$ thus $F^{cfk} = G^{cfk}$ by Proposition 11 (c). It follows that $F \equiv_s^{\tau} G$ for $\tau \in \{cf, naive\}$ by Theorem 5. Moreover, by Theorem 4, $F \equiv_s^{\sigma} G$ for $\sigma \in \{stb, stage\}$ implies $F^{sk} =_{sa} G^{sk}$, thus $F^{cfk} =_a G^{cfk}$ by Proposition 11 (d). Therefore $F \equiv_s^{\tau} G$ for $\tau \in \{cf, naive\}$ by Theorem 5 which completes the proof for statement (6). □

Note that above relations carry over to the other variants of strong equivalence considered in this paper. The same holds for our final result that shows that strong equivalence with respect to complete semantics holds exactly if strong equivalence for grounded and an admissible-based semantics holds. Again this generalizes a property of the AF kernels.

Proposition 12. *For SETAFs F, G, $F \equiv_s^{com} G$ if and only if $F \equiv_s^{\sigma} G$ for $\sigma \in \{adm, pref, sem, ideal, eager\}$ and $F \equiv_s^{grd} G$.*

Proof. Let $\sigma \in \{adm, pref, sem, ideal, eager\}$. $F \equiv_s^{com} G$ implies $F \equiv_s^{\sigma} G$ and $F \equiv_s^{grd} G$ by Corollary 2. It remains to show that $F \equiv_s^{\sigma} G$ and $F \equiv_s^{grd} G$ implies $F \equiv_s^{com} G$. Towards a contradiction assume $F \not\equiv_s^{com} G$. Then there exists a SETAF H such that $com(F \cup H) \neq com(G \cup H)$. Without loss of generality, there is $E \in com(F \cup H) \setminus com(G \cup H)$. Observe that $E \in adm(F \cup H)$ and thus $E \in adm(G \cup H)$ by assumption $F \equiv_s^{\sigma} G$. Moreover also $A_{F \cup H} = A_{G \cup H}$. Let $E' = E \cup \Gamma_G(E)$ and notice that E' is admissible in $G \cup H$ by the fundamental lemma, thus also $E \in adm(F \cup H)$.

For a fresh argument c we define

$$H' = (A_{F\cup H} \cup \{c\}, \{(c,b) \mid b \in E^+\} \cup \{(b,b) \mid b \in (A_{F\cup H} \setminus E')\}).$$

$E' \cup \{c\}$ is grounded in $G \cup H$: c is not attacked at all and c defends E' since c attacks every argument $b \in E^+$. We show that $E^+_{G\cup H} = E^+_{F\cup H}$: By Corollary 2, $F \equiv^{cf}_s G$ since $F \equiv^{\sigma}_s G$ thus $cf(F \cup H) = cf(G \cup H)$ and $cf(F) = cf(G)$. Moreover, $F^{ak} = G^{ak}$ since $F \equiv^{\sigma}_s G$, thus F and G have the same active attacks, i.e., $T^+_F \cup T^+_G$ for every $T \in cf(F) = cf(G)$. Consequently $E^+_{G\cup H} = E^+_{F\cup H}$. It follows that $E' \cup \{c\}$ is not grounded in $F \cup H$: For any $a \in E' \setminus E$ we have that a is not defended by c, otherwise E defends a in $F \cup H$ by $E^+ = c^+$ and thus E is not complete in $F \cup H$. □

5 Conclusion

In this work, we investigated different forms of attacks that can occur in SETAFs. We classified them along the main extension-based semantics in the sense whether or not they can be removed from a SETAF without changing the extensions. In addition, we identified situations where self-attacks can be replaced by proper attacks and vice versa, and proposed a normal form for SETAFs. We used the gained insights to provide characterizations for strong equivalence (and restricted variants thereof) which are in line with the concepts of kernels as used to characterize strong equivalence in standard AFs. Notice that our characterizations of strong equivalence also reveal that the complexity of testing strong equivalence between SETAFs remains tractable for all considered semantics (both obtaining the kernel and comparing kernels can be done in polynomial time).

