# Towards a Game-Theoretic Security Analysis of Off-Chain Protocols

Sophie Rain D *TU Wien*, Austria Georgia Avarikioti *TU Wien*, Austria Laura Kovács D *TU Wien*, Austria Matteo Maffei Christian Doppler Lab Blockchain Technologies for the Internet of Things TU Wien, Austria

Abstract—Off-chain protocols constitute one of the most promising approaches to solve the inherent scalability issue of blockchain technologies. The core idea is to let parties transact on-chain only once to establish a channel between them, leveraging later on the resulting channel paths to perform arbitrarily many peer-to-peer transactions off-chain. While significant progress has been made in terms of proof techniques for off-chain protocols, existing approaches do not capture the game-theoretic incentives at the core of their design, which led to overlooking significant attack vectors like the Wormhole attack in the past.

In this work we take a first step towards a principled gametheoretic security analysis of off-chain protocols by introducing the first game-theoretic model that is expressive enough to reason about their security. We advocate the use of Extensive Form Games (EFGs) and introduce two instances of EFGs to capture security properties of the closing and the routing of the Lightning Network. Specifically, we model the closing protocol, which relies on punishment mechanisms to disincentivize parties to upload old channel states on-chain. Moreover, we model the routing protocol, thereby formally characterizing the Wormhole attack, a vulnerability that undermines the fee-based incentive mechanism underlying the Lightning Network.

This work is the extended version of our CSF 2023 paper "Towards a Game-Theoretic Security Analysis of Off-Chain Protocols".

*Index Terms*—game theory, off-chain protocols, security, rational players, Lightning Network

#### I. INTRODUCTION

Blockchain technologies are emerging as a revolutionary paradigm to perform secure decentralized financial applications. Nevertheless, a widespread adoption of cryptocurrencies, such as Bitcoin [1] and Ethereum [2], is severely hindered by their inherent limitations on transaction throughput [3], [4]. For instance, while Bitcoin can support tens of transactions per second and the confirmation time is about an hour, traditional credit networks like Visa can comfortably handle up to 47,000 transactions per second.

Off-chain protocols [5] are recognized as one of the most promising scalability solutions, achieving a seemingly contradictory property: the bulk of transactions is performed offchain, and yet in a secure fashion. The idea is to leverage the blockchain only in case of disputes, resorting otherwise to offchain, peer-to-peer transactions. Bitcoin's Lightning Network [6] is the most widely adopted off-chain instantiation, hosting at the time of writing bitcoins worth more than 170M USD, in a total of more than 27,000 nodes and more than 76,000 channels. In a nutshell, parties deposit money in a shared address, called channel, and can later on perform arbitrarily many off-chain transactions with each other by redistributing the deposit on the channel. In the end, the channel can be closed and the latest state (i.e., deposit distribution) is posted on-chain. Off-chain transactions are not limited to the endpoint of the channel, but they can be routed over paths of channels (so-called multi-hop payments). Besides such payment channel networks, an entire ecosystem of off-chain protocols [5] (virtual channels, watchtowers, payment-channel hubs, state channels, side-chains, etc.) is under development for Bitcoin [7]–[12], Ethereum [13]–[16], as well as other cryptocurrencies [17].

The cryptographic protocols underlying these off-chain constructions are rather sophisticated and, most importantly, rely on game-theoretic arguments to discourage malicious behavior. For instance, the Lightning Network relies on a punishment mechanism to disincentivize parties to publish old states onchain and on an unlocking mechanism where parties first pay a neighbor and then retrieve the paid amount from the other to ensure the atomicity of multi-hop payments (i.e., either all channels are consistently updated or none is).

Off-chain protocols are typically subject to rigorous security analyses, which however concentrate on cryptographic properties and do not capture the game-theoretic ones. In particular, most protocols are proven secure in the Universal Composability framework [18], proving that the cryptographic realization simulates the ideal functionality. This framework, however, was developed to reason about security in the classical honest/Byzantine setting: in particular, the ideal functionality has to model all possible parties' behavior, rational and irrational, otherwise it would not be simulatable, but reasoning on whether or not certain behavior is rational is outside of the model and thus left to informal arguments. This is not just a theoretical issue, but a practical one, as there is the risk to let attacks pass undetected: for instance, the Wormhole attack [7] constitutes a rational behavior in the Lightning Network, which is thus admitted in any faithful model thereof although it undermines its incentive mechanism. The first step towards closing this gap in cryptographic proofs is to come up with a faithful game-theoretic model for off-chain protocols in order to reason about security in the presence of rational parties. We address this challenge in this paper, advocating the use of Extensive Form Games (EFGs) for the game-theoretic security analysis of off-chain protocols. In particular, we introduce two

instances of EFGs to model the closing and the routing of the Lightning Network.

# A. Related Work

A game-theoretic model for off-chain protocols is initiated and introduced in [19]. This work suffers, however, from several limitations, which make it unsuitable to conduct faithful security analyses. Firstly, the game model considers only honest closing of channels, i.e., all deviations - such as posting an old state - are ignored: this makes it impossible to reason about the security of basic channel operations. Secondly, the pay-offs are represented as constants, which neglects the dependency of the channel's balance on its security properties. Further, fees are not considered at all, thereby ignoring their impact on Lightning protocols. For instance, the routing game to model the security of multi-hop payments fails to capture already identified attacks in payment channel networks, like the Wormhole attack [7] that targets the fee distribution among players. Additionally, Lightning is vulnerable to the Griefing attack [20], where a significant amount of money is locked. In our work, we overcome the aforementioned limitations, by defining a stronger closing phase model, by aligning the utilities to the monetary outcome, by considering all possible deviations of parties during closing, and by revising the relevant security properties. We demonstrate the importance of precision in game-theoretic protocol models by modeling the Wormhole attack, as well as the Griefing attack.

Our work further complements other game-theoretic advancements in the area, most prominently the following lines of research.

*a)* Incentivizing Watchtowers: A major drawback of payment channel protocols is that channel participants must frequently be online and watch the blockchain to prevent cheating. To alleviate this issue, the parties can employ third parties, or so-called watchtowers, to act on their behalf in case their counterparty misbehaves. Correctly aligning the incentives of watchtowers to yield a secure payment channel protocol is, however, challenging. This is the main focus of several works [11], [15], [16], [21]. As their objective is to incentivize external parties, their models does not apply in our work.

b) Payment Channel Network Creation Games: Avarikioti et al. [22], [23] study payment channel networks as network creation games. Their goal is to determine which channels a rational node should establish to maximize its profit. Ersoy et al. [24] undertake a similar task; they formulate the same problem as an optimization problem, show it is NP-hard, and present a greedy algorithm to approximate it. Similarly to our work, all these works assume rational participants. However, we aim to model the security of the protocols, in contrast to these works that study the network creation problem graphtheoretically.

c) Blockchains with Rational Players: Blockchains incentivize miners to participate in the network via monetary rewards [1]. Therefore, analyzing blockchains under the lens of rational participants is critical for the security of the consensus layer. There are multiple works in this direction: Badertscher et al. [25] present a rational analysis of the Bitcoin protocol. Eyal and Garay [26] introduce an attack on the Nakamoto consensus, effectively demonstrating that rational miners will not faithfully follow the Bitcoin protocol. This attack is generalized in [27], [28]. Consequently, Kiayias et al. [29] analyze how miners can deviate from the protocol to optimize their expected outcome. Later, Chen et al. [30] investigate the reward allocation schemes in longest-chain protocols and identify Bitcoin's allocation rule as the only one that satisfies a specific set of desired properties. On a different note, several works study the dynamics of mining pools from a gametheoretic perspective [31], [32] or introduce network attacks that may increase the profit of rational miners [33], [34]. An overview of game-theoretic works on blockchain protocols can be found in [35].

All these works, however, focus on the consensus layer (Layer-1) of blockchains and as both the goals and assumptions are different from the application layer (Layer-2), the models introduced there cannot be employed for our purposes. For instance, payment channel protocols occur off-chain and thus game-based cryptographic assumptions of the blockchain do not apply. In addition, consensus protocols investigate the expected reward of miners which is a probabilistic problem, whereas we ask if any honest player could lose money, which depends on the behavior of the other players and is fundamentally deterministic.

Game-based definitions have also been proposed for the security analysis of smart contracts [36], [37]. These models, however, target an on-chain setting and are thus not suitable to reason about the specifics of off-chain constructions (e.g., closing games, routing games, etc.).

## B. Our Contributions

In this work, we take the first steps towards closing the gap between security and game-theoretic analysis of off-chain protocols. Specifically, we introduce the first game-theoretic models that are expressive enough to reason about the security of off-chain protocols. We model off-chain protocols as games and then analyze whether or not certain security properties are satisfied. The design of our models is driven by two principles: (a) all possible actions should be represented and (b) the utility function should mirror the monetary outcome realistically. We aim to ensure that honest participants do not suffer any damage (P1), whereas deviating from the protocol vields a worse outcome for the adversary (P2) We will use weak immunity (Definition II.4) to implement (P1), and collusion resilience (Definition III.8) together with practicality (Definition III.7) for (P2). While we believe that our approach of implementing principles (a) and (b) is easily extensible to other off-chain protocols, in this work we focus on the Bitcoin Lightning Network, which constitutes the most widely adopted off-chain protocol. Our technical contributions can be summarized as follows:

- We refine existing game-theoretical concepts in order to reason about the security of off-chain protocols (Section III).
- We introduce the Closing Game  $G_c$ , the first gametheoretic security model that accurately captures the closing phase of Lightning channels, encapsulating arbitrary deviations from the protocol specification (Section IV).
- We perform a detailed security analysis of  $G_c$ , formalize folklore security corner cases of Lightning, and present the strategy that rational parties should follow to close their channels in order to maximize their expected outcome relative to the current and previous distribution states (Section V).
- We identify limitations in prior work [19] on gamebased modeling of multi-hop payments, putting forward a new game-based definition that is precise enough to cover the Wormhole and the Griefing attack (Section VI). We further show how to model Fulgor protocol [38], a variant of Lightning's routing that prevents the Wormhole attack. Our formalization leverages game theory concepts introduced in Section III and Section IV, thereby demonstrating the theoretical expressiveness of our framework to analyze complex protocols.

