Defendable Security in Interaction Protocols

Wojciech Jamroga,[1] Matthijs Melissen,[1] and Henning Schnoor[2]

[1] Computer Science and Communication and Interdisciplinary Centre for Security,
Reliability, and Trust, University of Luxembourg
{wojtek.jamroga, matthijs.melissen}@uni.lu
[2] Arbeitsgruppe Theoretische Informatik, University of Kiel
henning.schnoor@email.uni-kiel.de

Abstract. We study the security of interaction protocols when incentives of participants are taken into account. We begin by formally defining correctness of a protocol, given a notion of rationality and utilities of participating agents. Based on that, we propose how to assess security when the precise incentives are unknown. Then, the security level can be defined in terms of *defender sets*, i.e., sets of participants who can effectively "defend" the security property as long as they are in favor of the property. In terms of technical results, we present a theoretical characterization of defendable protocols under Nash equilibrium, and study the computational complexity of related decision problems.

1 Introduction

Interaction protocols are ubiquitous in multi-agent systems: As soon as two machines communicate, a protocol is required. Protocols can be modeled as games, since every participant in the protocol has several strategies that she can employ. From a game-theoretic perspective, protocols are an interesting class of games since they have a *goal*, i.e., a set of outcomes that are preferred by the designer of the protocol. A subclass of protocols are *security protocols* which use cryptography to enforce their goals against any possible behavior of participants. Such a protocol is deemed correct with respect to its goal if the goal is achieved in all runs where a predefined subset of players follows the protocol.

We point out that this definition of correctness can be too strong, since violation of the goal may be achievable only by irrational responses from the other players. On the other hand, the definition may also prove too weak when the goal can be only achieved by an irrational strategy of agents supporting the goal, in other words: one that they should never choose to play. To describe and predict rational behavior of agents, game theory has proposed a number of *solution concepts* [18]. Each solution concept captures some notion of rationality which may be more or less applicable in different contexts. We do not fix a particular solution concept, but consider it to be a parameter of the problem.

Our main contributions are the following. First, in Section 3.1, we define a parametrized notion of *rational correctness* for security protocols, where the parameter is a suitable solution concept. Secondly, based on this notion, we define a concept of *defendability of security* in a protocol, where the security property is guaranteed under relatively weak assumptions (Section 3.3). Thirdly, we give complexity results for verification of rational protocol correctness and defendability (see Section 4). Fourthly, in Section 5, we propose a *characterization* of defendable security properties when rationality of participants is based on Nash equilibrium. Finally, we extend the results to mixed strategies in Section 6.

1.1 Related Work

There are several meeting points of security protocols and game theory. Some researchers have considered protocol execution as a game with the very pessimistic assumption that the only goal of the other participants ("adversaries") is to break the intended security property of the protocol. In this pessimistic analysis, a protocol is correct if the "honest" participants have a strategy such that, for all strategies of the other agents, the goal of the protocol is satisfied (cf. e.g. [14]). Recently, protocols have been analyzed with respect to some game theoretic notions of rationality [10, 2] where preferences of participants are taken into account. An overview of connections between cryptography and game theory is given in [8]. Another survey [1, 16] presents arguments suggesting that study of incentives in security applications is crucial.

Game theoretic concepts have been applied to analysis of specific security properties in a number of papers. Kremer and Raskin [15] used graph games to verify non-repudiation protocols. However, their method used neither a model of incentives nor of rationality. Buttyán, Hubaux and Čapkun [5] model games in a way similar to ours, and also use incentives to model the behavior of agents. However, they restrict their analysis to strongly Pareto-optimal Nash equilibria which is not necessarily a good solution concept for security protocols. First, it is unclear why agents would *individually* converge to a strongly Pareto-optimal play. Moreover, in many protocols it is unclear why agents would play a Nash equilibrium in the first place. Our method is more general, as we use the solution concept as a parameter to our analysis.

Asharov et al. (2011) [2] use game theory to study gradual-release fair exchange protocols, i.e., protocols in which at any round, the probability of any party to predict the item of the other player increases only by a negligible amount. They model this in a game-theoretical setting, where in every round, the player can either continue or abort. In every round, the item of the other player is predicted. The situation where the player predicts correctly and the other one does not has the highest utility, and the situation where the player predicts incorrectly and the other one predicts correctly the lowest. Then a protocol is said to be game-theoretically fair if the strategy that never aborts the protocol is a computational Nash-equilibrium (i.e., a configuration where no player can gain non-negligible advantage by polynomial-time computable unilateral deviations). They show that no protocol is both fair and effective, but fairness without effectiveness is achievable. They also show that their analysis allows for solutions that are not admitted by the traditional cryptographic definition.

Groce and Katz [12] show that if agents have a strict incentive to achieve fair exchange, then gradual-release fair exchange without trusted third party (TTP) is possible under the assumption that the other agents play rational. Chadha et. al [6] show that in any fair, optimistic, timely contract-signing protocol, there is a point where one player has a strategy to determine whether or not to complete the protocol and obtain a contract. Although they reason about strategies, they do not model incentives explicitly, and do not take different solution concepts into account. Syverson [19] presents a *rational exchange* protocol for which he shows that "enlightened, self-interested parties" have no reason to cheat.

Chatterjee & Raman [7] use assume-guarantee synthesis for synthesis of contract signing protocols. Finally, in [9], a logic for modeling coordination abilities between agents is presented, but incentives are not taken into account. [11] also studies coordination and applies iterated elimination of dominated strategies.

In summary, rationality-based correctness of protocols has been studied in a number of papers, but usually with a particular notion of rationality in mind. In contrast, we define a concept of correctness where a game-theoretic solution concept is a parameter of the problem. Even more importantly, our concept of *defendability* of a security property is completely novel. The same applies to our characterizations of defendable properties under Nash equilibrium.

2 Protocols and Games

We begin by recalling standard concepts used for modeling protocols on the one hand, and games on the other. We also point out where the two meet.