Our results show, on the one hand, that the identification of removable attacks in SETAFs is more opaque than for Dung AFs; we have given a comprehensive classification in Section 3. Notable, while in AFs all classes of removable attacks require self-attacks in the framework, we identify several classes of removable attacks for SETAFs that do not (necessarily) rely on self-attacks. On the other hand, relating strong equivalence in terms of the different semantics (see Corollary 2) for SETAFs provides exactly the same picture as known for Dung AFs, see [4, Section 4: Replaceability]. In our view, this also indicates that SETAFs are a sound generalization of Dung AFs since very fundamental properties that related the semantics to each other still hold.

Future work includes the following topics: For AFs, several weaker notions of strong equivalence have been analyzed [1, 2, 5, 4]. In particular, the parameterized notion from [6] has been identified as suitable for preprocessing AFs [12]. Given that the range of removable attacks in SETAFs is much larger compared to AFs (as we have seen in this work), preprocessing of SETAFs might be even more beneficial in practice. We note that first implementations of SETAF semantics that could benefit from such preprocessing have been presented recently [11]. Another question concerns strong equivalence with respect to labelling-based semantics for SETAFs as recently introduced [14]. We expect here similar results as observed for labelling-based semantics of AFs in [3].

Acknowledgments This research has been supported by WWTF through project ICT19-065, and FWF through projects I2854, P30168, and W1255-N23. The authors are grateful to the anonymous reviewers for their careful reviews and helpful comments.

References

[1] Ringo Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artif. Intell.*, 193:18–44, 2012.

[2] Ringo Baumann. Context-free and context-sensitive kernels: Update and deletion equivalence in abstract argumentation. In Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, editors, *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 63–68. IOS Press, 2014.

[3] Ringo Baumann. Characterizing equivalence notions for labelling-based semantics. In Chitta Baral, James P. Delgrande, and Frank Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 22–32. AAAI Press, 2016.

[4] Ringo Baumann. On the nature of argumentation semantics: Existence and uniqueness, expressibility, and replaceability. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, chapter 18. College Publications, February 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2779-2886.

[5] Ringo Baumann and Gerhard Brewka. The equivalence zoo for Dung-style semantics. *J. Log. Comput.*, 28(3):477–498, 2018.

[6] Ringo Baumann, Wolfgang Dvořák, Thomas Linsbichler, and Stefan Woltran. A general notion of equivalence for abstract argumentation. *Artif. Intell.*, 275:379–410, 2019.

[7] Gerhard Brewka, Sylwia Polberg, and Stefan Woltran. Generalizations of Dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29(1):30–38, 2014.

[8] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[9] Paul E. Dunne, Wolfgang Dvořák, and Stefan Woltran. Parametric properties of ideal semantics. *Artif. Intell.*, 202:1–28, 2013.

[10] Wolfgang Dvořák, Jorge Fandinno, and Stefan Woltran. On the expressive power of collective attacks. *Argument & Computation*, 10(2):191–230, 2019.

[11] Wolfgang Dvořák, Alexander Greßler, and Stefan Woltran. Evaluating SETAFs via Answer-Set Programming. In Matthias Thimm, Federico Cerutti, and Mauro Vallati, editors, *Proceedings of the Second International Workshop on Systems and Algorithms for Formal Argumentation (SAFA 2018) co-located with the 7th International Conference on Computational Models of Argument (COMMA 2018), Warsaw, Poland, September 11, 2018.*, volume 2171 of *CEUR Workshop Proceedings*, pages 10–21. CEUR-WS.org, 2018.

[12] Wolfgang Dvořák, Matti Järvisalo, Thomas Linsbichler, Andreas Niskanen, and Stefan Woltran. Preprocessing argumentation frameworks via replacement patterns. In Francesco Calimeri, Nicola Leone, and Marco Manna, editors, *Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings*, volume 11468 of *Lecture Notes in Computer Science*, pages 116–132. Springer, 2019.

[13] Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. Strong equivalence for argumentation frameworks with collective attacks. In Christoph Benzmüller and Heiner Stuckenschmidt, editors, *KI 2019: Advances in Artificial Intelligence - 42nd German Conference on AI, Kassel, Germany, September 23-26, 2019, Proceedings*, volume 11793 of *Lecture Notes in Computer Science*, pages 131–145. Springer, 2019.

[14] Giorgos Flouris and Antonis Bikakis. A comprehensive study of argumentation frameworks with sets of attacking arguments. *Int. J. Approx. Reasoning*, 109:55–86, 2019.