In conclusion, our work brings game-theoretical foundations to enforce security of off-chain protocols, by providing a rigorous analysis over security properties expressed through formal requirements over game strategies. We believe, the provided rigor in our paper opens up new venues for automating security analysis via game-theoretic arguments, a challenge which we aim to tackle in future work.<sup>1</sup>

## II. BACKGROUND AND PRELIMINARIES

#### A. Payment Channel Networks

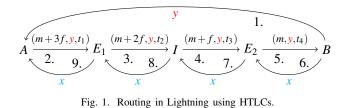
A payment channel [8] can be seen as an escrow (or multisignature), into which two parties Alice A and Bob B transfer their initial coins with the guarantee that their coins are not locked forever and the agreed balance can be withdrawn at any time. After that, A and B can pay each other off-chain by signing and exchanging messages that reflect the updated balances in the escrow. These signatures can be used at any time to close the channel and distribute the coins on-chain according to the last channel state. In order to discourage parties from posting an old state on-chain, a punishment mechanism is in place. In particular, in Lightning [6], once A closes the channel, she has to wait a mutually agreed time before getting her coins. Meanwhile, B has the opportunity to withdraw all the coins in the channel (by posting a so-called revocation transaction), including the ones assigned to A, if the state posted on-chain by A is not the last one they mutually agreed on. Such a punishment mechanism is of game-theoretic nature: parties can indeed post an old state on-chain, yet they are discouraged to do so.

In particular, Lightning payment channels operate as follows: First, Alice and Bob create a funding transaction where they input their respective coins; the funding transaction has a single output that can only be spent if both A and Bprovide their signature (2-out-of-2 multi-signature). Then, the two parties create the first commitment transaction, i.e., a transaction that spends the output of the funding transaction and returns the initial coins to both parties. In other words, the input of the commitment transaction is the output of the funding transaction while the output of the first commitment transaction is two-fold: the first output returns the coins to A and the second output to B. However, the commitment transaction each party holds is not the same. Specifically, the commitment transaction of A has an additional spending condition, a timelock t that signifies the revocation period and is pre-agreed between the two parties; in A's commitment transaction B's output is spendable immediately. Symmetrically, in B's commitment transaction B's output has a timelock t while A's output is spendable immediately. Note that a timelock t is a condition that allows the coins of the output to be spent on-chain only after time t has elapsed from the publication of the transaction. After A and B sign and exchange the respective first commitment transactions, they proceed to signing the funding transaction and publishing it on-chain. This order is important to avoid hostage situations<sup>2</sup>. As soon as the funding transaction is securely published on-chain, A and B can transact off-chain by creating every time a new commitment transaction that depicts the current balance of the joint capital among the two parties. Every time a new commitment transaction is created, the parties reveal a secret to their counterparty that allows their counterparty to spend their own coins immediately (e.g., A can spend B's coins from the previous commitment) if the previous commitment transaction appears on chain (revocation transaction). To close a Lightning channel, the two parties can either collaborate and spend the output of the funding transaction, or each of them can close the channel unilaterally by publishing the last commitment transaction. Since the commitment transactions each party hold have a timelock, in case of cheating, i.e., publication of a previous commitment transaction on-chain, the counterpart can immediately spend the cheating party's coins, claiming all the coins of the channel, thus punish the cheating party for misbehaving.

Technically, A and B do not just lock their initial funds but also a certain small amount which will be used as a fee for the closing transaction of the channel. Note that every onchain transaction requires such a fee f. The fee for the opening transaction is paid upon the opening of the channel and is thus irrelevant to our consideration. However, in case A posts an old state on-chain and B performs the revocation transaction – which is an on-chain transaction – to prove it, B has to carry the additional transaction fee alone. These facts have an important impact on our game-theoretic models.

<sup>&</sup>lt;sup>1</sup>We refer the interested reader to the appendix for the complete definitions and proofs.

 $<sup>^{2}</sup>$ If the funding transaction is published on-chain before the first commitment transactions are signed, a party may hold the other hostage since none of the parties can close the channel unilaterally but only in collaboration.



In the following, we refer to *honest closing* when a party unilaterally closes the channel by posting the last commitment transaction or when the parties close collaboratively, where both parties sign to spend the funding transaction output directly.

Off-chain transactions are not limited to the end-points of a channel, as they can be performed whenever sender and receiver are connected by a path of channels with enough capacity. The cryptographic approach to do so exploits hashtime-locked-contracts (HTLC) [12]. Assume players A and B do not share a channel. Instead, A has a channel with  $E_1$ ;  $E_1$ has a channel with I; I has a channel with  $E_2$ ; and  $E_2$  has a channel with B, as illustrated in Figure 1. Player A can now send an amount m to player B via the intermediaries  $E_1$ , I and  $E_2$ , where each intermediary charges a fee f for the routing service, hence A should pay m+3f. The core idea is that A pays  $E_1$ ,  $E_1$  pays I, and so forth, until B gets paid.

A key security property in multi-hop payments is atomicity: either all payments are successful, and the deposit in each channel is updated accordingly, or none is. To achieve this property, the Lighting protocol proceeds as follows. First, the receiver B generates a secret x and sends its hash h(x) = y to the sender A (see action 1 in Figure 1). Then A creates an HTLC for  $E_1$ , where she locks m + 3f with lock y and timeout  $t_1$ . That means only  $E_1$  can claim the money and only by providing a value whose hash is y within time  $t_1$  (action 2 in Figure 1). Although  $E_1$  does not know such a value yet and can therefore not unlock,  $E_1$  can nevertheless proceed by creating another HTLC for I also locked with y and a timeout  $t_2$  (action 3 in Figure 1). Thereafter, I and  $E_2$  continue in the same way (actions 4 - 5 in Figure 1). Actions 1 - 5of Figure 1 are called the *locking phase*. Note that in order to allow everybody to unlock their HTLCs in the subsequent steps, the time-outs have to be decreasing  $t_1 > t_2 > t_3 > t_4$ . Once B receives the conditional payment, he can reveal x to  $E_2$ and the conditional payment is unlocked (action 6 in Figure 1). The others can now unlock the HTLCs one after the other from right to left (actions 7 - 9 in Figure 1), which is called the unlocking phase. Finally, A paid m + 3f, B received m and each intermediary was rewarded with f.

We note that atomicity is achieved by a game-theoretic argument: intermediaries can, in principle, stop the protocol either in the locking phase or in the unlocking phase. In the former, they would lose the transaction fee f, while in the latter, they would lose the payment amount m, m + f, m + 2f respectively. Thus, they are incentivized to act once they have committed to participate.

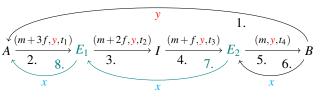


Fig. 2. Wormhole Attack in Lightning.

a) The Wormhole Attack: The aforementioned routing protocol is proven to be vulnerable to the Wormhole attack [7], which is depicted in Figure 2. The attack is as follows:  $E_1$  and  $E_2$  collude, and bypass I in the unlocking phase, thus stealing I's participation reward f. Until actions 6 in Figure 1 and Figure 2, the behavior is identical. Then,  $E_2$ , knowing x, forwards x to  $E_1$  (offline) instead of unlocking the HTLC from I (action 7 in Figure 2). This way,  $E_1$  can unlock A's HTLC and claim the money (action 8), but I will never be able to unlock. After a certain time the remaining HTLCs time-out and the locked money returns to the creators.

Therefore, the parties A and B are not affected. However,  $E_1$  and  $E_2$  collectively earn 3f instead of the 2f they deserve, stealing the fee f from I, who locked resources in the locking phase of the protocol. This attack undermines the incentive of intermediaries to route payments.

b) The Griefing Attack: It describes the scenario when a player, assume *B* for simplicity, ignores the proposed payment and refuses to proceed [20]. This way, money is locked in the conditional payments for a considerable amount of time. While [39] studies the Griefing attack through probabilistic modelling and [40] provides mitigation techniques, to the best of our knowledge there is no formal security analysis of this attack at present. Our work addresses this limitation and shows that Lightning's routing module is indeed susceptible to the Griefing attack.

In the sequel we consider the behavior as illustrated in Figure 1 as the only *honest routing* behavior.

#### B. Game-Theoretic Definitions

We now introduce the game-theoretic concepts relevant for our formalization. We denote real numbers by  $\mathbb{R}$  and tuples as  $\sigma = (\sigma_1, ..., \sigma_n)$ . We write  $\sigma[\sigma'_i/\sigma_i]$  to denote the tuple resulting from substituting  $\sigma_i$  by  $\sigma'_i$  in  $\sigma$ , that is  $\sigma[\sigma'_i/\sigma_i] =$  $(\sigma_1, ..., \sigma_{i-1}, \sigma'_i, \sigma_{i+1}, ..., \sigma_n)$ . We understand games as static objects in which finitely many players can choose finitely many times from a finite set of actions. A game yields a certain positive or negative utility for each player. We briefly overview the very common Normal Form Games, also called Strategic Games [41], in which each player chooses an action *only once*, called strategy.

**Definition II.1** (Normal Form Game – NFG). *A* Normal Form Game (NFG) *is a tuple*  $\Gamma = (N, \mathcal{S}, u)$ , *where N is the set of game* players,  $\mathcal{S} = X_{p \in N} \mathcal{S}_p$  *the set of* joint strategies  $\sigma$  *and u the* utility function:

•  $\mathscr{S}_p$  is the non-empty set of strategies player p can choose from. Thus, a joint strategy  $\sigma \in \mathscr{S}$  is a tuple of strategies  $\sigma = (\sigma_{p_1}, ..., \sigma_{p_{|\mathcal{N}|}})$ , with  $\sigma_{p_i} \in \mathscr{S}_{p_i}$ .

TABLE I
NFG $\Gamma_C$ with players $A, B$

B A	U	С	J
U	(1/2, 1/2)	(0,1)	(0,1)
С	(1,0)	(1,1)	(-1, -1)
I	(1,0)	(-1, -1)	(-1, -1)

*u* = (*u*<sub>p1</sub>,...,*u*<sub>pn</sub>), where *u*<sub>pi</sub> : *S* → ℝ assigns player *p<sub>i</sub>* its utility for every joint strategy *σ* ∈ *S*.

In what follows we fix an arbitrary game  $\Gamma$  and give all definitions relative to it. To formalize an optimal outcome on game strategies, we use Nash Equilibria.

**Definition II.2** (Nash Equilibrium – NE). A Nash Equilibrium is a joint strategy  $\sigma \in \mathscr{S}$  s.t. no player  $p_i$  can increase their utility by unilaterally deviating from  $\sigma = (\sigma_{p_1}, ..., \sigma_{p_{|N|}})$ , Formally,

$$\forall p \in N \,\forall \sigma'_p \in \mathscr{S}_p : \, u_p(\sigma) \ge u_p(\sigma[\sigma'_p/\sigma_p]) \,. \tag{1}$$

Another important concept is *weakly dominated strategies*, expressing the strategies a rational player would not play since they yield worse utilities.