2.1 Security Protocols

A protocol is a specification of how agents should interact. Protocols can contain *choice points* where several actions are available to the agents. An agent is *honest* if he follows the protocol specification, and *dishonest* otherwise. In the latter case, the agent is only restricted by the physical and logical actions that are available in the environment. For instance, in a cryptographic protocol, dishonest agents can do anything that satisfies properties of the cryptographic primitives, assuming perfect cryptography (as in [15]). The protocol specification, together with a model of the environment of action, a subset of agents who are assumed to be honest, and the operational semantics of action execution, defines a multi-agent transition system that we call the *model* of the protocol. In the rest of the paper, we focus on protocol models, and abstract away from how they arise. We also do not treat the usual "network adversary" that can intercept, delay and forge messages, but essentially assume the existence of secure channels. The issue of the "network adversary" is of course highly relevant for the protocols we consider, but orthogonal to the aspects we discuss in this paper. A complete analysis of a protocol needs to take both aspects into account.

As a running example, we consider two-party contract signing protocols. Two agents, Alice and Bob, intend to sign a contract. The two main objectives in such

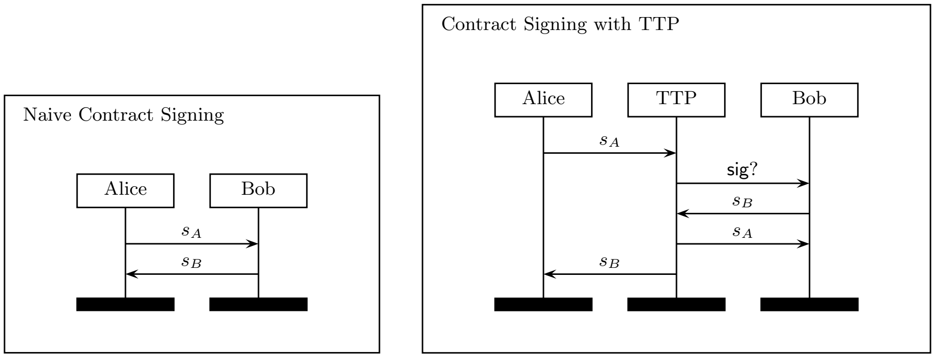

Fig. 1. Two contract-signing protocols

protocols are *fairness* and *effectivity.* Fairness requires that if one agent gets the signature of the other agent, the other agent will eventually get the signature of the first agent as well. A protocol run is *effective* if, at the end of the run, both agents have the signature of the other agent.

Example 1. Figure 1 displays two simple contract signing protocols. In the protocol on the left, Alice sends her signature to Bob, who responds with his signature. Alice and Bob can stop the protocol at any moment (thereby deviating from the protocol). Clearly, the run where Bob and Alice both send their signatures is fair. However, not all protocol runs are fair. In particular, if Bob is dishonest, he can stop the protocol right after receiving the signature of Alice.

The protocol on the right [4] uses a *trusted third party* (TTP) T, assumed to be honest. First, Alice sends her signature s_A to the TTP, then the TTP requests Bob's signature with the message $\mathsf{sig?}$. Subsequently, Bob sends his signature s_B to the TTP. Finally, the TTP forwards the signatures to Bob and Alice. Again, each participant can stop executing the protocol at any point. Fairness is guaranteed as long as the TTP is honest.

2.2 Game Theoretic Models of Interaction

We use *normal-form games* as abstract models of interaction in a protocol.

Definition 1 (Frames and games). *A* game frame *is a tuple $\Gamma = (N, \Sigma, \Omega, o)$, where $N = \{A_1, \ldots, A_{|N|}\}$ is a finite set of* agents, *$\Sigma = \Sigma_{A_1} \times \cdots \times \Sigma_{A_{|N|}}$ is a set of strategy profiles, Ω is the set of outcomes, and $o : \Sigma \to \Omega$ is a function mapping each strategy profile to an outcome.*

A normal-form (NF) game *is a game frame plus a* utility profile $u = \{u_1, \ldots, u_{|N|}\}$ *where $u_i : \Sigma \to \mathbb{R}$ is a utility function assigning utility values to strategy profiles.*

Game theory uses *solution concepts* to define which strategy profiles capture rational interactions. Let $\mathcal{G}$ be a class of games with the same strategy profiles Σ. Formally, a solution concept for $\mathcal{G}$ is a function $SC : \mathcal{G} \to \mathcal{P}(\Sigma)$ that,

$A\backslash B$	*Stop*	*Sign*
Stop	$\emptyset$	$\emptyset$
Sign	$\{\mathsf{sign}_B\}$	$\{\mathsf{sign}_A, \mathsf{sign}_B\}$

(a) Γ_1

$A\backslash B$	*Stop*	*Sign*
Stop	$\emptyset$	$\emptyset$
Sign	$\emptyset$	$\{\mathsf{sign}_A, \mathsf{sign}_B\}$

(b) Γ_2

Fig. 2. Game frames for contract signing: (a) naive protocol, (b) protocol with TTP

$A\backslash B$	*Stop*	*Sign*
Stop	$1, 1$	$1, 1$
Sign	$0, 3$	$2, 2$

(a) G_1

$A\backslash B$	*Stop*	*Sign*
Stop	$1, 1$	$1, 1$
Sign	$1, 1$	$2, 2$

(b) G_2

Fig. 3. NF games for contract signing: : (a) naive protocol, (b) protocol with TTP

given a game, returns a set of *rational* strategy profiles. Well-known solution concepts include e.g. Nash equilibrium (NE), dominant and undominated strategies, Stackelberg equilibrium, Pareto optimality etc. For a detailed discussion, see [18].

2.3 Protocols as Games

Let P be a model of a protocol. We will investigate properties of P through the game frame $\Gamma(P)$ in which strategies are *conditional plans* in P, i.e., functions that specify for each choice point which action to take. A set of strategies, one for each agent, uniquely determines a *run* of the protocol, i.e., a sequence of actions that the agents will take. $\Gamma(P)$ takes runs to be the outcomes in the game, and hence maps strategy profiles to runs.

