Risk balance in optimistic non-repudiation protocols

Report**Author(s):**

Torabi Dashti, Muhammad; Cederquist, J.; Wang, Y.

Publication date:

2011

Permanent link:

<https://doi.org/10.3929/ethz-a-006804456>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Risk balance in optimistic non-repudiation protocols

M. Torabi Dashti¹, J. Cederquist², and Y. Wang³

¹ ETH Zürich, Universitätstrasse 6, CBN F 109, 8092 Switzerland

² Dep. Engenharia Informática, Instituto Superior Técnico, Universidade Técnica de Lisboa
SQIG, Instituto de Telecomunicações, Portugal

³ Department of Philosophy, Peking University, 100871 Beijing, China

Abstract. We investigate how the behaviors of malicious trusted parties affect participants of optimistic non-repudiation protocols. We introduce a notion of *risk balance* for exchange protocols. Intuitively, risk balance refers to fairness in the amount of protection a protocol offers to the participants against malicious trustees. We explore how risk balance relates to the notions of accountable trustees and transparent trustees previously introduced by Asokan and Micali, respectively. As a case study, we investigate the consequences of malicious behaviors of trusted parties in the context of two fair non-repudiation protocols, proposed by Gürgens, Rudolph and Vogt (2005). We discover a number of security issues in these protocols and propose simple solutions for fixing them.

Keywords. Optimistic fair exchange; Non-repudiation; Trust; Risk; Game theory

1 Introduction

Context. Fair exchange protocols in general deal with exchanging items in a *fair* manner. Intuitively, fair means that either all the protocol participants receive a desired item in exchange for their own, or none of them does so. It has been shown that no exchange protocol can achieve fairness if there is no presumed trust among the protocol participants. Asokan, Shoup and Waidner pioneered asynchronous *optimistic* fair exchange protocols [2]. The idea of optimistic fair exchange is that the participants first commit to release their items to the opponent, and after exchanging these commitments, they actually release the items. An off-line trusted third party (TTP) supervises the exchanges. If the participants correctly follow the protocol, the TTP does not intervene in the exchange. However, if an error (maliciously or accidentally) occurs, the wronged party can resort to the TTP, and by presenting the commitment of the opponent, recover to a fair state. Assuming that errors are infrequent (which is an optimistic assumption), the optimistic approach reduces the computation and communication costs of the TTP, which otherwise could become a bottleneck.

The information that is subject to exchange is constrained in optimistic protocols. Sandholm and Wang have shown that optimistic fair exchange is possible only if at least one of the exchanged items is either *generatable* or *revocable* [14]. In this paper we will only consider generatable items. Intuitively, an item is generatable if the TTP can generate the item, given the commitments of a protocol participant to release that item.

In this paper, we are concerned with a specific class of fair exchange protocols, namely two-party asynchronous optimistic *non-repudiation* protocols. A non-repudiation protocol ensures that a party cannot deny having sent or received a message, if it has actually done so in the course of the protocol. To achieve this, protocol participants usually collect evidences – evidence of origin and evidence of receipt – which can later be presented to a judge. A non-repudiation protocol is fair iff these evidences are exchanged in a fair manner. We distinguish four parties in a fair non-repudiation protocol: A and B , representing Alice and Bob who wish to exchange items and evidences, T playing the role of the TTP, and J playing the role of the external judge to whom the evidences are presented. Any non-repudiation protocol must in particular specify what constitutes an evidence of receipt and what constitutes an evidence of origin, and give an algorithm for J to verify if an evidence presented by a participant is valid.

Many protocols (including the ones we take as our case study) have exploited the evidence verification algorithm of J to make evidences generatable. To make an evidence generatable it is sufficient to allow T to generate a substitute for the evidence, and to count the substitute as a valid evidence for J . Enabling the TTP to generate substitute evidences is desirable as it reduces the amount of cryptographic machinery used in the protocol. This however comes with at least two drawbacks:

1. The protocol allows for the affidavits endorsed by the TTP to pass as valid evidences, hence enforcing specific formats for the evidences collected by the participants. The protocol thus cannot be used for generic fair exchange of *any* set of evidences, hence being *invasive* [2]. As a solution to invasiveness, TTPs in optimistic protocols can be made *transparent* [10]. Transparency of the TTP means that the outcome of the protocol (i.e. the evidences collected by the participants) does not depend on whether the TTP is involved in the exchange or not.
2. This gives extra freedom to *malicious* trustees, e.g. to issue substitute evidences for exchanges that never took place. In general, trusted parties are merely *trusted*, and need not be *trustworthy*. As a convention, we call a trusted party who deviates from the protocol a malicious trusted party, whether the party intentionally misbehaves, or this happens due to external reasons, e.g. being subverted by an attacker. In practice, the TTP may, e.g., have its own interests in the protocol, or it may collude with some of the participants to cheat on the others. To decrease the TTP's incentives to misbehave, the notion of *accountability* in fair exchange protocols has been introduced [2]. The trustee in a protocol is accountable if whenever the TTP misbehaves, its misbehavior can (under certain conditions) be proved to an external court. Accountability is thus a prohibition mechanism, relying on the assumption that a TTP prefers not being detected as malicious.

Asokan observes that in asynchronous optimistic protocols transparency and accountability cannot be achieved simultaneously [1]. We come back to this point later.

Contributions. How a malicious TTP may affect fair exchange protocols, which participants are more vulnerable to its actions, and methods to limit a malicious TTP's abilities are not well studied. A notable exception to this is the concept of accountable TTPs, mentioned above. Our work is a further step towards understanding the abilities of malicious TTPs and the methods to limit them.