**Definition II.3** (Weakly Dominated Strategy). A strategy  $\sigma_p^d \in \mathscr{S}_p$  of player p is called weakly dominated by strategy  $\sigma_p' \in \mathscr{S}_p$ , if it always yields a utility at most as good as  $\sigma_p'$  and a strictly worse utility at least once:

$$\forall \, \boldsymbol{\sigma} \in \mathscr{S} : \, u_p(\, \boldsymbol{\sigma}[\boldsymbol{\sigma}_p^d/\boldsymbol{\sigma}_p]\,) \le u_p(\, \boldsymbol{\sigma}[\boldsymbol{\sigma}_p^\prime/\boldsymbol{\sigma}_p]\,) \, and \qquad (2)$$

$$\exists \sigma \in \mathscr{S} : u_p(\sigma[\sigma_p^d/\sigma_p]) < u_p(\sigma[\sigma_p^\prime/\sigma_p]).$$
(3)

**Example II.1.** Consider the NFG  $\Gamma_C$  in Table I, which was introduced in [19] to model closing. In this game  $\Gamma_C$ , there are two players  $N = \{A, B\}$  and each players can choose from the same strategy set  $\mathscr{S}_A = \mathscr{S}_B = \{U, C, \Im\}$ . Here, strategy U captures unilateral closing, that is publishing the latest state on-chain. Further, strategy C corresponds for closing collaboratively, that is publishing a mutually signed transaction. Finally, strategy  $\Im$  stands for ignoring, that is doing nothing. The utility for each joint strategy is given in Table I, where player A's strategies are listed in the left column of Table I.

Applying Definition II.2, the joint strategies (C,C),  $(U,\mathfrak{I})$ and  $(\mathfrak{I},U)$  are Nash Equilibria: for each of these joint strategies, neither A nor B can deviate in order to increase their own utility. Comparing the second and the third row of Table I, we see that A's utility is always as least as good in the second row as it is in the third row. Hence, strategy C weakly dominates strategy  $\mathfrak{I}$  for player A, by Definition II.3; the same property also holds for player B. By comparing the other pairs of rows/columns of Table I, we see that there is no other weak dominance in  $\Gamma_C$ .

### C. Game-Theoretic Security Properties of Off-Chain Protocols

We now present existing game-theoretic concepts [19], [41] implying security properties of off-chain protocols. In Section III, we extend these concepts towards another type of games, called Extensive Form Games, enabling our security analysis in Section IV. We focus on two security properties ensuring that (P1) honest players do not suffer damage, and (P2) subgroups of rational players do not deviate from a respective strategy. A protocol is compliant to these properties, if the strategy implementing the intended behavior satisfies them; we call such a strategy an *honest strategy*.

(P1) No Honest Loss. As the utility function of a game is supposed to display the monetary and intrinsic value of a certain joint strategy, property (P1) is expressed using weak immune strategies defined next.

**Definition II.4** (Weak Immunity). A joint strategy  $\sigma \in \mathscr{S}$  in an NFG  $\Gamma$  is called weak immune, if every player p that follows  $\sigma$  gets utility at least 0, regardless of how the other players behave:

$$\forall p \in N \ \forall \sigma' \in \mathscr{S} : \quad u_p(\sigma'[\sigma_p/\sigma'_p]) \ge 0.$$
(4)

**Example II.2.** In the game  $\Gamma_C$  of Table I, the only weakly immune strategy is (U,U). This is the case, because as long as A chooses U, player B can take any strategy and A will never get negative utility (similarly, vice-versa).

(P2) No Deviation. Even though the concept of Nash Equilibria seems to be a good candidate to ensure (P2) at first glance, they have two crucial shortcomings. First, a Nash Equilibrium only ensures that a single player cannot profit from deviating, but does not imply that two or more players cannot do so. Second, there might be Nash Equilibria, which are weakly dominated by another strategy for a specific player. Such Nash Equilibria will therefore not be played by rational parties and hence should not be considered to satisfy (P2).

The solution proposed for NFGs in [19] is to consider strategies  $\sigma$  compliant to (P2), if they are both *strongly resilient* (fixing the former shortcoming) and *practical* (fixing the latter) as defined subsequently.

Strong resilience extends Nash Equilibria by considering deviations of multiple players.

**Definition II.5** (Strong Resilience – SR). A joint strategy  $\sigma \in \mathscr{S}$  in an NFG  $\Gamma$  is strongly resilient (SR) if no proper subgroup of players  $S := \{s_1, ..., s_j\}$  has an incentive in deviating:

$$\forall S \subset N \quad \forall \sigma'_{s_i} \in \mathscr{S}_{s_i} \quad \forall p \in S : u_p(\sigma) \ge u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1},...,\sigma'_{s_i}/\sigma_{s_i}]) .$$
 (5)

We note that in games with two players (i.e. two-player games), strong resilience and Nash Equilibrium are identical. As such, in  $\Gamma_C$  from Table I, the joint strategies (C,C),  $(U,\mathfrak{I})$  and  $(\mathfrak{I},U)$  of Example II.1 are also strongly resilient.

To define practicality of a strategy, we first introduce the concept of *iterated deletion of weakly dominated strategies* (*IDWDS*).

TABLE II NFG  $\Gamma'_C$  obtained from IDWDS over Table I.

B A	U	С
U	(1/2, 1/2)	(0,1)
С	(1,0)	(1,1)

Definition II.6 (Iterated Deletion of Weakly Dominated Strategies - IDWDS). The iterated deletion of weakly dominated strategies (IDWDS) of an NFG  $\Gamma$  is defined as iteratively rewriting  $\Gamma$  by omitting all weakly dominated strategies of all players. This is repeated until no strategy is weakly dominated any more. The resulting game  $\Gamma'$  is thus a subgame of  $\Gamma$ .

Note that, when IDWDS is applied to a game  $\Gamma$ , then every Nash Equilibrium of the resulting game  $\Gamma'$  is also a Nash Equilibrium of  $\Gamma$ . Since all weakly dominated strategies of every player are removed at each step, the generated game is unique. Details and proofs can be found in [41].

We now define practical strategies, in order to ensure that no single strategy is weakly dominated at any iteration.

Definition II.7 (Practicality). A strategy is practical if it is a Nash Equilibrium of the NFG  $\Gamma'$  after iterated deletion of weakly dominated strategies.

**Example II.3.** Let us consider  $\Gamma_C$  from Table I. We know from Example II.1 that only  $\Im$  is weakly dominated for both A and B. Therefore, according to Definition II.6, strategy I has to be removed from both player's strategy set. This yields the game  $\Gamma'_C$  as listed Table II.

Note that there are no weakly dominated strategies in  $\Gamma'_C$ . Thus, any Nash Equilibrium of  $\Gamma'_C$  is also practical strategy of  $\Gamma_C$ . By comparing utilities, we derive that the only Nash Equilibrium of  $\Gamma'_C$  is the joint strategy (C,C).

An alternative approach for expressing (P2) is by requiring a strategy  $\sigma$  to be both a strong Nash Equilibrium (a property similar to SR) and practical, instead of SR and practical.

Definition II.8 (Strong Nash Equilibrium - sNE). A joint strategy  $\sigma$  is a strong Nash Equilibrium (SNE) if for every group of deviating players  $S := \{s_1, ..., s_i\}$  and all possible deviations  $\sigma'_{s_i} \in \mathscr{S}_{s_i}$ ,  $i \in \{1, ..., j\}$  at least one player  $p \in S$ has no incentive to participate, that is

$$\forall S \subseteq N, S \neq \emptyset \quad \forall \sigma_{s_i}' \in \mathscr{S}_{s_i} \quad \exists p \in S : \\ u_p(\sigma) \ge u_p(\sigma[\sigma_{s_1}'/\sigma_{s_1}, ..., \sigma_{s_i}'/\sigma_{s_i}]).$$
(6)

**Example II.4.** In  $\Gamma_C$  from Table I, all NE are also sNE. For the joint strategy (C,C), this is easy to see. However, it is also the case for  $(U,\mathfrak{I})$  and  $(\mathfrak{I},U)$ , since any deviation yields a utility of at most 1. Thus, at least one player's utility does not increase by deviating from  $(U,\mathfrak{I})$ ,  $(\mathfrak{I},U)$  respectively.

A detailed comparison of the various concepts ensuring (P2) including their strengths and weaknesses, is given in Section III.

# III. EFG-BASED MODELING OF OFF-CHAIN PROTOCOLS

So far we considered games in which each party takes only one action. We now extend our definitions to handle adaptive strategies, i.e., games in which parties take several actions and choose at each step which action to take based on the actions previously chosen by other parties. As we will see, this is necessary for faithfully modeling off-chain protocols and overcoming the limitations of previous work [19]. For that, we overview the concept of extensive form games (EFGs) in Section III-A. We show how to lift NFG-based security definitions to EFGs in Section III-B. Finally, we show that these definitions do not yet suffice to yield an accurate security model of off-chain protocols, and introduce a refined security definition based on the concept of collusion resilience in Section III-C.

#### A. Extensive Form Games (EFG)

To formalize strategies where players make multiple choices one after the other, we advocate the usage of Extensive Form Games (EFGs) [41], which extend NFGs as follows.

**Definition III.1** (Extensive Form Game – EFG). An Extensive Form Game (EFG) is a tuple  $\Gamma = (N, \mathcal{H}, P, u)$ , where N and u are as in NFGs. The set  $\mathcal{H}$  captures game histories,  $\mathcal{T} \subseteq \mathcal{H}$ is the set of terminal histories, and P denotes the next player function, satisfying the following properties.

- The set  $\mathscr{H}$  of histories is a set of sequences of actions with
- 1)  $\emptyset \in \mathscr{H};$
- 2) if the action sequence  $(a_k)_{k=1}^K \in \mathscr{H}$  and L < K, then
- also  $(a_k)_{k=1}^L \in \mathscr{H}$ ; 3) a history is terminal  $(a_k)_{k=1}^K \in \mathscr{T}$ , if there is no action  $a_{K+1}$  with  $(a_k)_{k=1}^{K+1} \in \mathscr{H}$ .
- The next player function P
- 1) assigns the next player  $p \in N$  to every non-terminal history  $(a_k)_{k=1}^K \in \mathscr{H} \setminus \mathscr{T}$ , that is  $P((a_k)_{k=1}^K) = p$ ;
- 2) after a non-terminal history  $h = (a_k)_{k=1}^K \in \mathscr{H}$ , it is player P(h)'s turn to choose an action from the action set  $A(h) = \{a : (h,a) \in \mathcal{H}\}.$

A strategy of player p is a function  $\sigma_p$  mapping every  $h \in \mathcal{H}$ with P(h) = p to an action from A(h). Formally,

$$\sigma_p: \{h \in \mathscr{H} : P(h) = p\} \to \{a: (h,a) \in \mathscr{H}, \forall h \in \mathscr{H}\},\$$

such that  $\sigma_p(h) \in A(h)$ . The set of all strategies of a player p is  $\mathscr{S}_p$ , and the set of all joint strategies is  $\mathscr{S} = X_{p \in \mathbb{N}} \mathscr{S}_p$ .