Example 2. Consider the protocols in Figure 1. Alice and Bob have the following strategies: *Stop* (stopping before sending the signature) and *Sign* (running the protocol honestly). The protocol can be modeled as game frame $\Gamma = (N, \Sigma, \Omega, o)$ with $N = \{A, B, T\}$, $\Sigma_T = \{-\}$ (the trusted third party T is deterministic), $\Sigma_A = \Sigma_B = \{Stop, Sign\}$, $\Omega = \mathcal{P}(\{\mathsf{sign}_A, \mathsf{sign}_B\})$, here sign_A (sign_B) denotes the event that Alice (Bob) gets a signed copy of the contract. For the protocol with TTP protocol we also have $o(\sigma) = \{\mathsf{sign}_A, \mathsf{sign}_B\}$ if $\sigma = (Sign, Sign, -)$, and $o(\sigma) = \emptyset$ otherwise. The game frame is displayed in Figure 2(b).

The "naive" protocol can be modeled in a similar way (Figure 2(a)). The available strategies are the same, but in this case $o(Sign, Stop, -) = \{\mathsf{sign}_B\}$.

We model agents' preferences with respect to outcomes by utility profiles.

Example 3. Assume the following utility function for A: $u_A(\{\mathsf{sign}_A\}) = 3$, $u_A(\{\mathsf{sign}_A, \mathsf{sign}_B\}) = 2$, $u_A(\emptyset) = 1$, $u_A(\{\mathsf{sign}_B\}) = 0$, and symmetrically for B. Thus, both agents prefer the exchange of signatures over no exchange; moreover, the most preferred option for an agent is to get the signature while the other agent does not, and the least preferred option is not to get the signatures while the other agent does. Combining this utility profile with the game frames from Figure 2 yields the normal-form games depicted in Figure 3.

Security protocols are designed to achieve one or more *security requirements* and/or *functionality requirements*. We only consider requirements that can be expressed in terms of individual runs having a certain property. We model this by a subset of outcomes, called the *objective of the protocol*.

Definition 2. *Given a game frame $\Gamma = (N, \Sigma, \Omega, o)$, an* objective *is a non-empty set $\gamma \subseteq \Omega$. We call γ* nontrivial *in Γ iff γ is neither impossible nor guaranteed in Γ, i.e., $\emptyset \neq \gamma \neq \Omega$.*

Example 4. Consider the following simple definition of fairness. A run is fair iff either both agents obtain the signature of the other agent, or none of them does. Moreover, the run is effective iff both agents obtain the other agent's signature. We can represent that by $\gamma_{\mathsf{fair}} = \{\emptyset, \{\mathsf{sign}_A, \mathsf{sign}_B\}\}$ for the fairness objective, and $\gamma_{\mathsf{eff}} = \{\{\mathsf{sign}_A, \mathsf{sign}_B\}\}$ for effectiveness.

3 Incentive-Based Security Analysis

In this section, we give a definition of correctness of security protocols that takes into account rational decisions of agents, based on their incentives. In NF games, it is often assumed that the mapping o is a bijection, i.e., every strategy profile determines a unique outcome. We therefore often identify Σ with Ω, and omit Ω and o from the representation of games to simplify notation.

3.1 Incentive-Based Correctness

As we have pointed out, the requirement that all strategy profiles satisfy the objective might be too strong. Instead, we will require that all *rational* runs satisfy the objective. In case there are no rational runs, all outcomes are equally rational; then, we adopt the usual pessimistic view and require that all outcomes must satisfy γ.

Definition 3. *A protocol model represented as game frame $\Gamma = (N, \Sigma)$ with utility profile u is* correct with respect to objective γ under solution concept *SC, written $(\Gamma, u) \models_{SC} \gamma$, iff:*

$$\begin{cases} SC(\Gamma, u) \subseteq \gamma & \textit{if } SC(\Gamma, u) \neq \emptyset \\ \gamma = \Sigma & \textit{otherwise.} \end{cases}$$

Protocol verification *is the following decision problem:*

- ***Input:*** *A protocol model P, a utility function u, an objective γ and a solution concept SC.*
- ***Question:*** *Does $(\Gamma(P), u) \models_{SC} \gamma$ hold?*

Example 5. Consider game G_1 from Figure 3(a) for the naive contract signing protocol. We saw that if Alice signs, Bob might stop the protocol, resulting in the worst possible utility for Alice. Therefore, Alice might consider it safer to never sign. This kind of reasoning can be captured by using Nash equilibrium as the solution concept, since $\mathrm{NE}(G_1) = \{(Stop, Stop)\}$. For $\gamma_{\mathsf{eff}} = \{(Sign, Sign)\}$, we have $\mathrm{NE}(G_1) \not\subseteq \gamma_{\mathsf{eff}}$, and thus G_1 is not effective under Nash equilibrium. On the other hand, for $\gamma_{\mathsf{fair}} = \{(Stop, Stop), (Sign, Sign)\}$, we have $\mathrm{NE}(G_1) \subseteq \gamma_{\mathsf{fair}}$, so G_1 guarantees fairness under Nash equilibrium.

Moreover, if we think that the players are willing to take risks in order to obtain a better outcome, then using e.g. Halpern and Rong's maximal perfect collaborative equilibrium [13] as the solution concept is more appropriate. Since $MPCE(G_1) = \{(Sign, Sign)\} \subseteq \gamma_{\mathsf{eff}} \subseteq \gamma_{\mathsf{fair}}$, we have that the protocol is both fair and effective under $MPCE$.

The above example highlights that, for different situations, different solution concepts are appropriate.

3.2 Unknown Incentives

In the previous section, we studied correctness of a protocol when a utility profile is given. However, the exact utility profiles are often unknown. One way out is to require the protocol to be correct for *all possible* utility profiles.

Definition 4. *A protocol model represented by game frame Γ is* valid *with respect to objective γ under solution concept SC (written $\Gamma \models_{SC} \gamma$) iff $(\Gamma, u) \models_{SC} \gamma$ for all utility profiles u.*

Protocol validity *is the following decision problem:*

- ***Input:*** *A protocol model P, an objective γ and a solution concept SC.*
- ***Question:*** *Does $\Gamma(P) \models_{SC} \gamma$ hold?*

It turns out that, under some reasonable assumptions, protocols are only valid for trivial objectives.