[15] Sarah Alice Gaggl and Stefan Woltran. Strong equivalence for argumentation semantics based on conflict-free sets. In Weiru Liu, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty - 11th European Conference, ECSQARU 2011, Belfast, UK, June 29-July 1, 2011. Proceedings*, volume 6717 of *Lecture Notes in Computer Science*, pages 38–49. Springer, 2011.

[16] Vladimir Lifschitz, David Pearce, and Agustín Valverde. Strongly Equivalent Logic Programs. *ACM Transactions on Computational Logic*, 2(4):526–541, 2001.

[17] Søren Holbech Nielsen and Simon Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Argumentation in Multi-Agent Systems, Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers*, volume 4766 of *Lecture Notes in Computer Science*, pages 54–73. Springer, 2006.

[18] Emilia Oikarinen and Stefan Woltran. Characterizing strong equivalence for argumentation frameworks. *Artif. Intell.*, 175(14-15):1985–2009, 2011.

[19] Sylwia Polberg. *Developing the abstract dialectical framework*. PhD thesis, TU Wien, Institute of Information Systems, 2017.

[20] Miroslaw Truszczynski. Strong and uniform equivalence of nonmonotonic theories - an algebraic approach. *Ann. Math. Artif. Intell.*, 48(3-4):245–265, 2006.

[21] Hudson Turner. Strong equivalence for causal theories. In Vladimir Lifschitz and Ilkka Niemelä, editors, *Proceedings of the 7th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR'04)*, volume 2923 of *Lecture Notes in Computer Science*, pages 289–301. Springer, 2004.

[22] Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Toward a more efficient generation of structured argumentation graphs. In Sanjay Modgil, Katarzyna Budzynska, and John Lawrence, editors, *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, pages 205–212. IOS Press, 2018.

[23] Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Ranking-based semantics for sets of attacking arguments. Accepted for publication in Proc. AAAI'20; preprint available under `https://hal-lirmm.ccsd.cnrs.fr/lirmm-02374712/`, 2020.

A Alternative Characterizations of Semantics

In this section we provide some equivalent characterization of preferred, semi-stable, ideal, and eager semantics that are in correspondence with well-known results for Dung AFs.

We first show that we can base preferred semantics on complete semantics instead of admissible semantics.

Proposition 13. *$S \in \mathit{pref}(F)$ if and only if $S \in \mathit{com}(F)$ and there is no $T \in \mathit{com}(F)$ s.t. $T \supset S$.*

Proof. only if: For $S \in \mathit{pref}(F)$ we have that $S \in \mathit{adm}(F)$ and thus there is $S' \in \mathit{com}(F)$ with $S \subseteq S'$. Notice that $S' \in \mathit{adm}(F)$ and by the maximality of preferred extensions we obtain $S = S' \in \mathit{com}(F)$. Towards a contradiction assume that there is a $T \in \mathit{com}(F)$ s.t. $T \supset S$. As $\mathit{com}(F) \subseteq \mathit{adm}(F)$ we obtain a contradiction to the initial assumption that $\nexists T \in \mathit{adm}(F)$ s.t. $T \supset S$.

if: From $S \in \mathit{com}(F)$ and $\mathit{com}(F) \subseteq \mathit{adm}(F)$ we obtain $S \in \mathit{adm}(F)$. Towards a contradiction assume that there is a $T \in \mathit{adm}(F)$ s.t. $T \supset S$. Then there is a $\hat{T} \in \mathit{com}[(F)$ with $T \subseteq \hat{T}$ We obtain a contradiction to the initial assumption that $\nexists T \in \mathit{com}(F)$ s.t. $T \supset S$. □

Next we show that we can base semi-stable semantics on complete or preferred semantics instead of on admissible semantics.

Proposition 14. *For every SETAF F, the following statements are equivalent:*

1. *$S \in \mathit{sem}(F)$*
2. *$S \in \mathit{com}(F)$ and $\nexists T \in \mathit{com}(F)$ s.t. $T^{\oplus}_F \supset S^{\oplus}_F$.*
3. *$S \in \mathit{pref}(F)$ and $\nexists T \in \mathit{pref}(F)$ s.t. $T^{\oplus}_F \supset S^{\oplus}_F$.*

Proof. (1) $\Rightarrow$ (3): For admissible sets A, B with $A \subset B$ we have that $A_F^{\oplus} \subset B_F^{\oplus}$ as each $a \in B \setminus A$ is neither contained in A nor attacked by A. Thus $S \in \mathit{pref}(F)$. Towards a contradiction assume that there is a $T \in \mathit{pref}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$. As $\mathit{pref}(F) \subseteq \mathit{com}(F)$ we obtain a contradiction to the initial assumption that $\nexists T \in \mathit{adm}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$.