Malicious TTPs can inevitably subvert an exchange protocol; this is indeed the definition of trusted parties. Therefore, when engaging in a protocol that requires TTPs a participant takes a *risk* which cannot be eliminated, namely the risk that the TTP is malicious and will act against the participant.

Here we give a definition of *risk balance* for optimistic exchange protocols. The aim is that a risk balanced protocol should offer about the same level of protection to its participants against malicious TTPs. Intuitively, a protocol is risk balanced if the amount of risk the protocol induces on the participants is (nearly) equal. Provided that the subjects of exchange are of the same “objective” value (which is typically the case for optimistic fair exchange protocols), in a risk balanced exchange protocols, the participants would thus incur the same amount of expected loss if the TTP is malicious.

We introduce *various notions of accountability*. In optimistic protocols, protection against malicious TTPs is tightly related to the level of accountability of TTPs. An accountable TTP, whose misbehaviors can be proved in an external court, is less likely to act against an agent *A* than a TTP who is not accountable. In this paper, we explore in more detail how risk balance relates to accountability, and, more generally, what each of these notions tell us about the security of a protocol, when the TTP may act maliciously.

We formalize exchange protocols in a game theoretical abstraction. Participants are considered as rational utility-maximizer agents who may follow the protocol honestly, or “compromise” the TTP by paying a certain cost. This abstract model in particular enables us to formally define the notion of risk balance.

We also define a *protocol selection game*, which is a framework for comparing security protocols. In this game, rational agents can choose between engaging in different protocols, and whether to play honest or not. By using the protocol selection game, we can (for instance) see that there are situations where agents favor risk balance more than accountability.

The notions of risk balance and various notions of accountability are demonstrated using a case study. We consider two protocols introduced by Gürgens, Rudolph and Vogt [8]: An optimistic fair non-repudiation protocol, and its adaptation to facilitate transparency of TTP. Through our analysis, we discover a number of security issues in these protocols and then propose simple solutions to fix them.

Related work. The notion of accountable TTPs has been introduced by Asokan [1], and transparent TTPs are introduced by Micali in [10]. In the context of the fair exchange protocols that require *online* TTPs (i.e. the TTP is directly involved in every exchange), Franklin and Reiter use a secret sharing scheme to limit what a TTP can learn about exchanged materials [7]. They assume that the participants cannot compromise the TTP, but the TTP may have its own interests in the matter. Previous analyzes of exchange protocol based on game theory assume trustees that cannot be compromised [5, 4, 9]. This is in contrast to the premise of our analysis: TTPs, by paying a cost, are compromiseable. An exception is our previous work on risk balance tailored for protocols with online TTPs [17]. Though the results of [17] are not directly applicable to optimistic protocols. To reduce the dependency of protocols on availability and sanity of a single TTP, distributed TTPs can be used [3, 13, 16]. Note that distributed TTPs in general need to ensure the consistency of their (distributed) state.

Structure of the paper. Section 2 gives an informal overview on optimistic fair exchange protocols. In section 3 we give an abstract model of two party exchange protocols based on strategic games. There, we also formalize the notion of risk balance for optimistic protocols. Two optimistic fair exchange protocols of [8] are briefly described in section 4. These protocols are analyzed in section 5 with respect to accountability and risk balance. The notion of protocol selection game is introduced in section 5. We conclude the paper in section 6. Proofs are relegated to an extended version of this paper.

2 Optimistic fair exchange

In this section, we give a high level description of a typical asynchronous optimistic fair exchange protocol; for concrete examples see section 4. Our goal is to explain the “logic” of these protocols without unnecessary cluttering. Therefore, we refrain from providing details. Optimistic protocols are comprehensively treated in [1].

We assume an asynchronous message passing communication system. The messages exchanged with the TTP are assumed to go over resilient channels, i.e. these messages eventually arrive at their destination. No such assumption is made about other channels. Below, we assume when a participant receives a message, it checks if the message conforms to the protocol. Receiving a bogus message is considered to be the same as not receiving it. An optimistic fair exchange protocol starts with A owning item $i(A)$ and B owning $i(B)$. A wishes to receive $i(B)$, and B wishes to receive $i(A)$. It is assumed that A knows a “sufficiently detailed” [1] description of $i(B)$ in order to verify whether a given item matches $i(B)$ that she has expected, or not. A similar assumption is made for B . Both A and B trust T as the TTP of the exchange.

Optimistic fair exchange protocols consist of three sub-protocols: An *optimistic* sub-protocol, a *resolve* sub-protocol, and an *abort* sub-protocol. In the optimistic sub-protocol, the participants, A and B , first exchange commitments $c(A)$ and $c(B)$, respectively, and then actually exchange their items $i(A)$ and $i(B)$. The sub-protocols, where A is the initiator and B the responder, are shown in figure 1. If no failure occurs, the optimistic sub-protocol (shown to the left in figure 1) ends successfully, with no TTP involvement. A participant who has $\{c(A), c(B)\}$ can initiate the resolve sub-protocol, shown in the center of figure 1, where $P \in \{A, B\}$. Ideally, this sub-protocol would be executed only if the participant has received the commitment, but not the item, from its opponent.

Fig. 1. Sub-protocols: (Left) optimistic, (Center) resolve $P \in \{A, B\}$, (Right) abort.