Note that the set of terminal histories  $\mathcal T$  is uniquely determined by  $\mathscr{H}$  and therefore does not explicitly occur in the tuple  $\Gamma$ . Since histories h are just sequences of actions  $h = (a_k)_{k=1}^K = (a_1, ..., a_K)$ , we denote histories by the variable h, the abstract sequence  $(a_k)_{k=1}^K$ , or the explicit sequence  $(a_1,...,a_K)$ , depending on the context in which they are used. We note that EFGs can conveniently be represented as trees, as described below.

**Definition III.2** (EFG as Tree). Considering an EFG  $\Gamma$  =  $(N, \mathcal{H}, P, u)$ , the following tree G = (V, E) represents  $\Gamma$ .

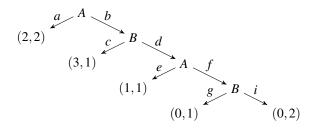


Fig. 3. An EFG  $\Gamma_E$ .

- For every history  $h \in \mathcal{H}$ , there exists exactly one node  $v_h \in V$ . This is labeled by P(h), the next player, if h is not terminal  $(h \notin \mathcal{T})$ , or by  $u(\sigma)$ , the joint utility of playing a game with history h, if h is terminal  $(h \in \mathcal{T})$  and the joint strategy  $\sigma$  yields history h.
- Two nodes  $v_h$ ,  $v_{h'} \in \mathscr{H}$  are connected via an oriented edge  $(v_h, v_{h'}) \in E$  iff h' = (h, a). This edge is labeled a.

Let us illustrate EFGs and their tree-based representation through the following example.

**Example III.1.** The game tree in Figure 3 results from the extensive form game  $\Gamma_E = (N, \mathcal{H}, P, u)$  with the two players  $N = \{A, B\}$ , where the set of histories is  $\mathscr{H} = \{\emptyset, (a), (b), (b, c), (b, d), (b, d, e), (b, d, f), (b, d, f, g), \}$  $\{(b,d,f,i)\}$ . The next player function P assigns player A after histories  $\emptyset$  and (b,d), and player B after (b) and (b,d,f). Finally, the utility function u assigns joint utility (2,2) to strategies that yield history (a), utility (3,1) for strategies with history (b,c), utility (1,1) for strategies with history (b,d,e), and (0,2) for strategies resulting in (b,d,f,i). A strategy  $\sigma = (\sigma_A, \sigma_B)$  in  $\Gamma_E$  is for example: A chooses a after history  $\emptyset$ :  $\sigma_A(\emptyset) = a$ ; and f after (b,d):  $\sigma_A((b,d)) = f$ ; B takes c after (b):  $\sigma_B((b)) = c$ , and g after (b,d,f):  $\sigma_B((b,d,f)) = g$ . Following this strategy until we read a leaf yields history (a). A different strategy  $\sigma' = (\sigma'_A, \sigma'_B)$ , which also yields history (a), is for example  $\sigma'_A(\emptyset) = a$ ,  $\sigma'_A((b,d)) = e$  and  $\sigma'_B = \sigma_B$ .

As depicted in the tree-based representation of Figure 3, we note that the utility of joint strategies in an EFG is uniquely determined by their associated history (i.e., path). In the context of EFGs, the concept of *Nash Equilibria* remains as given in Definition II.2. In addition to Nash Equilibria, another useful concept for EFGs is the *Subgame Perfect Equilibrium*, which we will use to characterize the strategies played in practice by rational parties. To this end, we first introduce the notion of subgames of EFGs. A subgame of an EFGs can be seen as a subtree determined by a certain history (i.e., whose root note is the last history node), and is formalized below.

**Definition III.3** (Subgame of EFG). The subgame of an EFG  $\Gamma = (N, \mathcal{H}, P, u)$  associated to history  $h \in \mathcal{H}$  is the EFG  $\Gamma(h) = (N, \mathcal{H}_{|h}, P_{|h}, u_{|h})$  defined as follows:  $\mathcal{H}_{|h} := \{h' \mid (h,h') \in \mathcal{H}\}, P_{|h}(h') := P(h,h'), and u_{|h}(h') := u(h,h').$ 

**Example III.2.** Consider the EFG  $\Gamma_E$  from Figure 3. The subgame of  $\Gamma_E$  associated to history (b,d) is the subtree rooted in A.

By adjusting the concept of Nash Equilibrium to subgames, we derive the following property of joint strategies.

**Definition III.4** (Subgame Perfect Equilibrium). A subgame perfect equilibrium *is a joint strategy*  $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_n) \in \mathscr{S}$ , *s.t.*  $\sigma_{|h} = (\sigma_{1|h}, ..., \sigma_{n|h})$  *is a Nash Equilibrium of the subgame*  $\Gamma(h)$ , for every  $h \in \mathscr{H}$ . The strategies  $\sigma_{i|h}$  are functions that map every  $h' \in \mathscr{H}_{|h}$  with  $P_{|h}(h') = i$  to an action from  $A_{|h}(h')$ .

## B. EFG Extensions for Security Properties

While EFGs enable us to incorporate choices made at different times yielding different options for the next player, they come with the following limitation. The intended (i.e., honest) behaviors in off-chain protocols only specify a terminal history (i.e., a path from root to leaf), rather than a strategy. For instance, an honest history may specify to close the channel collaboratively, but it does not capture a player's behavior once a player deviated. To address this limitation, we introduce the following notion of an extended strategy in EFGs.

**Definition III.5** (Extended Strategy). Let  $\beta$  be a terminal history in an EFG  $\Gamma$ . Then, all strategies  $\sigma_{\beta}$  that result in history  $\beta$  are extended strategies of  $\beta$ .

**Example III.3.** Recall Figure 3. In Example III.1, we consider the terminal history (a) and provide two extended strategies of (a), they are  $\sigma$  and  $\sigma'$ . A strategy, which is not an extended strategy of (a) is for instance  $\sigma'' = (\sigma''_A, \sigma''_B)$ , where  $\sigma''_A(\emptyset) =$ b,  $\sigma''_A((b,d)) = e$  and  $\sigma''_B = \sigma_B$ . This is the case because by following the choices of A and B in  $\sigma''$ , we end up in (b,c).

While EFGs can in principle be translated to NFGs, as explained in [41], analyzing the security properties (P1)-(P2) over the translated NFGs may yield unexpected results. We shortly exemplify this point in Example III.4, but similar issues occur also in larger games. We thus lift NFG-based definitions to EFGs, enabling the analysis of (P1) and (P2). Since EFGs have a utility function just as NFGs do, which assigns values after the game, the NFG concepts of weak immunity, strong resilience and **sNE** remain the same for EFGs.

**Definition III.6** (EFG Properties). A joint strategy  $\sigma \in \mathcal{H}$ of an EFG  $\Gamma$  is called weak immune, strongly resilient, or a strong Nash Equilibrium, if it satisfies the formulae of Definition II.4, Definition II.5 or Definition II.8 respectively.

Practicality in NFGs, however, relies on IDWDS, which fails to incorporate the sequential nature of EFGs, and hence must be adjusted for EFGs. This is because NFG actions happen simultaneously, while EFG players choose their actions sequentially. We first present an example to showcase that applying the NFG definition of practicality to an EFG, by using its translation to an NFG, leads to overlooking rational strategies.

**Example III.4.** Let us consider the EFG  $\Gamma_E$  from Figure 3, with two players A and B. The compact translation of  $\Gamma_E$  to an NFG  $\Gamma_N$  is given in Table III. Histories of Figure 3, where players choose twice, such as (b,d,f), are translated

TABLE III COMPACT VIEW OF  $\Gamma_E,$  TRANSLATED TO AN NFG  $\Gamma_N.$ 

B	С	d;g	d;i
а	(2,2)	(2,2)	(2,2)
b;e	(3,1)	(1,1)	(1,1)
b;f	(3,1)	(0,1)	(0,2)

to Table III as the joint strategy (b; f, d). Hence, the NFG strategy b; f of player A means choosing action b first, and, if A gets to choose again, A takes f. Player A's strategies are displayed in the rows, whereas player B's are shown in the columns of Table III. Strategy d;g for example denotes choosing d in the first turn and g in the second turn, unless the game ends before. For readability, strategies with identical utilities in any case are merged together, e.g., having only a instead of both a;e and a;f.

According to definition of practicality for NFGs (see Definition II.7), the only practical strategy in  $\Gamma_N$  is (a, d; i), which results in a utility of (2,2). This is because for A strategy b; e weakly dominates b; f and for B strategy d; i weakly dominates both c and d; g. After deleting those (in blue), the red strategy b; e of A becomes weakly dominated by a. Thus, after removing b; e only the joint strategy (a, d; i) remains and is therefore a Nash Equilibrium of the resulting game.

However, in the EFG  $\Gamma_E$  the comparison of strategies has a certain order, as not all choices are made simultaneously. Thus, when it comes to B choosing between option c and d, choosing c is also a rational action because in any case B gets utility 1. This is the case, since the subgame following after d, will end in the subgame perfect and practical (1,1), if played by rational players. Following this argumentation, we claim that (b;e,c), yielding history (b,c) should also be considered rational and thus practical.

Example III.4 demonstrates that it is advisable to adapt the NFG concept of practicality for EFGs, and that a näive application can be problematic since information may be lost during the transformation from EFG to NFG [41]. We therefore propose to use subgame perfect equilibria for comparing EFG strategies, and define *practicality for EFGs* as follows.

**Definition III.7** (Practicality for EFG). A strategy of an EFG  $\Gamma$  is practical if it is a subgame perfect equilibrium of  $\Gamma$ .

## C. Security Strategies for Off-Chain Protocols

We now leverage the previously introduced EFG-based definitions (Section III-B) to faithfully model the security of offchain protocols. In particular, we propose the novel concept of *collusion resilience* for addressing (P2), and compare it to existing formalizations of property (P2).

In [19], strong resilience and practicality were used to model the no deviation property of (P2): We identify unwanted properties of strong resilience and we thus investigate variations of it. Specifically, we show that strong Nash Equilibria do not imply strong resilience nor vice-versa (Lemma III.1), and therefore define the *collusion resilience* property of a joint strategy. Intuitively, collusion resilience considers the sum of the utilities of the deviating parties, since rational players may collude or be controlled by the same entity.