Definition 5. *Let $G = (N, \Sigma, (u_1, \ldots, u_n))$. Let $\pi = (\pi_1, \ldots, \pi_n)$, where for all $i \in N$, $\pi_i : \Sigma_i \to \Sigma_i$ is a permutation on Σ_i. We slightly abuse the notation by writing $\pi((s_1, \ldots, s_n))$ for $(\pi_1(s_1), \ldots, \pi_n(s_n))$. A solution concept is* closed under permutation *iff $s \in SC((N, \Sigma, (u'_1, \ldots, u'_n)))$ if and only if $\pi(s) \in SC((N, \Sigma, (u'_1 \circ \pi_1^{-1}, \ldots, u'_n \circ \pi_n^{-1})))$.*

Being closed under permutation is a very natural property. Essentially, it means that "renaming" of strategies does not have an effect on the output of the game. All solution concepts that we know of are closed under permutation.

Theorem 1. *If SC is closed under permutation, then $\Gamma \models_{SC} \gamma$ iff $\gamma = \Sigma$.*

Proof. Let Γ be a game frame, SC be a solution concept closed under permutation of utilities, and γ be an objective. Fix u such that $u_i(s) = 0$ for all $i \in N$ and $s \in \Sigma$. First, if $SC(\Gamma, u) = \emptyset$ then γ must be equal to Σ by Definition 3. On the other hand, suppose that $s \in SC(\Gamma, u)$, and consider any other $s' \in \Sigma$. It is easy to see that there is a permutation π such that $\pi(s) = s'$. Also, $(\Gamma, u \circ \pi)$ is the same game as (Γ, u) for this very special utility function u. By the closure property, $s' \in SC(\Gamma, u \circ \pi) = SC(\Gamma, u)$, which concludes the proof.

Thus, correctness for all distributions of incentives is equivalent to correctness in all possible runs. This characterization is natural: If we do not make any assumptions about the incentives of the participating agents, then no run can be regarded as "irrational," hence all runs need to be taken into account. Clearly, incentive-based analysis needs some assumptions about the incentives of the agents participating in a protocol.[3]

In the next section we look at the case where a subset of agents D, called the *defenders* of the protocol, have a genuine interest in achieving the objective of the protocol.

3.3 Defendability of Protocols

Typical analysis of a protocol implicitly assumes some participants to be aligned with its purpose. E.g., one usually assumes that communicating parties are interested in exchanging a secret without the eavesdropper getting hold of it, that a bank wants to prevent web banking fraud etc. In this section, we formalize this idea by assuming a subset of agents, called the *defenders* of the protocol, to be in favor of its objective. Our new definition of correctness says that a protocol is correct with respect to some objective γ if and only if it is correct with respect to every utility profile in which the preferences of all defenders comply with γ.

Definition 6. *A group of agents $D \in N$* supports *the objective γ in game (N, Σ, u) iff for all $i \in D$, if $s \in \gamma$ and $s' \in \Sigma \setminus \gamma$ then $u_i(s) > u_i(s')$.*

A protocol model represented as game frame Γ is defended *by agents D, written $\Gamma \models_{SC} [D]\gamma$, iff $(\Gamma, u) \models_{SC} \gamma$ for all utility profiles u such that D supports γ in game (Γ, u).*

Protocol defendability *is the following decision problem:*

- ***Input:*** *Protocol model P, objective γ, set of agents D, solution concept SC.*
- ***Question:*** *Does $\Gamma(P) \models_{SC} [D]\gamma$ hold?*

For example, it makes sense to assume that if Alice signs the contract then she prefers to get it also signed by Bob. In other words, Alice supports fairness for herself. Note that the issue of support is different from that of honest execution of a protocol. The former is about preferences of a party; the latter about the actions that the party is bound to select. In particular, there might be protocols in which the objective can be only obtained by deviating from the protocol.[4] We do not go deeper into that, and focus only on defendability. We first note

[3] An interesting special case of Theorem 1 appears in a study of rational secret sharing: Asharov and Lindell [3] proved that the *length* (number of rounds) of a protocol for rational secret sharing must depend on the utilities of the involved agents for the possible protocol outcomes. In particular, there can be no single protocol which works for every possible set of incentives of the agents. Their result even holds under some plausibility assumptions on the agents' incentives (i.e., agents prefer to learn the secret over not learning it, etc.).

[4] Arguably, that would mean that the protocol is badly designed.

that $\mathit{\Gamma}(P) \models_{SC} [D]\gamma$ is *not* equivalent to D having a strategy to achieve γ: It is possible that such a strategy exists, but is not rational in the sense of the solution concept SC. We begin by investigating the borderline cases, where either none or all of the agents are defenders. Clearly, if there are no defenders, then defendability is equivalent to ordinary protocol validity.

Proposition 1. *If $\mathit{\Gamma}$ is a game frame and SC is a solution concept, we have that $\mathit{\Gamma} \models_{SC} [\emptyset]\gamma$ iff $\mathit{\Gamma} \models_{SC} \gamma$.*

If all agents are defenders, any protocol is correct, as long as the solution concept does not select *strongly Pareto-dominated* strategy profiles.

Definition 7. *A solution concept is* weakly Pareto *iff it never selects a strongly Pareto dominated outcome (i.e., such that there exists another outcome strictly preferred by all the players). It is* efficient *iff it never returns the empty set.*

Theorem 2. *If $\mathit{\Gamma}$ is a game frame and SC is an efficient weakly Pareto solution concept then $\mathit{\Gamma} \models_{SC} [N]\gamma$.*

Proof. Let $\mathit{\Gamma}$ be a game frame, and let u be a utility function such that D supports γ in game $(\mathit{\Gamma}, u)$. We have $SC(\mathit{\Gamma}, u) \neq \emptyset$ by assumption. Now we prove that $SC(\mathit{\Gamma}, u) \subseteq \gamma$. Assume $s \in SC(\mathit{\Gamma}, u)$ and $s \notin \gamma$. Let $s' \in \gamma$. Then $u_i(s') > u_i(s)$ for all $i \in N$. However, this implies that $s \notin SC(\mathit{\Gamma}, u)$, which is a contradiction.