The TTP has a secure persistent database. Upon receiving $c(A)$ and $c(B)$, the TTP stores a fingerprint of this exchange in the database, and tags the exchange as being *resolved* (see remark 1 below). If $P = B$, TTP extracts $i(A)$ from these commitments, and sends it back to P as part of $i(T)$ (similarly, for the case $P = A$). Extracting $i(A)$ from $c(A)$ often introduces heavy cryptographic machinery in the protocol. Therefore, in many protocols, and particularly in our case studies, $i(T)$ merely consists in an affidavit stating that: T testifies that A has sent an item described by $c(A)$ to B . Naturally, $c(A)$ is assumed to uniquely described $i(A)$.

The abort sub-protocol is meant to guarantee termination. Note that after A sends $c(A)$, she has to wait for B to reply, and cannot meanwhile resort to the TTP using the resolve sub-protocol (she needs $c(B)$ for that). This is unfair because B has the possibility to continue the optimistic protocol, resolve or do nothing at all. To avoid this problem, A can execute the abort sub-protocol, shown in the right exchange of figure 1. If the TTP responds with `abort.token` to A , then it would store a fingerprint of the exchange in its database, and tag it as *aborted*.

Remark 1. A few remarks are due:

1. In the resolve sub-protocol, the TTP performs a check so that previously aborted exchanges will not be resolved; instead an abort token will be sent back. This ensures that the TTP's reactions are consistent.
2. In case the TTP needs $c(B)$ to compute $i(T)$, it has to store it along with the fingerprint of the exchange in the resolve sub-protocol. This is because if A tries to abort an exchange which has already been resolved by B , the TTP needs to send $i(T)$ to A , but the abort request from A does not contain $c(B)$.
3. After a successful exchange, A can send an abort request and receive an abort token from the TTP. This abort token therefore has no evidential value: it is a promise by the TTP stipulating that the TTP never resolves this exchange.

An exchange protocol has to specify what are the subjects of exchange. If no failure occurs, A gets $i(B)$ and B gets $i(A)$. These are thus the basic subjects of exchange. As a convention we refer to subjects of exchange as *evidences* Ev and denote these by $Ev(A) = i(B)$ and $Ev(B) = i(A)$. If the TTP is involved, then $i(T)$ too is often considered as a valuable token. That is, $Ev(A) = i(B) \mid i(T)$, and $Ev(B) = i(A) \mid i(T)$ ($\mid$ denotes 'or'). A protocol is said to be *transparent* for A if $i(B)$ is the only subject of exchange from A 's point of view, even if T is involved in the exchange, see [10]. For this, A (or T) must be able to extract $i(B)$ from $i(T)$. Transparency for B is symmetrically defined. To achieve transparency one can often embed $i(A)$ and $i(B)$ inside $c(A)$ and $c(B)$, respectively, encrypted for the TTP. As a side note, a protocol that is transparent from A 's point of view does not necessarily guarantee that A cannot prove the involvement of T in the protocol to an outsider; this means that an outsider cannot infer whether T was involved in the protocol or not by solely inspecting the subject of exchange collected by A ; cf. *abuse-freeness* [5].

Optimistic fair exchange protocols usually aim at the following basic goals:

- (Non-triviality) If no failure occurs, and communication channels are reliable, then A and B can successfully exchange their items, with no TTP involvement.

- (Fairness) If A receives her desired subject of exchange $Ev(A)$ from B or T , then if B follows the protocol he also receives his desired subject of exchange $Ev(B)$ from A or T , and vice versa.
- (Timeliness) No honest participant can be forced to wait indefinitely.

Intuitively, the optimistic sub-protocol serves non-triviality. Fairness is achieved by the logic of optimistic sub-protocol, and also consistent behavior of the TTP. To achieve timeliness, A and B are provided with resolve and abort sub-protocols. Timeliness can be seen as fairness in being able to unilaterally terminate the exchange.

The TTP is accountable from A 's point of view iff whenever the TTP misbehaves such that A loses fairness, then A can prove this in a court, possibly external to the protocol. A specific way for the TTP to misbehave is to ignore all resolve and abort request from a participant. For checking accountability however it is assumed that the TTP is forced to answer to any valid resolve or abort request [1]. The intuition is that if the TTP cheats A by giving $i(T)$ to B , but sending an abort token to A , then this is detectable in an external court: it is against the protocol for the TTP to be inconsistent.

Note that in case the protocol is transparent for B , then, intuitively, $i(T)$ contains $i(A)$. Therefore, the TTP can send $i(T)$ to B and the abort token to A , and claim in the external court that A herself has sent $i(A)$ (which B obtains from $i(T)$) to B , while TTP has sent abort tokens to both A and B . Accountability of the TTP is indeed adversely related to transparency of the protocol [1]: the TTP cannot be held accountable for its actions if its involvement is not recognizable.

3 Game abstraction for exchange protocols

From a game theoretical point of view, a two-party exchange protocol with a compromiseable TTP can be seen as a two-party strategic game, in which the agents can either follow the protocol faithfully or compromise the TTP. If both parties play faithfully, then they normally “earn” the goods from the opponent and “lose” their own goods. However, when engaging in the exchange, each agent has to take some risk due to the fact that the opponent may compromise the TTP. In such cases, the agent who compromises the TTP can earn the amount that the other (honest) party risks, and lose only the cost of compromising the TTP. We omit basics of game theory, and instead refer the reader to [11]. Formally, we have the following game abstraction.