**Definition III.8** (Collusion Resilience – CR). A joint strategy  $\sigma \in \mathscr{S}$  in an EFG/NFG  $\Gamma$  is called collusion resilient (CR) if no strict subgroup of players  $S := \{s_1, ..., s_j\}$  has a joint incentive in deviating from  $\sigma$ . That is,

$$\forall S \subset N \quad \forall \sigma'_{s_i} \in \mathscr{S}_{s_i} : \\ \sum_{p \in S} u_p(\sigma) \ge \sum_{p \in S} u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1}, ..., \sigma'_{s_j}/\sigma_{s_j}]).$$
(7)

In addition, we also consider a slight adaption of strong resilience,  $SR_{\subseteq}$ , where the deviation of the entire set of players N is also allowed, as it is for sNE.

**Definition III.9** (Strong Subset Resilience  $-SR_{\subseteq}$ ). *A joint* strategy  $\sigma \in \mathscr{S}$  is called strongly subset resilient ( $SR_{\subseteq}$ ), if no player of any subgroup  $S \subseteq N$ ,  $S := \{s_1, ..., s_j\}$  has an incentive to deviate from  $\sigma$ :

$$\forall S \subseteq N \quad \forall \sigma'_{s_i} \in \mathscr{S}_{s_i} \quad \forall p \in S : \\ u_p(\sigma) \ge u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1},...,\sigma'_{s_i}/\sigma_{s_i}]) .$$

$$(8)$$

We now formalize how the resilience properties relate to each other, which motivates our definition of (P2).

**Lemma III.1** (Resilience Properties). Let  $\sigma \in \mathscr{S}$  be a joint strategy. The following and only the following implications hold.

The next example further motivates why we decided to formalize (P2) in terms of *collusion resilience*.

**Example III.5.** Consider the games  $\Gamma_1$  and  $\Gamma_2$ , respectively defined in Tables IV-V. The games  $\Gamma_1$  and  $\Gamma_2$  show that there exist cases where both strong resilience and strong Nash Equilibria fail to correctly state whether rational players will deviate, while collusion resilience does not.

Let us study  $\Gamma_1$  first. There are three players  $P_1$  on the left,  $P_2$  in the "3rd dimension" who only has one possible strategy, and  $P_3$  at the top. Let us consider the joint strategy  $\sigma = (H_1, H_2, H_3)$ . Since  $P_2$  does not have another choice,  $P_2$ can never deviate. Player  $P_1$  deviating alone yields the same utility as  $\sigma$  and is thus irrelevant. The same holds for  $P_3$ . The only deviation that makes a difference, is if  $P_1$  and  $P_3$  change strategy together to  $(D_1, H_2, D_3)$ . By doing so,  $P_1$  profits and receives 5 instead of 1, but  $P_3$  looses by getting -2 instead of 1. Thus,  $P_3$  does not have an incentive to do so, unless the two players collude for their mutual benefit and share their payoffs. This way they receive 1.5 each instead of 1 each, which poses a serious thread to  $\sigma$  and should thus not be considered satisfying (P2). However,  $(H_1, H_2, H_3)$  is **SNE**, since  $P_3$  has no incentive in deviating with  $P_1$ , if their utilities

TABLE IV	TABLE V
Three Player Game $\Gamma_1$ .	Three Player Game $\Gamma_2$ .

			_			
ねょ	$H_3$	$D_3$		れい	$H_3$	$D_3$
$H_1$	(1,1,1)	(1, 1, 1)		$H_1$	(1, 1, 1)	(1, 1, 1)
$D_1$	(1, 1, 1)	(5,0,-2)		$D_1$	(1, 1, 1)	(3,0,-2)

are not shared, but it is not CR, since

$$2 = u_{P_1}(H_1, H_2, H_3) + u_{P_3}(H_1, H_2, H_3)$$
(9)

$$< u_{P_1}(D_1, H_2, D_3) + u_{P_3}(D_1, H_2, D_3) = 3.$$
 (10)

In the similar game  $\Gamma_2$ , on the contrary,  $P_3$  has no incentive in deviating from  $\sigma = (H_1, H_2, H_3)$  together with  $P_1$ , also if their utilities are shared. Such a deviation yields 0.5 each, instead of 1 each in  $\sigma$ . Hence, there is no incentive to change strategy for one or more players and therefore  $(H_1, H_2, H_3)$ should be considered satisfying (P2). Nevertheless, according to Definition II.5,  $(H_1, H_2, H_3)$  is not SR, since at least one of the deviating parties  $P_1$ ,  $P_3$  profits from choosing  $(D_1, H_2, D_3)$ , although  $P_3$  has no reason to play along. However, in  $\Gamma_2$ ,  $(H_1, H_2, H_3)$  is CR as

$$2 = u_{P_1}(H_1, H_2, H_3) + u_{P_3}(H_1, H_2, H_3)$$
(11)

$$\geq u_{P_1}(D_1, H_2, D_3) + u_{P_3}(D_1, H_2, D_3) = 1.$$
 (12)

**Remark 1** (Formalizing ((P1) and (P2)). Based on the resilience properties of Lemma III.1, we say (P2) is satisfied by a joint strategy  $\sigma$ , if  $\sigma$  is CR and practical. In addition, a joint strategy  $\sigma$  satisfies (P1), if  $\sigma$  is weak immune, as in [19].

We conclude this section by defining secure game strategies/histories, as follows.

**Definition III.10** (Secure Strategy). A strategy  $\sigma$  of an NFG/EFG is secure if it is weak immune, practical and CR.

When discussing security in the setting of EFGs, we are interested mainly in assessing whether a *history* is secure, as the protocol only defines an honest history instead of a full strategy. By applying Definition III.5, we state the following security characterization.

**Definition III.11** (Secure History). A terminal history  $\beta$  of an EFG is secure if there exist extended strategies  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  of  $\beta$ , such that  $\sigma_1$  is weak immune,  $\sigma_2$  is practical and  $\sigma_3$  is CR.

We note that we do not have to find a secure extended strategy for the history to be secure, as aiming for one joint secure strategy in an EFG would be unnecessarily restrictive. Instead, our goal is to make sure that rational parties follow the honest history, no matter what their actual strategy is. In particular, an honest player follows the honest history by default, a rational player does so because of practicality and collusion resistance. Weak immunity further ensures that honest players as well as rational one cannot be damaged by Byzantine players while following the honest history. Hence, the strategy each player has in mind does not matter, since in a secure protocol weak immune, practical, and collusion resistant, strategies are overlapping along the honest history. This is the case because in Definition III.11 we require  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$  to all yield the same history, namely  $\beta$ . We can therefore admit that an honest player has a weak immune strategy in mind, while a rational player has a practical one, as long as these overlap on the honest history.

#### IV. CLOSING GAMES OF OFF-CHAIN PROTOCOLS

We now define a new two-player EFG, called the *Closing Game*  $G_c$ , in order to model closing phase properties of offchain protocols, in particular of the Lightning Network. As explained in Section II-A, to close a channel a party can unilaterally publish a channel state on-chain, which does not necessarily have to be the latest one. The one who closes, however, has to wait a certain amount of time until the money can be used. Meanwhile, the other party can steal all the money from the channel in case the state published on-chain is not the latest one: this ensures that rational players close their channel only with the latest state. Alternatively, the parties can collaboratively sign a new transaction to split the money. In this case no one has to wait.

Our closing game overcomes the limitations of previous work [19] in representing dishonest closing attempts, by modeling how closing can be achieved after a failed collaborative closing attempt and by also considering the additional fee fto be paid in a revocation transaction.

To the best of our knowledge, our closing game  $G_c$  is the most accurate model for the security analysis of off-chain protocols, notably of the Lightning Network. In our model of the closing phase we make the following assumptions for a channel between A and B at the moment where the closing phase is initiated.

- The fair split of the channel's funds is  $a \rightarrow A$ ,  $b \rightarrow B$  and a > 0, b > 0.
- The benefit of closing the channel is α. Closing a channel yields a benefit, since it unlocks assets.
- The opportunity cost of having to wait for one's funds upon closing is ε.
- When both players agree to update the channel we assume a fair deal in the background which yields a profit of  $\rho$  for both parties.
- Publishing a revocation transaction on-chain costs a fee f > 0.

Further, to properly model utilities in the closing game  $G_c$ , we define the following total order, which is crucial for analyzing security properties of  $G_c$ . For capturing total order properties in the setting of  $G_c$ , we extend the set  $\mathbb{R}$  of real numbers by the infinitesimal numbers  $\alpha$ ,  $\varepsilon$  and  $\rho$ .

**Definition IV.1** (Utility Order). We consider the total order  $(\mathbb{U}, \preccurlyeq)$ , where  $\mathbb{U}$  is the group resulting from closing  $\mathbb{R} \cup \{\alpha, \varepsilon, \rho\}$  under addition. The total ordering  $\preccurlyeq$  is uniquely defined by the following conditions.

1) On  $\mathbb{R}$ , the relation  $\preccurlyeq$  is the usual less than or equal relation  $\preccurlyeq \mid_{\mathbb{R}} := \le$ .

TABLE VI POSSIBLE ACTIONS IN  $G_c(A)$ .

п	Class unilaterally and key such without reacting to a previous move
Η	Close unilaterally and <i>honestly</i> without reacting to a previous move,
	such as a collaborative closing attempt.
D	Close unilaterally but <i>dishonestly</i> (without reacting to a previous move)
	with a profit of $d_A \in (0,b]$ in A's case, $d_B \in (0,a]$ in B's case.
$C_h$	Try to close <i>collaboratively</i> and <i>honestly</i> , that is proposing a fair split.
$C_c$	Try to close <i>collaboratively</i> but by <i>cheating</i> the other party by $c \in$
	(0,b], that means proposing an unfair split.
S	Signing the collaborative closing attempt of the other player.
J	Ignore the previous action and do nothing.
P	Prove other party tried to close dishonestly. That means stating a
	revocation transaction. We assume its publication requires a fee of
	f > 0 and that the attempt to do so is always successful, that is that
	the miners behave honestly.
$U^+$	Propose an <i>update</i> of the channel where player <i>A</i> 's balance is <i>increased</i>
	by $p_A \in (0, b]$ .
$U^{-}$	Propose an <i>update</i> where player A's balance is <i>decreased</i> by $p_B \in$
	(0,a].
A	Agree to a proposed update.

2) The values  $\alpha$ ,  $\varepsilon$  and  $\rho$  are greater than 0,

$$\forall \xi \in \{\alpha, \varepsilon, \rho\} : -\xi \prec 0 \prec \xi . \tag{13}$$

- 3) The values  $\alpha$ ,  $\varepsilon$  and  $\rho$  are closer to 0 than any real number,
  - $\forall x \in \mathbb{R}, \xi \in \{\alpha, \varepsilon, \rho\}, x > 0: \xi \prec x, -x \prec -\xi.$ (14)
- 4) Additionally,  $\alpha$ ,  $\varepsilon$  and  $\rho$  have the order  $\rho \prec \varepsilon \prec \alpha$ .