(3) $\Rightarrow$ (2): By $\mathit{pref}(F) \subseteq \mathit{com}(F)$ we obtain that $S \in \mathit{com}(F)$. Towards a contradiction assume that there is a $T \in \mathit{com}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$. Then there is a $\hat{T} \in \mathit{pref}(F)$ with $T \subseteq \hat{T}$ and thus $\hat{T}_F^{\oplus} \supseteq T_F^{\oplus} \supset S_F^{\oplus}$. We obtain a contradiction to the initial assumption that $\nexists T \in \mathit{pref}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$.

(2) $\Rightarrow$ (1): By $\mathit{com}(F) \subseteq \mathit{adm}(F)$ we obtain that $S \in \mathit{adm}(F)$. Towards a contradiction assume that there is a $T \in \mathit{adm}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$. Then there is a $\hat{T} \in \mathit{com}[(F)$ with $T \subseteq \hat{T}$ and thus $\hat{T}_F^{\oplus} \supseteq T_F^{\oplus} \supset S_F^{\oplus}$. We obtain a contradiction to the initial assumption that $\nexists T \in \mathit{com}(F)$ s.t. $T_F^{\oplus} \supset S_F^{\oplus}$. $\square$

We next consider generalizations of the different characterizations of ideal semantics.

Proposition 15. *For every SETAF F, the following statements are equivalent for $\sigma \in \{\mathit{adm}, \mathit{com}, \mathit{pref}\}$:*

1. *$S \in \mathit{ideal}(F)$*
2. *S is $\subseteq$ maximal among $\{S \in \mathit{com}(F) \mid S \subseteq \bigcap_{E \in \mathit{pref}(F)} E\}$*
3. *S is $\subseteq$ maximal among $\{S \in \mathit{adm}(F) \mid \nexists T \in \sigma(F)$ s.t. T attacks $S\}$*
4. *S is $\subseteq$ maximal among $\{S \in \mathit{com}(F) \mid \nexists T \in \sigma(F)$ s.t. T attacks $S\}$*

Proof. (1)$\Leftrightarrow$(2) We first consider the ideal sets $idealsets(F) = \{S \in \mathit{adm}(F) \mid S \subseteq \bigcap_{E \in \mathit{pref}(F)} E\}$ and show that for each $S \in idealsets(F)$ there is a $S' \in idealsets(F) \cap \mathit{com}(F)$ such that $S \subseteq S'$. Towards a contradiction assume $S \in idealsets(F)$ and there is no $S' \in idealsets(F) \cap \mathit{com}(F)$ such that $S \subseteq S'$. Without loss of generality, we can assume that S is $\subseteq$-maximal in $idealsets(F)$, i.e., $S \in \mathit{ideal}(F)$. Thus there is some $a \notin S$ that is defended by S. As for each preferred extension E we have $S \subset E$ we have that each $E \in \mathit{pref}(F)$ defends a and thus also contains a. That is $S \cup \{a\}$ is admissible and contained in each preferred extension which is in contradiction to the maximality of S. Now as $\mathit{com}(F) \subseteq \mathit{adm}(F)$ we obtain the equivalence between (1) and (2).

(1,2) $\Leftrightarrow$ (3): We show that $idealsets(F) = \{S \in \mathit{adm}(F) \mid \nexists T \in \sigma(F)$ s.t. T attacks $S\}$. First, let $S \in idealsets(F)$. By definition we have that $S \in \mathit{adm}(F)$. Towards a contradiction assume there is an extension $T \in \sigma(F)$ such that T attacks S. That implies that there is a $E \in \mathit{pref}(F)$ such that E attacks $a \in S$. We then have that $a \notin E$ and thus $a \notin S$, a contradiction to our initial assumption.

Now consider a set $S \in \mathit{adm}(F)$ such that there is no $T \in \sigma(F)$ such that T attacks S. That is there is no $E \in \mathit{pref}(F)$ such that E attacks S. Now consider some $E \in \mathit{pref}(F)$. As both S and E are admissible, by [10, Lemma 2], we either have that (a) $S \cup E$ is admissible, or (b) E attacks an argument of S (and vice versa). In the former case, by the maximality of E we have $S \subseteq E$. As observed above we have that for each $E \in \mathit{pref}(F)$ (b) is false and we thus have that S is contained in each preferred extension, i.e., $E \in idealsets(F)$.