Definition 1. (Protocol game) Given a two-party exchange protocol Prot with a TTP, the strategic game $G(\text{Prot})$ is defined as follows:

$A \backslash B$	$\mathcal{H}_B$	$\mathcal{D}_B$
$\mathcal{H}_A$	$g_B^A - g_A^A, g_A^B - g_B^B$	$-r_A^A, r_A^B - c_B$
$\mathcal{D}_A$	$r_B^A - c_A, -r_B^B$	$r_B^A - r_A^A - c_A, r_A^B - r_B^B - c_B$

Here, $\mathcal{H}_x$ is the strategy of an agent x that conforms to the protocol; $\mathcal{D}_x$ is the strategy of x in which x compromises the TTP and may stop following the normal course of the

protocol when it has to release its goods; g_x^y is y 's evaluation of the goods that x wants to exchange; r_x^y is y 's evaluation of the risk that x has if the TTP is compromised by the opponent of x ; and c_x is the expected cost x pays to compromise the TTP.

Below we assume:

- Agents have incentives to exchange goods: $g_x^y > g_y^y$ if $x \neq y$. For simplicity, we assume that there is a fixed exchange rate $\rho > 1$ such that $g_x^y = \rho g_x^x$, $x \neq y$.
- The risks of the agents comply with the same exchange rate: $r_x^y = \rho r_x^x$, $x \neq y$.
- The subjective values of the goods are the same: $g_A^A = g_B^B = g > 0$.

Then, $G(\text{Prot})$ can be simplified to $SG(\text{Prot})$ below; here $a = r_A^A$ and $b = r_B^B$.

$A \setminus B$	$\mathcal{H}_B$	$\mathcal{D}_B$
$\mathcal{H}_A$	$(\rho - 1)g, (\rho - 1)g$	$-a, \rho a - c_B$
$\mathcal{D}_A$	$\rho b - c_A, -b$	$\rho b - a - c_A, \rho a - b - c_B$

To apply game theoretical analysis, we assume that the agents are rational utility-maximizers. A strategy profile is a joint strategy that determines a unique utility pair; for example $(\mathcal{H}_A, \mathcal{H}_B)$ is a strategy profile while $((\rho - 1)g, (\rho - 1)g)$ is the corresponding utility pair. A strategy profile (S_A, S_B) is called a *Nash equilibrium* if no agent gets a higher utility by switching to another strategy, given the strategy of the other agent according to the profile. In this paper, we consider the Nash equilibria of a simplified protocol game as the expected executions of the corresponding protocol by rational agents. We write $\text{Utility}_x(S_A, S_B)$ as the utility of x if the agents select the strategy profile (S_A, S_B) , and define $\Delta_U(S_A, S_B) = |\text{Utility}_A(S_A, S_B) - \text{Utility}_B(S_A, S_B)|$.

3.1 Risk balance

We study the protocol game $SG(\text{Prot})$ under the assumption that the objective values of the exchanged items are the same, i.e. $a = b$, while the cost of compromising the TTP might be different for A and B , i.e. $c_A \neq c_B$. Intuitively, the amount of protection for A against malicious TTPs is proportional to c_B in the protocol game. Similarly, the amount of protection for B against malicious TTPs is proportional to c_A .

The difference between c_A and c_B can in particular be due to accountability: Suppose in an exchange protocol, the TTP is accountable to B , but not accountable from A 's point of view. If B compromises the TTP, their illegitimate collaborate cannot (easily) be detected. Therefore, the collusion between TTP and B is of low risk for them. However, if A compromises the TTP, they can be caught red handed by an external court, since the TTP is accountable to B . That is, the collusion between the TTP and A is of high risk for them. Thus, if the TTP is accountable to B , but not to A , then $c_A > c_B$.

Definition 2. Protocol Prot is called risk balanced iff $|c_A - c_B| < (\rho - 1)g$ in the protocol game $SG(\text{Prot})$.

This definition puts a bound on $|c_A - c_B|$, which intuitively reflects the difference between the risks A and B take to participate in the protocol. Does this definition of risk balanced correspond to balance in the losses the protocol participants incur when the TTP is malicious? The following theorem answers this question positively, by showing that rational agents in such protocols have the same incentives with respect to compromising the TTP, and the damages a malicious TTP can cause for the participants is nearly equal.

Theorem 1. *Under the assumption $a = b = g$, for any risk balanced protocol Prot , there are Nash equilibria in $SG(\text{Prot})$. These equilibria can only occur in the profiles $(\mathcal{H}_A, \mathcal{H}_B)$ and $(\mathcal{D}_A, \mathcal{D}_B)$. Moreover, for each such Nash equilibrium (S_A, S_B) , $\Delta_U(S_A, S_B) \leq |c_A - c_B| < (\rho - 1)g$ holds.*

The theorem intuitively states that the difference between participants' expected utilities is bounded by the difference between the protection the protocol provides for the participants against malicious TTPs. The limit on utility differences turns out to be independent of the exact costs of compromising the TTP. This is a desirable property since it ensures that no matter how trustworthy the TTP might be in an execution, the difference between participants' utilities is limited to a value independent of c_A and c_B ; hence none of the participants would hugely benefit by compromising the TTP, compared to the other one.