Theorem 2 says that if our notion of rationality is efficient and weakly Pareto then designing a protocol for friendly agents is very easy. That is, rather unsurprisingly, if all players are defenders of the a goal γ then as long as there is *some* way of achieving the goal, the players will identify a working strategy. Many solution concepts are both efficient and weakly Pareto, for example: Stackelberg equilibrium, maximum-perfect cooperative equilibrium, backward induction and subgame-perfect Nash equilibrium in perfect information games. However, using such a solution concept is based on optimistic assumptions about both the players' goodwill and their ability to coordinate their strategies. In practice, other solution concepts are used, which in general do not satisfy the preconditions of Theorem 2. For example, Nash equilibrium is neither weakly Pareto nor efficient,[5] and equilibrium in dominant strategies is weakly Pareto but not necessarily efficient.

Given a protocol model, a solution concept and an objective, we can determine the smallest set of defenders for which the protocol is correct. Clearly, defendability of a protocol is monotonic with respect to the set of defenders.

Proposition 2. *For every $D \subseteq D' \subseteq N$, if $\mathit{\Gamma} \models_{SC} [D]\gamma$ then $\mathit{\Gamma} \models_{SC} [D']\gamma$.*

This justifies the following definition.

Definition 8. *The* game-theoretic security level *of protocol P is the antichain of minimal sets of defenders that make the protocol correct.*

[5] We will look closer at defendability under Nash equilibrium in Section 5.

Intuitively, the game-theoretic security level is the set of minimal coalitions $C \subseteq N$ such that if C supports the goal, then every rational play will fulfill it. Note that due to Proposition 2, the game-theoretic security level of a protocol is nonempty (i.e., the goal of the protocol can be defended) if and only if the "grand coalition" N of all players can defend the goal. We will concentrate on defendability by the grand coalition in Section 5.

4 Computational Complexity: General Case

In this section, we study the complexity of protocol verification, validity, and defendability for the general case when the solution concept is a parameter of the problem (and hence a part of the input). Algorithms for this case are useful to evaluate a protocol with respect to different solution concepts. Also, they give upper bounds for every specific subclass of the problem.[6] Since our definitions of correctness are parametrized by a solution concept and a security property, our results are relative to the complexity of verification for the two.

Theorem 3. *Let us measure the complexity w.r.t. the size of the NF game (i.e., the number of strategy profiles), and let $\mathfrak{SC}$, $\mathfrak{Obj}$ be the complexity of verification for the solution concept and the objective of the protocol, respectively. Then:*

1. Protocol verification *is in* $\mathbf{P}^{\mathfrak{SC} \cup \mathfrak{Obj}}$ *which is the class of decision problems that can be solved by a deterministic Turing machine running in polynomial time and making calls to an oracle for problems in* $\mathfrak{SC} \cup \mathfrak{Obj}$*;*
2. Protocol validity *is in* $\mathbf{coNP}^{\mathfrak{SC} \cup \mathfrak{Obj}}$ *(problems solvable by a TM for* **coNP***, calling an oracle for* $\mathfrak{SC} \cup \mathfrak{Obj}$*);*
3. Protocol defendability *is also in* $\mathbf{coNP}^{\mathfrak{SC} \cup \mathfrak{Obj}}$.

Proof. **Ad. 1.** For protocol verification, it suffices to check the outcome of every strategy profile whether it is not accepted by the solution concept or accepted by the objective. This can be done by a deterministic algorithm running in polynomial time (wrt to the number of strategy profiles) and making calls to oracles verifying the solution concept and the objective, respectively;

Ad. 2. We can reduce protocol validity to an instance of **coSAT** making calls to oracles for the solution concept and the objective. This is because every solution concept can be equivalently rephrased in terms of preference relations over outcomes rather than utility profiles. Since there are exponentially many such relations, they can be encoded by polynomially many binary variables. Then, a protocol is valid iff it is correct for *all possible valuations of the variables*;

Ad. 3. Protocol defendability reduces analogously.

The next theorem proposes a lower bound.

Theorem 4. *Protocol validity and protocol defendability are* **coNP***-hard.*

[6] In Section 5, we will give complexity results for the specific case when the notion of rationality is based on Nash equilibrium.

Proof. We prove hardness by a reduction of **coSAT** to protocol validity. Since protocol validity is a special case of defendability (Theorem 1), **coNP**-hardness for defendability follows as well.

Let $x_1, \ldots, x_n$ be Boolean variables and φ a formula in DNF. We construct an instance of protocol validity by simulating valuations of $x_1, \ldots, x_n$ by utility profiles, and formula φ by the solution concept. Formally, let Γ consist of $N = \{1, \ldots, n\}$ and $\Sigma = \{s^0, s^1\}$, and SC be defined as: $SC(\Gamma, u) = \{s^1\}$ if $\varphi\big((u_1(s^1) \geq u_1(s^0)), \ldots, (u_n(s^1) \geq u_n(s^0))\big)$ and $\{s^0\}$ otherwise. Finally, let $\gamma = \{s^1\}$. Note that membership in $SC(\Gamma, u)$ and γ can be verified in polynomial time. Now, $\mathbf{coSAT}(x_1, \ldots, x_n, \varphi)$ iff $\Gamma \models_{SC} \gamma$.

In practice, a protocol model is rarely given as a normal form game, but rather as a sequence of transitions (cf. for example Figure 1). For this representation, the following theorem gives the complexity of protocol verification and validity:

Theorem 5. *Let* $\mathfrak{SC}$, $\mathfrak{Obj}$ *be as above.* Protocol verification, protocol validity, *and* protocol defendability *are in* $\mathbf{coNP}^{\mathfrak{SC} \cup \mathfrak{Obj}}$ *wrt the number of possible transitions in the protocol model.*

Proof. We observe that a strategy profile in an extensive game can be encoded by an array of choices, one per agent and game position (i.e., protocol state in our case). Since the array has polynomial size wrt the size of the game tree, we obtain the result by analogous reasoning to the proof of Theorem 3.

We note that some natural solution concepts can be verified in deterministic polynomial time (e.g., Stackelberg equilibrium in NF games, subgame-perfect Nash equilibrium in EF games, etc.). Also, many objectives can be verified in polynomial time. Then, we obtain the following.

Theorem 6. *If the solution concept and the objective can be verified in polynomial time then:*

1. Protocol verification *is* **P**-*complete wrt to the size of the NF game and* **coNP**-*complete wrt the number of transitions in the protocol;*
2. Protocol validity *and* protocol defendability *are* **coNP**-*complete with respect to both types of input.*

5 Characterizing Defendability under Nash Equilibrium

In this section, we turn to properties that can be defended if agents' rationality is based on Nash equilibrium or Optimal Nash Equilibrium.