In general, unlocking funds gives additional financial freedom even if there is some processing delay; therefore, we choose  $\varepsilon \prec \alpha$  in Definition IV.1. Additionally, once the parties initiate the closing phase, it is reasonable to assume that no potential update significantly benefits both parties. In contrast, both parties are interested in avoiding the opportunity cost, i.e., the cost of having to wait for their funds upon closing, therefore, we set  $\rho \prec \varepsilon$  in Definition IV.1.

**Remark 2.** While the ordering conditions of Definition IV.1 may seem to be restrictive, lifting them comes with the burden of considering a high number of possible variable orderings. In particular, one would need to consider (number of variables)! orderings, which would highly complicate the formal analysis task. Approximating or clustering the number of orderings, while weakening conditions in Definition IV.1, is an interesting venue for future work.

Based on the utility ordering of Definition IV.1, we introduce our *Closing Game for Player A* below.

**Definition IV.2** (Closing Game  $G_c(A)$  of Player A). The Closing Game  $G_c(A) = (N, \mathcal{H}, P, u)$  is an EFG with two players  $N = \{A, B\}$ . The tree representation of  $G_c(A)$  in Figure 4 defines  $\mathcal{H}$ , P and  $u^3$ , with the actions of the game being summarized in Table VI.

Note that the *utility function u* of  $G_c(A)$  in Figure 4 assigns player  $p \in N$  the money player p received minus the money player p deserved based on the latest channel state. The values of closing  $(\alpha)$ , updating  $(\rho)$  and waiting  $(-\varepsilon)$  are also

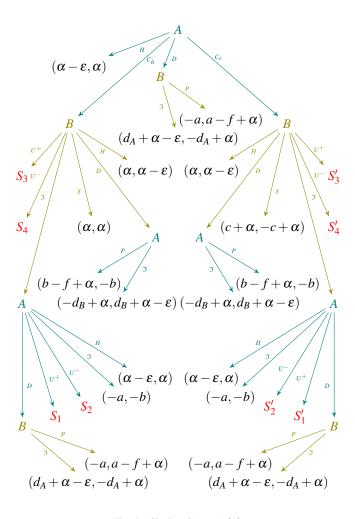


Fig. 4. Closing Game  $G_c(A)$ .

considered in Figure 4. As discussed in Section II-A, the fee needed for the closing transaction is assumed to have been reserved among the locked funds in the channel all the time and is spent upon closing, therefore not affecting the players' channel balance.

The closing game for player B,  $G_c(B)$  is defined similarly to  $G_c(A)$ , with the roles of A and B being swapped in Definition IV.2. Based on the closing games  $G_c(A)$  and  $G_c(B)$ , we consider the closing phase in an off-chain channel as given in Figure 5 and defined below.

**Definition IV.3** (Closing Phase). The closing phase of an offchain channel modeled by a closing game  $G_c(A)$  is initiated in one of three ways: (i) A starts with a closing action C, and thus triggers the closing game  $G_c(A)$ ; (ii) A does not start a closing action, thus performing action ignore  $\Im$ , but B starts with a closing action C and triggers  $G_c(B)$ ; or (iii) none of the players A and B ever start closing, that is B also choosing action  $\Im$ , in which case the money stays locked in the channel. Then, we get the EFG  $\Gamma_C$  from Figure 5 modeling the closing phase of  $G_c(A)$  and  $G_c(B)$ .

<sup>&</sup>lt;sup>3</sup>The subgames  $S_i$ ,  $S'_i$  are given in the appendix.

$$\begin{array}{c|c} A & \xrightarrow{\mathfrak{I}} & B & \xrightarrow{\mathfrak{I}} & (-a, -b) \\ c \downarrow & c \downarrow \\ G_c(A) & G_c(B) \end{array}$$

#### Fig. 5. Closing Phase $\Gamma_C$ .

V. CLOSING GAMES FOR SECURE LIGHTNING CHANNELS

We now show that the closing games from Definition IV.2 precisely capture secure closing phases in Lightning channels [6]. Namely, the following two terminal histories of closing games model the honest behavior of Lightning: (i) history (*H*) from Figure 4 represents unilateral honest closing of *A*, yielding utility  $(\alpha - \varepsilon, \alpha)$ ; and (ii) history (*C<sub>h</sub>*, *S*) captures the attempt of *A* to close collaboratively and honestly, while *B* signs, with a utility of  $(\alpha, \alpha)$ . Our security analysis focuses on these two honest histories of Lightning channels.

**Definition V.1** (Honest Closing). *The only* honest histories *in* the closing game  $G_c(A)$  are the terminal histories (H) honest unilateral closing and  $(C_h, S)$  honest collaborative closing. All strategies yielding one of the two histories are considered honest strategies.

In the following, the values  $d_A$  (resp.  $d_B$ ) defined in Table VI (line *D*) represent the difference of funds between the latest state and the old one that is dishonestly posted on chain by *A* (resp. *B*). In other words, if (a,b) is the latest state, the one posted on chain is  $(a + d_A, b - d_A)$  (resp.  $(a - d_B, b + d_B)$ ), thus enabling dishonest closing attempts of profit  $d_A$  for *A* (resp.  $d_B$  for *B*). The values  $p_{A,B}$  (Table VI, lines  $U^+$ ,  $U^-$ ) and *c* (Table VI, line  $C_c$ ) can respectively be chosen by *A* and *B* at the time of the action and do not depend on previous distribution states. Based on this setting, we derive the security properties (P1) and (P2) of Lightning channels as given below. The omitted proofs are given in the appendix.

**Theorem V.1** (Weak Immunity of Honest Behavior – (P1)). The terminal histories (H) of honest unilateral closing, and  $(C_h, S)$  of honest collaborative closing of  $G_c(A)$  are weak immune, if the channel balances are higher than the fee required in a revocation transaction, that is if  $a \ge f$  and  $b \ge f$ .

Theorem V.1 implies that as long as both players have a minimal balance of f in the channel, no honest player can lose money. As such, Theorem V.1 establishes the security property (P1) ensuring "no honest loss" in the channel.

Further, to ensure the security property (P2) of "no deviation", we require that

$$a - p_B + d_A \ge f \quad \text{and} \tag{15}$$

$$b - p_A + d_B \ge f . \tag{16}$$

To understand the inequations (15)-(16) consider the history  $(C_h, \Im, U^-, \mathfrak{A}, D)$  in Figure 4, respectively  $S'_1$ . This history formalizes the case where *A* attempts honest collaborative closing (action  $C_h$ ) and *B* ignores it (action  $\Im$ ). Then *A* proposes an update (action  $U^-$ ) from state (a,b) to state  $(a - p_B, b + p_B)$  and *B* agrees (action  $\mathfrak{A}$ ). Finally, *A* closes dishonestly (action

D) using the old distribution state  $(a - p_B + d_A, b + p_B - d_A)$ . Let us also study the options B has. By ignoring A's behavior (action  $\Im$ ), B receives  $b + p_B - d_A$  instead of the fair amount  $b + p_B$ , leaving B with a loss of  $d_A$ . By publishing the revocation transaction (action P), B receives a + b but has to pay the fee f for pushing it on the blockchain, which leads to a win of  $a - p_B - f$ . Therefore, the win should be greater than the loss; hence

$$a - p_B - f \ge -d_A \quad \Leftrightarrow \quad a - p_B + d_A \ge f$$
, (17)

in order for a rational *B* to publish the revocation transaction. This in turn yields a loss for *A* and hence discourages *A* from closing dishonestly, which is necessary for the incentive compatibility (P2) of Lightning's closing phase. By swapping *A*'s and *B*'s roles, we get the prerequisite formulated in (16). These extreme cases of dishonest closing subsume the others. Thus, the only preconditions we need in the following Theorem V.2 are (15)–(16). In summary, formulas (15)–(16) ensure that ignoring any dishonest closing attempt is worse than publishing the revocation transaction. Property (P2) is then established by the following theorem.

**Theorem V.2** (Incentive-Compatibility – (P2)). If  $a - p_B + d_A \ge f$  and  $b - p_A + d_B \ge f$ , then

- 1) honest unilateral closing (H) is CR, but not practical.
- 2) honest collaborative closing  $(C_h, S)$  is CR. It is practical iff  $c \neq p_A$ .

**Remark 3** (Explanation of  $c \neq p_A$ ). The condition  $c \neq p_A$ in Theorem V.2 has the following relevance. Player A can in principle choose to propose dishonest collaborative closing (action  $C_c$ ), providing A an unfair advantage of value c. Then, either B (action  $U^+$ ) or A (B choosing action  $\Im$  to ignore first, then A taking action  $U^+$ ) can propose a channel update  $(a,b) \mapsto (a+c,b-c)$ . The value of the update  $p_A$  is now equal to the amount player A cheated with in  $C_c$ :  $p_A = c$ . In this special case, the closing game behaves differently. The described histories  $(C_c, \Im, U^+)$  and  $(C_c, U^+)$  lead to the subgames  $S'_1$  and  $S'_3$  respectively. Let us consider  $S'_3$  with  $p_A = c$ .

Assume A agrees to the update, action  $\mathfrak{A}$ , and player B signs the initially unfair collaborative closing attempt of A. Since in the meantime the channel was updated by the exact amount that A tried to cheat with, the pending collaborative closing now contains the fair split. Therefore, both players profit from this course of action, yielding utility ( $\rho + \alpha, \rho + \alpha$ ). The analog can be achieved in subgame  $S'_1$  with the history ( $\mathfrak{A}, \mathfrak{I}, S$ ). In fact, for  $p_A = c$ , those histories are the only practical ones and provide the mutually best outcome possible.

However, updating to (a + c, b - c) first and then closing honestly and collaboratively yields the exact same result. This is why we study the closing game without the possibility of updating after a closing attempt in the next section Section V-A.

We now state our first main security theorem. Since (H) is not practical, a rational player will not play it. Hence, the terminal history (H) is not secure. We get the following

security result instead for  $(C_h, S)$ .

**Theorem V.3** (Security of  $G_c(A)$ ). If  $a \ge f$ ,  $b \ge f$ ,  $a - p_B + d_A \ge f$ ,  $b - p_A + d_B \ge f$ , and  $c \ne p_A$ , then the closing game  $G_c(A)$  together with the honest behavior  $(C_h, S)$  is secure.

*Proof.* As  $a \ge f$  and  $b \ge f$ , we have that  $(C_h, S)$  is weak immune (Theorem V.1). Since  $a - p_B + d_A \ge f$ , we derive  $b - p_A + d_B \ge f$  and  $c \ne p_A$ , we have that  $(C_h, S)$  is also practical and CR (Theorem V.2). Hence, by Definition III.11,  $(C_h, S)$  is secure.