The converse of Theorem 1 does not hold. But it can be shown that, if c_A (c_B) is small enough and c_B (c_A) big enough, then the risks are indeed un-balanced:

Theorem 2. *If $c_A \leq g$ and $c_B \geq \rho g$ then $(\mathcal{D}_A, \mathcal{H}_B)$ is a Nash equilibrium. Similarly, if $c_B \leq g$ and $c_A \geq \rho g$ then $(\mathcal{D}_B, \mathcal{H}_A)$ is a Nash equilibrium.*

In our game abstraction, the participants are allowed to play dishonest at the same time. This corresponds to the case where both parties gain access to TTP's machine. Limiting this situation to the cases where only one of the participants can subvert the TTP for arbitrary long periods of time would affect our results. We leave further research in this direction for future work. Note that the misbehaviors of the participants when they both subvert the TTP cannot be detected by any external court since the TTP is not deviating from the protocol; it is rather following a particular flow of the protocol, e.g. through the resolve sub-protocol. Nevertheless, the expected costs of compromising the TTP is not changed for the profile $(\mathcal{D}_A, \mathcal{D}_B)$. This reflects the fact that participants are in general ignorant of whether their opponent tries to subvert the TTP, or not.

4 Protocols of Gürgens, Rudolph and Vogt

We briefly describe the protocols of [8]. These will serve as our case studies in this paper. Details which are not of importance to our analysis are pruned away. A thorough presentation can be found in the original text.

As in [8], we assume ideal cryptography à la Dolev and Yao [6]. Our analysis is however not bound to this assumption. We write m_K for the encryption of message m with symmetric key K . Process P can decrypt m_K only if K is known to P . Similarly, P can compose m_K from m only using K . We assume a deployed secure public

key infrastructure. The notation $\{m\}_P$, with P being the name of a principal, represents the encryption of message m with the public key of P . The signature of P on m is denoted $[m]_P$. As usual, m, m' stands for concatenation of messages m and m' . The principals also have access to a secure hash function $\mathcal{H}$.

Protocol I. This protocol consists of three sub-protocols. We first describe the optimistic sub-protocol. In this protocol A wants to send message (e.g. a certified email) M to B , and expects a receipt. Conversely, B expects to receive M and an evidence that A has indeed sent M to B . Below, A starts by choosing a random key K , and constructing the label $L = \mathcal{H}(A, B, T, \mathcal{H}(M_K), \mathcal{H}(K))$:

1. $A \rightarrow B : A, B, T, M_K, \mathcal{H}(K), \{f_1, L, K\}_T, [f_2, L, \{f_1, L, K\}_T]_A$
2. $B \rightarrow A : [f_3, L, \{f_1, L, K\}_T]_B$
3. $A \rightarrow B : K, [f_4, L, K]_A$
4. $B \rightarrow A : [f_5, L, K]_B$

Here each f_i is a unique flag to determine the intention of the accompanying terms. If no failure occurs, the optimistic sub-protocol will finish successfully, with no TTP intervention. In case A does not receive message 4 after sending message 3, or B does not receive message 3, they can run the resolve sub-protocol (below $P \in \{A, B\}$):

- 1r. $P \rightarrow T : A, B, \mathcal{H}(M_K), \mathcal{H}(K), \{f_1, L, K\}_T, [f_2, L, \{f_1, L, K\}_T]_A, [f_3, L, \{f_1, L, K\}_T]_B$
- 2r. $T \rightarrow P : [f_{\text{aborted}}, L]_T$
- or
- 2'r. $T \rightarrow P : K, \text{con}, \quad \text{where con} = [f_6, L, K]_T$

What T sends back depends on whether the protocol has been aborted before or not. If it has not been aborted, then T sends back 2'r, and stores $(L, \text{resolved}, K, \text{con})$ in its persistent secure database. However, if the exchange has been aborted, T sends back message 2r. The TTP can verify whether a message has been resolved or aborted previously by simply searching for the fingerprint of the exchange (i.e. L) in its database. In case a message is not valid (e.g. the signatures are not correct), T sends back an error message. We assume that these error messages are entangled with the messages they refer to, and thus cannot be replayed by malicious participants.

Now we turn to the abort sub-protocol. If A does not receive message 2 in the optimistic sub-protocol in a reasonable amount of time (which A unilaterally determines), A starts the abort sub-protocol:

- 1a. $A \rightarrow T : A, B, \mathcal{H}(M_K), \mathcal{H}(K), [f_{\text{abort}}, L]_A,$
- 2a. $T \rightarrow A : \text{con}, \quad \text{where con} = [f_6, L, K]_T$
- or
- 2'a. $T \rightarrow A : [f_{\text{aborted}}, L]_T$

In case the exchange has been resolved before, T sends back message 2a. Otherwise, T sends back message 2'a, and stores $(L, \text{aborted}, [f_{\text{aborted}}, L]_T)$ in its database.

For A , the set $\{A, B, T, M, M_K, K, e_r\}$ would constitute a receipt that B has received M , where $e_r = [f_5, L, K]_B$ or $e_r = \text{con}$. That is, these two possible message sets

constitute $Ev(A)$. Similarly, for B , the set $\{A, B, T, M, M_K, K, e_o\}$ would constitute an evidence that A has sent M to B , where $e_o = [f_4, L, K]_A$ or $e_o = \text{con}$.

Protocol II. This protocol is exactly like protocol I, except for the following changes:

1. All instances of $\{f_1, L, K\}_T$ in protocol I are replaced with $\{f_1, L, K, [f_4, L, K]_A\}_T$. Note that the evidence obtained by B in case the TTP is not involved, is now embedded for the TTP in the message.
2. In the resolve sub-protocol, instead of con , T sends $[f_4, L, K]_A$ to B . This message is computable for T , since message 1 contains $[f_4, L, K]_A$, encrypted for T .
3. For B , the set that constitutes an evidence that A has sent M to B is

$$\{A, B, T, M, M_K, K, [f_4, L, K]_A\}.$$

According to [8], the purpose of this variant is to provide transparency for B . Indeed, from B 's point of view what constitutes $Ev(B)$ in protocol II does not depend on whether T was involved in the exchange, or not.