5.1 Defendability under Nash Equilibrium

From Theorem 1, we know that no protocol is valid under Nash equilibrium (NE) for any nontrivial objective, since NE is closed under permutation. Do things get better if we assume some agents to be in favor of the security objective? We now

	t_1	t_2
s_1	hi, hi	$0, 0$
s_2	$0, 0$	lo, lo

(a)

	t_1	t_2	t_3
s_1	hi, lo	lo, hi	$0, 0$
s_2	lo, hi	hi, lo	$0, 0$
s_3	$0, 0$	$0, 0$	$0, 0$

(b)

Fig. 4. (a) HiLo game for 2 players; (b) Extended matching pennies. In both games, we assume that $hi > lo > 0$, e.g., $hi = 100$ and $lo = 1$

look at the extreme variant of the question, i.e., defendability by the grand coalition N. Note that, by Proposition 2, nondefendability by N implies that the objective is not defendable by any coalition at all.

Our first result in this respect is negative: we show that in every game frame there are nontrivial objectives that are not defendable under NE.

Theorem 7. *Let Γ be a game frame with at least two players and at least two strategies per player. Moreover, let γ be a singleton objective, i.e., $\gamma = \{\omega\}$ for some $\omega \in \Sigma$. Then, $\Gamma \not\models_{\mathrm{NE}} [N]\gamma$.*

Proof. Assume wlog that $N = 2$, $\Sigma_1 = \{s_1, s_2\}$, $\Sigma_2 = \{t_1, t_2\}$, and $\gamma = \{(s_1, t_1)\}$. Now, consider the utility function u^{hl} of the well known HiLo game (Figure 4(a)). Clearly, N support γ in u^{hl}. Moreover, $\mathrm{NE}(\Gamma, u^{hl}) \neq \emptyset$. On the other hand, $\mathrm{NE}(\Gamma, u^{hl}) = \{(s_1, t_1), (s_2, t_2)\} \not\subseteq \gamma$, which concludes the proof.

In particular, the construction from the above proof shows that, as mentioned before, there are cases where the "defending" coalition has a strategy to achieve a goal γ, but there are still rational plays in which the goal is not achieved.

To present the general result that characterizes defendability of security objectives under Nash equilibrium, we need to introduce additional concepts. In what follows, we use $s[t_i/i]$ to denote $(s_1, \ldots, s_{i-1}, t_i, s_{i+1}, \ldots, s_N)$, i.e., the strategy profile that is obtained from s when player i changes her strategy to t_i.

Definition 9. *Let γ be a set of outcomes (strategy profiles) in Γ. The* deviation closure *of γ is defined as $Cl(\gamma) = \{s \in \Sigma \mid \exists i \in N, t_i \in \Sigma_i \,.\, s[t_i/i] \in \gamma\}$.*

$Cl(\gamma)$ extends γ with the strategy profiles that are reachable by unilateral deviations from γ. Thus, $Cl(\gamma)$ can be seen as the closure of γ with the outcomes that are relevant for Nash equilibrium. Moreover, the following notion captures strategy profiles that can be used to construct sequences of unilateral deviations ending up in a cycle.

Definition 10. *A* strategic knot *in γ is a subset of strategy profiles $S \subseteq \gamma$ such that there is a permutation $(s^1, \ldots, s^k)$ of S where: (a) for all $1 \leq j < k$, $s^{j+1} = s^j[s_i^{j+1}/i]$ for some $i \in N$, and (b) $s^j = s^k[s_i^1/i]$ for some $i \in N, j < k$.*

Essentially, this means that every strategy s^{j+1} is obtained from s^j by a unilateral deviation of a single agent. If these deviations are rational (i.e., increase the utility of the deviating agent), then the knot represents a possible endless loop of rational, unilateral deviations which precludes a group of agents from reaching a stable joint strategy. We now state the main result of this section.

Theorem 8. *Let Γ be a finite game frame and γ a nontrivial objective in Γ. Then, $\Gamma \models_{\mathrm{NE}} [N]\gamma$ iff $Cl(\gamma) = \Sigma$ and there is a strategy profile in γ that belongs to no strategic knots in γ.*

Proof. **"⇒"** Let $\Gamma \models_{\mathrm{NE}} [N]\gamma$, and suppose that $Cl(\gamma) \neq \Sigma$. Thus, there exists $s_0 \in \Sigma$ which is not in $Cl(\gamma)$. Consider a HiLo-style utility function $u(s) = hi$ if $s \in \gamma$, lo if $s = s_0$, and 0 otherwise (for some values $hi > lo > 0$). Clearly, s_0 is a Nash equilibrium in (Γ, u), and thus $\mathrm{NE}(\Gamma, u) \neq \emptyset$ but also $\mathrm{NE}(\Gamma, u) \not\subseteq \gamma$, which is a contradiction.

Suppose now that $Cl(\gamma) = \Sigma$ but every $s \in \gamma$ belongs to a strategic knot. We construct the utility function akin to the extended matching pennies game (Figure 4(b)), i.e., for every node s in a strategic knot $u_i(s) = hi$ for the agent i who has just deviated, and lo for the other agents.[7] Moreover, $u_i(s) = 0$ for all $i \in N, s \notin \gamma$. Clearly, u_i is consistent with γ for every $i \in N$. On the other hand, no $s \in \Sigma$ is a Nash equilibrium: if s is outside of γ then there is a profitable unilateral deviation into γ, and every s inside γ lies on an infinite path of rational unilateral deviations. Thus, $\mathrm{NE}(\Gamma, u) = \emptyset$. Since γ is nontrivial, we have $\Gamma \not\models_{\mathrm{NE}} [N]\gamma$, a contradiction again.