Theorem V.3 implies that for honest and rational players the action of collaborative closing followed by signing  $(C_h, S)$  is the best way to close an off-chain channel. It also implies, that rational adversaries will cooperate. Further, Byzantine players represent no threat as long as their channel balances are high enough and they do not engage in special cases of channel updates after a collaborative closing attempt.

We note that for proving our security properties (P1)-(P2) in Theorem V.1–Theorem V.3, we rely on a succinct analysis of the finite graph properties of the closing game  $G_C(A)$ from Figure 4. While automated approaches analyzing a finite number of graph properties exist, see e.g. [42], [43], these approaches cannot handle (game) graphs where graph leaves contain variables, instead of specific numerical values, which is the case of  $G_C(A)$ . For such cases, automated reasoning tools, such as theorem provers, need to be combined with graph-theoretic manipulations of  $G_C(A)$ , an approach we aim to investigate as a future work towards automating the security analysis (and proofs) of closing games.

#### A. Closing Games without Updates

We will now consider a variation of closing games without updates, as updating is not beneficial for at least one player upon closing. Furthermore, we avoid special cases such as the one described in Remark 3, which should be equivalent to updating before initiating  $G_c(A)$ , and then closing honestly and collaboratively. As such, the *closing game*  $G_c(A)$  without updates results from removing all actions  $U^+$  and  $U^-$  in Figure 4. For the resulting closing game  $G_c(A)$  without updates we get the following security result similar to Theorem V.3.

**Theorem V.4** (Security of  $G_c(A)$  without Updates). If  $a \ge f$ and  $b \ge f$ , then the closing game  $G_c(A)$  without updates and together with both honest histories (H) and ( $C_h$ , S) is secure.

*Proof.* We respectively fix honest strategies  $\sigma$  and  $\sigma'$  for histories (*H*) and (*C<sub>h</sub>*,*S*); let  $\sigma'$  have *A* choosing *C<sub>h</sub>* initially, *P* after (*C<sub>h</sub>*,*D*) and *H* after (*C<sub>h</sub>*,  $\mathfrak{I}$ ), and then *B* choosing *S* after (*C<sub>h</sub>*), *P* after (*D*) and *H* after (*C<sub>c</sub>*). Infer that the deviation of *A* causes negative utility for *B*, whereas the deviation of *B* leads to non-negative utility for *A* as  $b - f \ge 0$ . By Theorem V.1 we thus have that  $\sigma'$ , and therefore (*C<sub>h</sub>*,*S*), are weak immune. In addition, Theorem V.1 implies that also (*H*) is weak immune.

To show practicality, we compute all subgame perfect terminal histories. From  $a \ge f$  and  $b \ge f$  we have  $a + d_A \ge f$  and  $b + d_B \ge f$ . Since closing with a dishonest behavior yields

utility  $a - f + \alpha$ ,  $b - f + \alpha$  respectively, whereas ignoring a dishonest behavior leads to  $-d_A + \alpha$  and  $-d_B + \alpha$ , we conclude that the best choice after action *D* is always *P*. Thus, *A*'s best choice after  $(C_h, \mathfrak{I})$  and  $(C_c, \mathfrak{I})$  is *H*. Therefore, *B* has the two subgame perfect options  $\mathfrak{I}$  and *S* after  $(C_h)$ , and only  $\mathfrak{I}$  after  $(C_c)$ , yielding thus the following practical histories: history  $(C_h, S)$  with utility  $(\alpha, \alpha)$ ; and  $(C_h, \mathfrak{I}, H)$ ,  $(C_c, \mathfrak{I}, H)$ , and (H) each with utility  $(\alpha - \varepsilon, \alpha)$ . Therefore, both (H) and  $(C_h, S)$  are practical.

Note that every practical terminal history is a Nash Equilibrium, since if a deviation could benefit a player, the player would have chosen differently already. As CR is equivalent to Nash Equilibria in two-player games (by Definition II.5 and Lemma III.1), we use Definition III.7 and Lemma III.1 to conclude that practicality of (H) and  $(C_h, S)$  implies collusion resistance CR of (H) and  $(C_h, S)$ . As (H) and  $(C_h, S)$  are both weak immune, practical and CR, by Definition III.11 we infer that they are also secure.

**Remark 4.** Note that the analysis of utilities in the closing game  $G_c(A)$  crucially depends on constraints of the underlining ordering that we set in Definition IV.1, and thus on the values of variables  $a, b, c, d_{A,B}$ , f in Table VI. In general, the bigger  $\varepsilon$  gets in Definition IV.1, the more discouraged is closing unilaterally in Table VI, and hence in Figure 4. Further, B is more likely to accept a dishonest collaborative closing attempt  $C_c$ , as it is better to lose c than to lose  $\varepsilon$ .

We further study what happens if a player has almost no funds left in a channel. In particular, we show that security properties, in particular weak immunity and practicality, are violated in this case, thereby formalizing the following folklore in the community.

**Theorem V.5** (Little Funds). If a < f, then only terminal histories that involve an explicit cheating attempt are weak immune in the closing game  $G_c(A)$  without updates. A terminal history involves an explicit cheating attempts if one of its actions is  $C_c$  or D.

*Proof.* Let  $\sigma$  be any strategy, yielding a history that does not involve an explicit cheating attempt. Then *A* can deviate to a strategy where *A* chooses *D* as its first action. In this case, the honest *B* gets negative utility, no matter whether *B* chooses *P* or  $\Im$ , since a < f. Hence, only histories that involve explicit cheating attempts can be weak immune.

We next derive the following results on security properties.

**Corollary 1.** If there exists an old channel state  $(a+d_A, b-d_A)$ , with  $a+d_A < f$ , then neither history (H) nor  $(C_h, S)$  is weak immune nor practical, but CR.

**Corollary 2.** A rational party should never, in any channel, let the opponent's balance fall below f, because at that point the other party can always cause financial loss by closing dishonestly and unilaterally<sup>4</sup>.

<sup>4</sup>The special edge cases a = 0 or b = 0 are considered in the appendix.

*Proof.* Once the opponent's balance is below f, that party can start the closing game, therefore the opponent becoming A. Thus, by applying Theorem V.5, it follows that the opponent can make the rational player lose money by closing unilaterally and dishonestly. If it is not the first time that A's balance is below f and the respective old state contains a higher balance for A than the latest one, then we are even in the situation of Corollary 1. It is thus rational of A (practical) to close dishonestly.

# B. Optimal Strategy for Closing Off-Chain

To summarize, our security analysis based on closing games for Lightning channels yields the following results. Theorem V.4-Theorem V.5, together with with Corollary 1-Corollary 2, allow us to derive the optimal strategy for closing an off-chain channel for a rational and suspicious player. We next describe and illustrate this optimal strategy, highlighting the main steps of our security analysis based on Theorem V.4-Theorem V.5.

Without loss of generality, we assume the current state of the channel is (a,b).

The player, assumed to be player *A*, who initiated the closing phase shall:

• try to close honestly and collaboratively (action  $C_h$ ), if there does not exist an old state  $(a + d_A, b - d_A)$ , where  $d_A > 0$  and  $a + d_A < 0$ . In case the other player, that is player *B*, does not sign (action *S*), player *A* shall close honestly and unilaterally (action *H*).

If player *B* closed dishonestly and unilaterally (action *D*), player *A* shall:

- state the revocation transaction (action *P*), if the state used for cheating was  $(a d_B, b + d_B)$ , where  $d_B > 0$  and  $b + d_B \ge f$ .
- ignore the cheating otherwise (action  $\Im$ ), as it yields less loss.
- close dishonestly and unilaterally (action *D*), if there exists an old state  $(a + d_A, b d_A)$ , where  $d_A > 0$  and  $a + d_A < f$ . In this case, player *A* shall use the old distribution state  $(a + d'_A, b d_A)$ , with the highest  $d'_A > 0$  that still satisfies  $a + d'_A < f$ .

The reacting player, in this case assumed to be player *B*, shall:

- sign the collaborative honest closing attempt (action S), if applicable, if there is no old state  $(a - d_B, b + d_B), d_B > 0$ in which the funds of player B are less then f, that is if  $b + d_B < f$ .
- close honestly and unilaterally (action *H*), in case of a dishonest collaborative closing attempt (action  $C_c$ ). This holds, if there is no old state  $(a d_B, b + d_B)$ ,  $d_B > 0$  in which the player *B*'s funds are less then *f*, that is  $b + d_B < f$ .
- otherwise ignore (action  $\mathfrak{I}$ ) the collaborative and honest/dishonest closing attempt, if applicable, and close dishonestly and unilaterally (action *D*), using the old state  $(a-d'_B,b+d'_B)$ , with the highest  $d'_B > 0$  that still satisfies  $b+d'_B < f$ .

- state the revocation transaction (action *P*), if player *A* tried to close dishonestly and unilaterally (action *D*) with state  $(a + d_A, b d_A)$ , where  $d_A > 0$  and  $a + d_A \ge f$ .
- ignore (action  $\Im$ ) if player A closed dishonestly (action D), in the case where  $a + d_A < f$ , as it yields less loss.

**Example V.1.** Let players A and B share a channel with initial balance (5,5) and let us assume the fee for publishing a revocation transaction f = 2. After the first update let their state be (3,7). The optimal way for A to close now is  $C_h$  and for B to sign. Dishonest closing would cause B to publish the revocation transaction, yielding a loss of 3 for A and a profit of 3-2=1 for B.

The next update could be (1.8, 8.2). The best way to close for A is still C<sub>h</sub>. Dishonest closing using (3,7), for example, would still cause B to publish the revocation transaction. Player B would in this case lose 1.8 - 2 = -0.2, but he would lose more, 7 - 8.2 = -1.2, by ignoring it.

Another update could be (1,9). Now the optimal strategy for A to close is D, using the old state (1.8,8.2). Ignoring the dishonest closing (action  $\Im$ ) brings B - 0.8, but proving A's cheating (action P) leads to 1-2 = -1. Hence, a rational B will choose to ignore (action  $\Im$ ), that means B does not publish the revocation transaction.

#### VI. BEYOND CLOSING GAMES FOR OFF-CHAIN SECURITY

Our game-theoretic analysis so far focused on using closing games to capture security properties of off-chain channels (Section IV), and in particular of Lightning channels (Section V). In this section, we show that our game-theoretic formalism from Section III is expressive enough to analyse more complex protocols than just closing phases in Lightning channels. In particular, we introduce a new EFG, called the *Routing Game* in Section VI-A, and use this game in Section VI-B to disprove security of Lightning's routing mechanism amid the Wormhole and Griefing attacks [7], [20]. We also discuss a natural extension of our analysis to model other off-chain protocols in Section VI-C.