5 Analysis

In this section we analyze the protocols I and II, described in the previous section, with respect to accountability and risk balance. Fairness and timeliness for these protocols have been studied in [8].

Our analysis in this section is qualitative, rather than quantitative. That is, we do not assign actual values to the parameters of the protocol games (defined in section 3). Our goal is here to understand *comparative* merits of protocols in terms of the risk they induce on the participants. We then use our game-based formal model to explain how risk balance affects choices of rational participants in different protocols. The main hypothesis of our risk analysis is that the probability that a malicious TTP would act against A is adversely related to the amount of protection the protocol provides for A against malicious TTPs. We take the level of accountability of the TTPs in a protocol as the main component of the protection the participants have against malicious TTPs. To carry out our qualitative risk analysis, we give a linear order on different levels of accountability below.

Protocol I.

We start with studying the TTP's accountability. Below, $A + T$ describes the scenario where A compromises (or, colludes with) T , and B remains honest. Similarly is defined $B + T$.

Accountability: Case $A + T$. We note that A and T can ignore B entirely. This is because in con there is no message component contributed by B . Therefore, T can construct con without B being aware that an "exchange" is going on. This is a notable shortcoming in the protocol, since with the trustee's help A can choose an arbitrary

agent $\mathcal{X}$ and an arbitrary message M and produce “evidences” showing that $\mathcal{X}$ has received M from A . Furthermore, for generating such evidences the trustee and A do not need to be involved in any exchange with the wronged parties (here, B). That is, $\mathcal{X}$ can be potentially off-line. Any prudent court would therefore treat the evidences produced in this protocol with skepticism. This is a flaw in the protocol far more serious than unfairness or unaccountability of the trustee.

Now suppose that B wants to prove T ’s misbehavior in an external court. Obviously, this can only happen after B has learned that A has an evidence proving that B has received a message from A , while in fact B has not done so. In this case B cannot send a valid resolve request to T , because B does not own A ’s commitment to the exchange (see the resolve sub-protocol in section 4). Therefore, T is not forced to reply to B (see assumption for checking accountability, in the end of section 2). That is, T ’s misbehavior cannot be proved by B . Therefore, T is not accountable from B ’s point of view.

In case A sends (or, as A would call it: *resend*) message 1 of the optimistic sub-protocol to B , then B can send a valid resolve request to T . If T replies with K , con to B , then B attains fairness. However, if T decides to reply with an abort token, its misbehavior can be detected by the court: The court would observe that T has provided A with con (this is clear by the evidences A has presented to the court) and has sent an abort token to B . This inconsistent behavior violates the protocol for T . Note that T ’s signature in con is the key for detecting T ’s misbehaviors. We conclude that T is accountable. However, as accountability of T from B ’s point of view is achieved only when A collaborates with B , we refer to this level of accountability as *discretionary*.

Accountability: Case $B + T$. Two scenarios for subverting A are conceivable in this case. (1) A starts an exchange, intending to send a message M to B , but B and T do not provide A with suitable replies, and (2) B fabricates a message and with T ’s help produces the evidence that A has sent M to B . We only consider scenario (2) here, which poses a more compelling threat. Indeed, scenario (2) is identical to $A + T$.

Now, suppose in an external court A wants to prove T ’s misbehavior. Obviously, this can only happen after A has learned that B has an evidence proving that A has sent a message M to B , while in fact A has not done so. In this case A cannot send a valid resolve or abort request to T , because A does not own M and K to construct these messages. Therefore, according to the definition of accountability T is not forced to reply to A . That is, from A ’s point of view, T is not accountable.

We however note that M and K can be constructed from the information B has to collect in his evidence; thus the court can provide A with these messages merely by inspecting the evidences B presents. If the court does so, A can send a valid request to T ; then, it is straightforward that T is accountable from A ’s point of view. However, since the court’s help is necessary here we refer to this level of accountability as *weak*; cf. the similar concept of *weak* fairness in [12].

We order the levels of the trustee’s accountability as: “discretionary accountability” is weaker than “weak accountability”, which is in turn weaker than “accountability”. This ordering is not arbitrary. In discretionary accountability, the trustee is accountable to, say, A only if A ’s opponent provides A with certain messages. That is, A relies on her opponent to prove the TTP’s misbehaviors. In weak accountability, however, A

relies on the (supposedly neutral) external court to provide her with certain messages in order to be able to prove TTP's misbehaviors to the very same court. It is reasonable to take weak accountability as a stronger guarantee for A compared to discretionary accountability. In the case of accountability as defined in section 2, A does not need any help from other parties to prove TTP's misbehaviors to the court. This ordering helps us in our qualitative risk analysis below.

Risk balance. We contend that in protocol I, compromising T is cheaper for A than B . This is because, when A compromises T , from B 's point of view T is accountable only at the level of discretionary accountability. However, when B compromises T , from A 's point of view T is accountable at the level of weak accountability (see discretionary vs. weak accountability above). Therefore, $c_A < c_B$ (see section 3.1). Now, assume the value of items subject to exchange (i.e. evidences) is the same from A and B 's viewpoint, and equals g , and let ρ be the exchange rate (see section 3.1). If $c_B - c_A \geq (\rho - 1)g$ then the protocol is not risk balanced, by definition. If, in addition, $c_A \leq g$ and $c_B \geq \rho g$ then a rational A has incentives to compromise T , while B has not; cf. theorem 2. Then the damage caused to B when T colludes with A is not proportional to the damage $B + T$ can cause for A .