"⇐" Assume $Cl(\gamma) = \Sigma$ and $s \in \gamma$ belongs to no strategic knot in γ. Let u be a utility function such that for every $i \in N, s \in \gamma, s' \in \Sigma \setminus \gamma$ it holds that $u_i(s) > u_i(s')$. Take any $\omega \notin \gamma$. Since $\omega \in Cl(\gamma)$, there is an agent i with a unilateral deviation to some $s \in \gamma$. Note that $u_i(\omega) < u_i(s)$, so $\omega \notin \mathrm{NE}(\Gamma, u)$. Thus, $\mathrm{NE}(\Gamma, u) \subseteq \gamma$. Moreover, s is a Nash equilibrium or there is a sequence of unilateral deviations leading from s to a Nash equilibrium (since Γ is finite and s does not lie on a knot). Thus, also $\mathrm{NE}(\Gamma, u) \neq \emptyset$, which concludes the proof.

Example 6. Consider contract signing with TTP, cf. Figure 1 (right). The properties of effectiveness and fairness can be defined as $\gamma_{\mathsf{eff}} = \{(Sign, Sign)\}$ and $\gamma_{\mathsf{fair}} = \{(Stop, Stop), (Sign, Sign)\}$. By Theorem 8, fairness in the protocol is N-defendable under Nash equilibrium. On the other hand, effectiveness is not.

It is important to note that the above result makes verification of defendability significantly easier than the general results from Section 4 suggest:

Theorem 9. *Let Γ be a finite game frame and γ a nontrivial objective in Γ. Then, checking $\Gamma \models_{\mathrm{NE}} [N]\gamma$ can be done in polynomial time wrt the size of γ.*

Proof (sketch). Checking if $Cl(\gamma) = \Sigma$: we look at every $s \notin \gamma$ and check if it can be "moved" to γ by a flip of an individual strategy.

Checking strategic knots: (i) Take Θ to be the deviation grid for γ, i.e., the graph containing strategy profiles from γ as vertices and individually rational deviations as edges; (ii) Construct the minimal spanning graph $MSG(\Theta)$

[7] If a node lies on several knots, we need to assign several different hi values in a careful way; we omit the details here due to lack of space.

(Kruskal or a similar algorithm); (iii) Let $Knotty = \emptyset$. For every edge (s, s') in $\Theta \setminus MSG(\Theta)$: add the vertices on the path from s to s' in $MSG(\Theta)$ to $Knotty$; (iv) For every $s \in \Theta \setminus Knotty$: add it to $Knotty$ iff there is a path in Θ between s and some $s' \in Knotty$. (v) The answer is "yes" iff $\Theta \setminus Knotty \neq \emptyset$.

5.2 Optimal Nash Equilibria

Nash equilibrium is a natural solution concept for a game played repeatedly until the behavior of all players converges to a stable point. For a one-shot game, NE possibly captures convergence of the process of deliberation. It can be argued that, among the available solutions, no player should contemplate those which are strictly worse for everybody when compared to another stable point. This gives rise to the following refinement of Nash equilibrium: $\mathrm{OptNE}(\Gamma, u)$ is the set of *optimal Nash equilibria* in game (Γ, u), defined as those equilibria *that are not strongly Pareto-dominated by another Nash equilibrium.* Defendability by the grand coalition under OptNE has the following simple characterization.

Theorem 10. *Let Γ be a finite game frame and γ a nontrivial objective in Γ. Then, $\Gamma \models_{\mathrm{OptNE}} [N]\gamma$ iff there is a strategy profile in γ that belongs to no strategic knots in γ.*

Proof. **"⇒"** If all $s \in \gamma$ lie on strategic knots in γ then there is u such that no $s \in \gamma$ is a Nash equilibrium in (Γ, u), cf. the proof of Theorem 8. Since $\mathrm{OptNE}(\Gamma, u) \subseteq \mathrm{NE}(\Gamma, u)$ and γ is nontrivial, this implies that $\Gamma \not\models_{\mathrm{OptNE}} [N]\gamma$.

"⇐" Let u be any utility profile. By analogous reasoning to Theorem 8, there must be a strategy profile $s \in \gamma$ in game (Γ, u) which is an optimal Nash equilibrium. Thus, $\mathrm{OptNE}(\Gamma, u) \neq \emptyset$. Suppose that there exists another optimal NE $s' \notin \gamma$. But then s' would be strictly Pareto-dominated, which cannot be the case. Thus, also $\mathrm{OptNE}(\Gamma, u) \subseteq \gamma$, and hence $\Gamma \models_{\mathrm{OptNE}} [N]\gamma$.

It is easy to see that checking N-defendability under OptNE is in **P**.

6 Defendability in Mixed Strategies

So far, we considered only deterministic (pure) strategies. It is well known that for many games and solution concepts, rational strategies exist only when taking mixed strategies into account. We now extend our definition of correctness to mixed strategies, i.e., randomized conditional plans represented by probability distributions over pure strategies from Σ_{A_i}. Let $dom(s)$ be the support (domain) of a mixed strategy profile s, i.e., the set of pure strategy profiles that have nonzero probability in s. We extend the notion to sets of mixed strategy profiles in the obvious way. By SC^m we denote the variant of SC in mixed strategy profiles. A protocol is correct in mixed strategies iff all the possible behaviors resulting from a rational (mixed) strategy profile satisfy the goal γ; formally: $\Gamma, u \models^m_{SC} \gamma$ iff $dom(SC^m(\Gamma, u)) \subseteq \gamma$ when $SC^m(\Gamma, u) \neq \emptyset$ and $\gamma = \Sigma_\Gamma$ otherwise. The definitions of protocol validity and defendability in mixed strategies ($\Gamma \models^m_{SC}$

γ and $\Gamma \models^m_{SC} [D]\gamma$) are analogous. For defendability in mixed strategies under Nash equilibrium, we have the following, rather pessimistic result.

Theorem 11. *Let Γ be a finite game frame, and γ an objective in it. Then, $\Gamma, u \models^m_{\mathrm{NE}} [N]\gamma$ iff $\gamma = \Sigma$.*

Proof. "$\Leftarrow$" Straightforward. For "$\Rightarrow$", we observe the following:

(i) $Cl(\gamma) = \Sigma$ by the same reasoning as for pure strategies.