(1,2,3) $\Leftrightarrow$ (4): From the above we have that $idealsets(F) \cap \mathit{com}(F) = \{S \in \mathit{adm}(F) \mid \nexists T \in \sigma(F)$ s.t. T attacks $S\} \cap \mathit{com}(F)$. The former equals $\{S \in \mathit{com}(F) \mid S \subseteq \bigcap_{E \in \mathit{pref}(F)} E\}$ and the latter equals $\{S \in \mathit{com}(F) \mid \nexists T \in \sigma(F)$ s.t. T attacks $S\}$. We this obtain (2)$\Leftrightarrow$(4) and the claim follows. $\square$

Finally, we consider generalizations of the different characterizations of eager semantics.

Proposition 16. *For every SETAF F, the following statements are equivalent:*

1. *$S \in \mathit{eager}(F)$*
2. *S is $\subseteq$ maximal among $\{S \in \mathit{com}(F) \mid S \subseteq \bigcap_{E \in \mathit{sem}(F)} E\}$*
3. *S is $\subseteq$ maximal among $\{S \in \mathit{adm}(F) \mid \nexists T \in \mathit{sem}(F)$ s.t. T attacks $S\}$*
4. *S is $\subseteq$ maximal among $\{S \in \mathit{com}(F) \mid \nexists T \in \mathit{sem}(F)$ s.t. T attacks $S\}$*

Proof. (1)⇔(2) We first consider the eager sets $eagersets(F) = \{S \in \mathit{adm}(F) \mid S \subseteq \bigcap_{E \in \mathit{pref}(F)} E\}$. and show that for each $S \in eagersets(F)$ there is a $S' \in eagersets(F) \cap \mathit{com}(F)$ such that $S \subseteq S'$. Towards a contradiction assume $S \in eagersets(F)$ and there is no $S' \in eagersets(F) \cap \mathit{com}(F)$ such that $S \subseteq S'$. Without loss of generality, we can assume that S is $\subseteq$-maximal in $eagersets(F)$, i.e., $S \in \mathit{eager}(F)$. Thus there is some $a \notin S$ that is defended by S. As for each semi-stable extension E we have $S \subset E$ we have that each $E \in \mathit{sem}(F)$ defends a and thus also contains a. That is $S \cup \{a\}$ is admissible and contained in each semi-stable extension which is in contradiction to the maximality of S. Now as $\mathit{com}(F) \subseteq \mathit{adm}(F)$ we obtain the equivalence between (1) and (2).

(1,2) ⇔ (3): We show that $eagersets(F) = \{S \in \mathit{adm}(F) \mid \nexists T \in \mathit{sem}(F)$ s.t. T attacks $S\}$. First, let $S \in eagersets(F)$. By definition we have that $S \in \mathit{adm}(F)$. Towards a contradiction assume there is an extension $T \in \mathit{sem}(F)$ such that T attacks S. We then have that $a \notin T$ and thus $a \notin S$, a contradiction to our initial assumption.

Now consider a set $S \in \mathit{adm}(F)$ such that there is no $T \in \mathit{sem}(F)$ such that T attacks S. Now consider some $T \in \mathit{sem}(F)$. As both S and T are admissible, by [10, Lemma 2], we either have that (a) $S \cup T$ is admissible, or (b) T attacks an argument of S (and vice versa). In the former case, by the maximality of T we have $S \subseteq T$. As observed above we have that for each $T \in \mathit{sem}(F)$ (b) is false and we thus have that S is contained in each semi-stable extension, i.e., $E \in eagersets(F)$.

(1,2,3) ⇔ (4): From the above we have that $eagersets(F) \cap \mathit{com}(F) = \{S \in \mathit{adm}(F) \mid \nexists T \in \mathit{eager}(F)$ s.t. T attacks $S\} \cap \mathit{com}(F)$. The former equals $\{S \in \mathit{com}(F) \mid S \subseteq \bigcap_{E \in \mathit{eager}(F)} E\}$ and the latter equals $\{S \in \mathit{com}(F) \mid \nexists T \in \mathit{sem}(F)$ s.t. T attacks $S\}$. We this obtain (2)⇔(4) and the claim follows. □