Protocol II.

Accountability Below, $A + T$ describes the scenario where A colludes with T , and B remains honest. Similarly is defined $B + T$.

- Case $A + T$: Exactly similar to the case of protocol I: From B 's point of view T is accountable in this protocol only at the level of discretionary accountability.
- Case $B + T$: In this case, T can construct $K, [f_4, L, K]_A$ for B , once A has sent message 1 of the optimistic protocol out (destined to B). Now, if B does not reply message 1, A will abort the exchange, and T can safely send an abort token to A . In this scenario, neither T nor B can be accused of violating the protocol; see the discussion on transparency versus accountability in section 2. Therefore, from A 's point of view, T is not accountable in this protocol.

Risk balance. We contend that B can compromise T in a cheaper way, compared to A . This is because, if B compromises T , then A cannot prove T 's misbehavior to an external court. While from B 's point of view T is accountable only at the level of discretionary accountability. Consequently $c_A > c_B$. Similar to the argument for protocol I, if the difference between c_A and c_B is big enough, protocol II is not risk balanced.

Fixing the accountability problems.

The fact that the TTP is not accountable from A 's point of view in protocol II is a direct result of the TTP being transparent from B 's point of view. Therefore, without forgoing the point of protocol II (which is to provide TTP transparency for B) the accountability issue for A cannot be fixed. The situation is however different for protocol I: The key observation to solve the accountability issues in protocol I reported above is to augment

evidences with messages produced by both originator A and recipient B . Towards this goal, we update the definition of con to

$$\text{con} = [f_6, L, K]_T, [f_2, L, \{f_1, L, K\}_T]_A, [f_3, L, \{f_1, L, K\}_T]_B$$

Note that this change affects the evidences collected by A and B in the protocol. We refer to the resulting protocol as protocol I_f .

The new con does not negatively affect the fairness and timeliness of protocol I. However, it reduces the power of malicious trustees. This however comes at the cost of the trustee storing further information about each disputed exchange in its database; recall that for resolved exchanges the trustee needs to store con along with the fingerprint of the exchange in its persistent database. The trustee indeed needs to be stateful in asynchronous optimistic protocols [15]; see also item 2 in remark 1.

Comparing the protocols.

To limit the damages malicious TTPs can cause, our analysis implies that protocol I is preferable over protocol II. Risk balance however does not approve any of these two protocols: They are both not risk balanced. If protocol I is strengthened to protocol I_f as described above, then clearly protocol I_f would be preferable over protocols I and II, by both the accountability and the risk balance criteria (from the transparency point of view, however, protocol II would be preferable over protocols I and I_f ; transparency is however not a security concern per se). An interesting conflict arises in case we are given a (fictitious) third protocol, protocol III, in which the TTP is accountable from both A and B 's points of view only at the discretionary level. According to accountability, protocol I is preferable over protocol III. Risk balance would however advise us to choose protocol III over protocol I. Whether accountability or risk balance is chosen as the overriding criterion depends on the setting in which the protocols are going to be deployed, as described below.

Under the assumption that A and B are provided only with the protocol we choose for them, then it is perhaps reasonable to prioritize the accountability criterion and choose protocol I over protocol III. This is because in absolute terms protocol I provides more protection against malicious TTPs, compared to protocol III. However, if A and B are assumed to be rational agents who can execute a variety of protocols, then we argue that it is reasonable to accept the risk balance criterion, i.e. choose protocol III over protocol I. Observe that B has the same guarantee over accountability of TTP in both protocol I and protocol III, namely discretionary accountability. However, B has a disadvantage compared to A when engaging in protocol I: He knows that he always has less protection against a malicious TTP, compared to A . Therefore, B needs to take more risk when playing protocol I, in comparison to A . Whereas, in protocol III the situation is fair: A and B take the same risk when executing the protocol. A rational B would therefore have an inclination towards protocol III. From A 's point of view however protocol I is preferable; nonetheless protocol III is not "unfair" to A .

This informal argument is supported by the following *protocol selection game*. We prove that if rational A and B choose to execute protocol I, then at least one execution of protocol III is also a Nash equilibrium. That is, whenever executing protocol I is a

rational choice for A and B , then so is executing protocol III. The converse however does not hold: There are situations where executing protocol III is a rational choice for A and B , while executing protocol I is not. Protocol III is thus preferable over protocol I from the point of view of feasibility. This is intuitively because there exist exchanges whose payoff is not “sufficiently high” (w.r.t. the induced risk) for B to execute protocol I, while A would be willing to execute protocol I. Then, obviously they will not exchange at all. However, given the same situation, if A and B have the choice to execute protocol III, then they would both engage in the exchange. It is worth noting that which protocol A and B will finally execute in general depends on the variety of the protocols that are available to them, and their payoffs in successful exchanges.

Protocol selection game. Let us assume that for any X and Y , when they execute protocol I or protocol III, compromising the TTP costs c (with $c > 0$) for X , if the level of the TTP’s accountability from Y ’s point of view is discretionary. However, if the TTP is weakly accountable from Y ’s point of view, then compromising the TTP would cost $c + \delta$ (with $\delta > 0$) for X . Following the assumptions of our analysis in section 3.1, we get the game matrix shown in table 1, which models strategies and utilities of A and B when selecting a protocol and choosing between playing honest and dishonest.