(ii) Let $Conv(\gamma)$ be the convex closure of γ, i.e., the set of strategy profiles obtained by combining individual strategies occurring in γ. Then, $Conv(\gamma) = \gamma$. (*Proof:* suppose that it is not the case, then there must be $s, s' \in \gamma$ such that one of their convex combinations s'' is not in γ. We play the Coordination game with 1 assigned to s, s', 0 to the other nodes in γ and -1 to the rest of nodes. The strategy profile $([s_{A_1}/0.5, s'_{A_1}/0.5], \ldots, [s_{A_{|N|}}/0.5, s'_{A_{|N|}}/0.5])$ is a mixed strategy Nash equilibrium, and clearly s'' is in its support. Since $\Gamma, u \models^m_{SC} [N]\gamma$, we have that $s'' \in \gamma$, which is a contradiction.)

(iii) By (i), every i's strategy must be a part of some strategy profile in γ. Thus, $Conv(\gamma) = \Sigma$, and hence $\gamma = \Sigma$.

On the other hand, it turns out that *optimal Nash equilibrium* yields a simple and appealing characteristics of N-defendable properties:

Theorem 12. *$\Gamma \models^m_{\mathrm{OptNE}} [N]\gamma$ iff $\gamma = Conv(\gamma)$, i.e., γ is closed under convex combination of strategies.*

Proof (sketch). "$\Rightarrow$" Analogous to point (ii) in the proof of Theorem 11.

"$\Leftarrow$" Consider any utility profile u. By the result of Nash [17], (Γ, u) has a Nash equilibrium in mixed strategies. Moreover, $\gamma = Conv(\gamma)$ implies that all the Nash equilibria s such that $dom(s) \not\subseteq \gamma$ are strongly Pareto-dominated by a mixed NE in γ. Hence, $\mathrm{OptNE}(\Gamma, u)$ is nonempty and entirely contained in γ.

Corollary 1. *$\Gamma \models^m_{\mathrm{OptNE}} [N]\gamma$ iff there exist subsets of individual strategies $\chi_1 \subseteq \Sigma_1, \ldots, \chi_{|N|} \subseteq \Sigma_{|N|}$ such that $\gamma = \chi_1 \times \cdots \times \chi_{|N|}$.*

That is, security property γ is defendable by the grand coalition in Γ iff γ can be *decomposed into constraints on individual behavior of particular agents.*

7 Conclusions

We propose a framework for analyzing security protocols (and other interaction protocols), that takes into account the incentives of agents. In particular, we consider a novel notion of *defendability* that guarantees that all the runs of the protocol are correct as long as a given subset of the participants (the "defenders") is in favor of the security property. We have obtained some characterization results for defendability under Nash equilibria and optimal Nash equilibria. We also studied the computational complexity of the corresponding decision problems, both in the generic case and in some special cases based on Nash equilibrium. In the future, we plan to combine our framework with results for protocol verification using game logics (such as ATL), especially for those solution concepts that can be expressed in that kind of logics.

Acknowledgements. Wojciech Jamroga acknowledges support of the National Research Fund Luxembourg (FNR) under project GaLOT – INTER/DFG/12/06.

References

1. R. Anderson, T. Moore, S. Nagaraja, and A. Ozment. Incentives and information security. In *Algorithmic Game Theory*. 2007.
2. G. Asharov, R. Canetti, and C. Hazay. Towards a game theoretic view of secure computation. In K. Paterson, editor, *EUROCRYPT*, volume 6632 of *Lecture Notes in Computer Science*, pages 426–445. Springer, 2011.
3. G. Asharov and Y. Lindell. Utility dependence in correct and fair rational secret sharing. In S. Halevi, editor, *CRYPTO*, volume 5677 of *Lecture Notes in Computer Science*, pages 559–576. Springer, 2009.
4. M Ben-Or, O Goldreich, S Micali, and R Rivest. A fair protocol for signing contracts. *IEEE Transactions on Information Theory*, IT-36(1):40–46, 1990.
5. L. Buttyán, J. Hubaux, and S. Čapkun. A formal model of rational exchange and its application to the analysis of Syverson's protocol. *Journal of Computer Security*, 12(3,4):551–587, May 2004.
6. R. Chadha, J. Mitchell, A. Scedrov, and V. Shmatikov. Contract signing, optimism and advantage. *Journal of Logic and Algebraic Programming*, 64(2):189–218, August 2005.
7. Krishnendu Chatterjee and Vishwanath Raman. Assume-guarantee synthesis for digital contract signing. *CoRR*, abs/1004.2697, 2010.
8. Y. Dodis and T. Rabin. Cryptography and game theory. In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors, *Algorithmic Game Theory*, chapter 8, pages 181–208. 2007.
9. B. Finkbeiner and S. Schewe. Coordination logic. In Anuj Dawar and Helmut Veith, editors, *CSL*, volume 6247 of *Lecture Notes in Computer Science*, pages 305–319. Springer, 2010.
10. G. Fuchsbauer, J. Katz, and D. Naccache. Efficient rational secret sharing in standard communication networks. In D. Micciancio, editor, *TCC*, volume 5978 of *Lecture Notes in Computer Science*, pages 419–436. Springer, 2010.
11. H. Ghaderi, H. Levesque, and Y. Lespérance. *A logical theory of coordination and joint ability.* ACM Press, New York, New York, USA, May 2007.
12. A. Groce and J. Katz. Fair Computation with Rational Players. In *EUROCRYPT*, pages 81–98, 2012.
13. J.Y. Halpern and N. Rong. Cooperative equilibrium (extended abstract). In *Proceedings of AAMAS 2010*, pages 1465–1466, 2010.
14. S. Kremer and J. Raskin. Game analysis of abuse-free contract signing. In *Proceedings of the 15th IEEE Computer Security Foundations Workshop (CSFW'02)*, pages 206–220. IEEE Computer Society Press, 2002.
15. S. Kremer and J. Raskin. A game-based verification of non-repudiation and fair exchange protocols. *Journal of Computer Security*, 11(3), 2003.
16. T. Moore and R. Anderson. Economics and internet security: a survey of recent analytical, empirical and behavioral research. Technical Report TR-03-11, Computer Science Group, Harvard University, 2011.
17. J. Nash. *Non-cooperative games.* PhD thesis, Princeton, 1950.
18. M. Osborne and A. Rubinstein. *A Course in Game Theory.* MIT Press, 1994.
19. P. Syverson. Weakly secret bit commitment: Applications to lotteries and fair exchange. In *CSFW*, pages 2–13, 1998.