$A \setminus B$	PI- $\mathcal{H}_B$	PI- $\mathcal{D}_B$	PIII- $\mathcal{H}_B$	PIII- $\mathcal{D}_B$
PI- $\mathcal{H}_A$	$\rho g - g, \rho g - g$	$-g, \rho g - c - \delta$	$0, 0$	$0, 0$
PI- $\mathcal{D}_A$	$\rho g - c, -g$	$\rho g - g - c, \rho g - g - c - \delta$	$0, 0$	$0, 0$
PIII- $\mathcal{H}_A$	$0, 0$	$0, 0$	$\rho g - g, \rho g - g$	$-g, \rho g - c$
PIII- $\mathcal{D}_A$	$0, 0$	$0, 0$	$\rho g - c, -g$	$\rho g - g - c, \rho g - g - c$

Table 1. A protocol selection game

Note that in the protocol selection game, players can explicitly choose not to engage in any exchange by selecting a protocol different from their opponent. This is indeed reflected in the game matrix (all ‘0, 0’ utilities).

Theorem 3. *For any ρ, c, δ, g such that a profile is a Nash equilibrium for protocol I, that profile is also a Nash equilibrium for protocol III. Moreover, there exist ρ, c, δ, g such that there is a Nash equilibrium for protocol III, but none for protocol I.*

This theorem indicates that whenever two rational agents execute protocol I, they would also execute protocol III. The converse however does not hold: There are situations in which two rational agents would engage in protocol III, but not in protocol I.

6 Concluding remarks

We have studied the notions of accountability and risk balance in optimistic fair exchange protocols. First, a game-theoretical definition of risk balance is given. The basic notion of fairness in exchange has already been extended to timeliness (i.e. fairness in

being able to finalize the exchange) and abuse-freeness (i.e. fairness in being able to prove to an outsider that the participant can force the outcome of the exchange: successful, or unsuccessful). Risk balance, intuitively, refers to fairness in the amount of risk the participants take in a protocol. Second, two protocols designed by Gürgens, Rudolph and Vogt [8] are analyzed with respect to accountability and risk balance. We have also pointed out security concerns in these protocols, and proposed solutions to fix the problems.

Accountability strives to minimize the risk the participants take when engaging in a protocol with a potentially malicious trusted party. Risk balance however aims at distributing the risks of the participants evenly. These criteria may clash. We define a protocol selection game and show that, under certain assumptions, a risk balance protocol is more “feasible” than a protocol which induces less, but unevenly distributed, risks. We believe that protocol selection games can give a general framework for formally comparing security protocols beyond the simple secure versus insecure dichotomy. Further investigations in this direction is left for future work.

References

1. N. Asokan. *Fairness in electronic commerce*. PhD thesis, University of Waterloo, 1998.
2. N. Asokan, V. Shoup, and M. Waidner. Asynchronous protocols for optimistic fair exchange. In *IEEE Security and Privacy '98*, pages 86–99. IEEE CS, 1998.
3. G. Ateniese, B. de Medeiros, and M. Goodrich. TRICERT: A distributed certified e-mail scheme. In *NDSS '01*. Internet Society, February 2001.
4. L. Buttyán, J. Hubaux, and S. Capkun. A formal model of rational exchange and its application to the analysis of Syverson’s protocol. *J. Computer Security*, 12(3-4):551–587, 2004.
5. R. Chadha, J. Mitchell, A. Scedrov, and V. Shmatikov. Contract signing, optimism, and advantage. In *CONCUR '03*, volume 2761 of *LNCS*, pages 366–382. Springer, 2003.
6. D. Dolev and A. Yao. On the security of public key protocols. *IEEE Trans. on Information Theory*, IT-29(2):198–208, 1983.
7. M. Franklin and M. Reiter. Fair exchange with a semi-trusted third party (extended abstract). In *ACM CCS '97*, pages 1–5. ACM Press, 1997.
8. S. Gürgens, C. Rudolph, and H. Vogt. On the security of fair non-repudiation protocols. *Int. J. Inf. Sec.*, 4(4):253–262, 2005.
9. K. Imamoto, J. Zhou, and K. Sakurai. An evenhanded certified email system for contract signing. In *ICICS '05*, volume 3783 of *LNCS*, pages 1–13. Springer, 2005.
10. S. Micali. Simple and fast optimistic protocols for fair electronic exchange. In *PODC '03*, pages 12–19. ACM Press, 2003.
11. M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1999.
12. H. Pagnia, H. Vogt, and F. Gärtner. Fair exchange. *The Computer Journal*, 46(1):55–7, 2003.
13. I. Ray, I. Ray, and N. Natarajan. An anonymous and failure resilient fair-exchange e-commerce protocol. *Decision Support Systems*, 39(3):267–292, 2005.
14. T. Sandholm and X. Wang. (Im)possibility of safe exchange mechanism design. In *8th national conference on artificial intelligence*, pages 338–344. AAAI, 2002.
15. M. Schunter. *Optimistic fair exchange*. PhD thesis, Universität des Saarlandes, 2000.
16. M. Srivatsa, L. Xiong, and L. Liu. ExchangeGuard: A distributed protocol for electronic fair-exchange. In *IPDPS '05*, page 105b. IEEE CS, 2005.
17. M. Torabi Dashti and Y. Wang. Risk balance in exchange protocols. In *Asian '07*, volume 4846 of *LNCS*, pages 70–77. Springer, 2007.