## A. Routing Games for Lightning's Routing Module

We first propose a new EFG, called the *Routing Game*, showing that EFGs can capture actual attacks, in this case the Wormhole attack [7] and the Griefing attack [20], which were overlooked for example in [19]. Specifically, the below defined routing game considers fees f, and supports actions allowing the intermediaries to choose not to claim their money using the secret x but instead to forward it to another intermediary (as explained in Section II-A). Additionally, other deviations such as creating a conditional payment (i.e. HTLC) with a different hash value, a different amount, or a different time-out than expected are also considered. For simplicity, we chose to model our routing game below with five players; however, an arbitrary number of intermediaries can be modeled.

**Definition VI.1** (Routing Game  $G_{\text{rout}}$ ). The routing game  $G_{\text{rout}} = (N_r, \mathcal{H}_r, P_r, u_r)$  is an EFG with five players  $N = \{A, E_1, I, E_2, B\}$ , where

$$(\rho, m+3f-\varepsilon, -\varepsilon, -m, \rho) \xrightarrow{\circ} I \xrightarrow{v} (\rho, f, m+2f-\varepsilon, -m, \rho)$$

$$(m+3f+\rho-\varepsilon, -\varepsilon, -\varepsilon, -m, \rho) \xrightarrow{\circ} (\rho, m+3f-\varepsilon, -\varepsilon, -m, \rho)$$

$$(m+3f+\rho-\varepsilon, -\varepsilon, -\varepsilon, -m, \rho) \xrightarrow{\circ} E_{1} \xrightarrow{s_{s_{1}}} (\rho, m+3f-\varepsilon, -\varepsilon, -m, \rho)$$

$$(m+3f+\rho-\varepsilon, -\varepsilon, -\varepsilon, -m, \rho) \xrightarrow{\circ} E_{1} \xrightarrow{s_{s_{1}}} (\rho, m+3f-\varepsilon, -\varepsilon, -m, \rho)$$

$$(0,0,0,0,0) \xrightarrow{s_{1}} \downarrow^{t_{A}} E_{1} \xrightarrow{t_{r_{f}/t_{H}}} S_{3} S_{3}$$

$$(0,0,0,0,0) \xrightarrow{s_{1}} \downarrow^{t_{A}} E_{1} \xrightarrow{t_{r_{f}/t_{H}}} S_{4} S_{4} \xrightarrow{s_{s_{1}}} (m+3f+\rho-\varepsilon, -m-2f, m+2f-\varepsilon, -m, \rho)$$

$$(0,0,0,0,0) \xrightarrow{s_{1}} \downarrow^{t_{A}} E_{2} \xrightarrow{t_{A}} I \xrightarrow{t_{r_{f}/t_{H}}} E_{2} \xrightarrow{t_{r_{f}/t_{H}}} S_{3} \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{v} (\rho, f, m+2f-\varepsilon, -m, \rho)$$

$$(n+3f+\rho-\varepsilon, -\varepsilon, -\varepsilon, -m, \rho) \xrightarrow{s_{s_{1}}} B \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{s_{s_{1}}} I \xrightarrow{v} I \xrightarrow{v$$

Fig. 6. Partial Definition of the Lightning's Routing  $G_{rout}$  and the Fulgor Model  $G_{Ful}$ . The olive colored subtree only applies for  $G_{rout}$ .

- the histories  $\mathcal{H}_r$ , the next player function  $P_r$ , and the utility function  $u_r$  are defined via the tree representation of Figure 6. The utility tuples in Figure 6 assign the first value to A, the second to  $E_1$ , the third to I, the fourth to  $E_2$ , and the last to B;
- the actions of  $G_{rout}$  are as listed in Table VII.

We note that our Routing Game  $G_{\text{rout}}$  has four types of subgames, as modeled in Figure 6 and described next: (i) subgames that result from sending the secret to another player  $\mathfrak{S}_i$ ; (ii) subgames that result from locking a wrong amount of money in the HTLC  $\mathbb{S}_i$ ; (iii) subgames that result from using a wrong time-out in an HTLC  $\mathbf{S}_i$ ; and (iv) subgames that result from using a wrong hash value as lock in the HTLC  $\mathbf{S}_i$ . We further note that Figure 6 only gives a partial model, as not all subgames are presented in Figure 6. However, within one type of subgame, the game trees are similar. Therefore, we provide only one instance of each type in the appendix, which are the subgames  $\mathbb{S}_1$ ,  $\mathbb{S}_2$ , and  $\mathbb{S}_3$ . An instance of a secret forwarding subgame capturing the Wormhole attack can be seen in  $G_{\text{rout}}$ , as the subtree after history  $(S_H, L, L, L, U, S_{S_{E_1}})$ .

Let us emphasize that the utility function  $u_r$  of  $G_{rout}$  assigns each player  $p \in N$  the relative profit of their routing actions and does not mirror the individual channel balances. It also takes the value  $\rho$  of a successful payment and the opportunity cost  $\varepsilon$  into account.

As in the closing games  $G_c(A)$  and  $G_c(B)$ , we aim to align utility and monetary outcome as tight as possible. We adjust the ordering  $(\mathbb{U},\preccurlyeq)$  of Definition IV.1 by not assuming that  $\rho \prec \varepsilon$ , since achieving an update is the ultimate goal of the routing protocol. We also consider the utility relative to the amount due to each party.

#### B. Security Analysis of Lightning's Routing Module

Let us recall Figure 1 and Figure 2, where player A wants to pay another player B money of value m. Since, A and B do not share a channel, the three intermediaries  $E_1$ , I, and  $E_2$ support the payment, with each receiving a fee f > 0 for their collaboration if the payment is successful. Each player who creates an HTLC locks her money for a given time, yielding an opportunity cost of  $\varepsilon$  if the money is returned. If the transaction fails, before anyone has unlocked an HTLC, all parties get utility 0 or  $-\varepsilon$ , depending on whether they created an HTLC or not. Otherwise, the intermediaries' utilities are according to their financial win/loss. The parties A and B both receive  $\rho$  once B is paid. Should the transaction fail after B is paid, but before A has paid, she has utility  $m+3f+\rho-\varepsilon$ ; once  $E_1$  collects the money, A's utility is  $\rho$ .

In the sequel, we consider the behavior from Figure 1 as the only *honest* history in  $G_{rout}$ , as also formalized next.

**Definition VI.2** (Honest Routing). *The only* honest history *in the routing game*  $G_{rout}$  *is the history*  $(S_H, L, L, L, L, U, U, U, U)$ . *All strategies yielding this history are considered* honest strategies.

Using our model  $G_{\text{rout}}$  and its honest behavior, we derive the following result.

**Theorem VI.1** (Vulnerability of  $G_{\text{rout}}$  to Wormhole Attacks). The honest behavior  $(S_H, L, L, L, L, U, U, U, U)$  of the Routing Game  $G_{\text{rout}}$  is not CR.

*Proof.* The utility of the honest behavior of the routing module  $(S_H, L, L, L, L, U, U, U, U)$  is  $(\rho, f, f, f, \rho)$  (as indicated in red in Figure 6). Let us compare this behavior and utility to the deviating terminal history  $(S_H, L, L, L, L, U, S_{S_{E_1}}, U, \Im)$  with a utility of  $(\rho, m+3f-\varepsilon, -\varepsilon, -m, \rho)$  (given in blue in Figure 6). It is not hard to argue that the collusion of  $E_1$  and  $E_2$  (and *B* by not sending the secret to *I*) strictly profits from the deviation, which yields a joint utility of  $3f - \varepsilon + \rho$ , whereas the honest behavior only yields a joint utility of  $2f + \rho$ . As such, collusion resistance CR is violated, since no honest player can prevent the Wormhole attack from happening by following any honest strategy (that is, a strategy  $\sigma$  whose history is the honest behavior  $(S_H, L, L, L, L, U, U, U, U)$ ).

In conclusion, Theorem VI.1 formally proves that Lightning's routing module is susceptible to the Wormhole attack. We further extend this result by noting that not only can  $G_{rout}$  capture the Wormhole attack, but also the Griefing attack, as stated below.

**Theorem VI.2** (Vulnerability of  $G_{\text{rout}}$  to Griefing Attack). The honest behavior  $(S_H, L, L, L, L, U, U, U, U)$  of the Routing Game  $G_{\text{rout}}$  is not weak immune.

*Proof.* For showing that history  $(S_H, L, L, L, L, U, U, U, U)$  is not weak immune, we prove that no strategy which yields this history is weak immune. Let us consider any such strategy

TABLE VII Possible Actions in  $G_{\text{ROUT}}$ .

$S_H$	Sharing the secret's Hash to enable the others to create HTLCs (action
	1 in Figure 1, Section II-A).
L	Lock money, as defined in actions 2–5 in Figure 1, in an HTLC.
U	Unlocking the money from an HTLC (actions 6–9 in Figure 1). Thereby
	the secret is revealed to the HTLC's creator.
J	Ignoring all the previous actions and do nothing. If applicable, until
	the unlockable HTLC has timed out.
$S_S$	Sending the Secret to another player. If it is sent to a specific player
	(not leading to $\mathfrak{S}_i$ ) this player is indicated by another subscript.
$L_H$	Locking money in an HTLC, that uses a different Hash-lock than
	described in Figure 1.
$L_A$	Locking a different Amount of money in an HTLC, than described in
	Figure 1.
$L_T$	Locking money in an HTLC, whose Time-out is different from the
	values described in Figure 1.

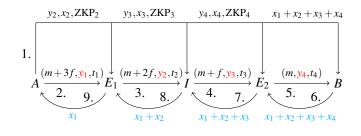


Fig. 7. Routing in Fulgor.

σ. Then, player *A* has to choose action *L* after *B* sent her the secret, that is history (*S<sub>H</sub>*). Assume now *E*<sub>1</sub> deviates and chooses to ignore (action ℑ). Then *A*'s utility is -ε ≺ 0. Hence, history (*S<sub>H</sub>*,*L*,*L*,*L*,*U*,*U*,*U*,*U*) is not weak immune.

We also obtain the following result as an immediate consequence of Theorem VI.1 and Theorem VI.2.

**Corollary 3** (Security of Routing Module). The honest behavior  $(S_H, L, L, L, L, U, U, U, U)$  of the Routing Game  $G_{rout}$  is not secure. Hence, the Routing Game  $G_{rout}$  is not secure.

## C. Further Routing Protocols Beyond the Lightning Network

We conclude this paper by arguing that our EFG games, either closing or routing games, are not restricted to Lightning networks but can be used for other protocols as well. In the remaining of this section, we illustrate how to model Fulgor [38], a payment channel network protocol that fixes the Wormhole attack, but not the Griefing attack.

The routing mechanisms used in Fulgor is similar to Lightning's routing, and is similarly based on HTLCs. The main difference lies in the structure of the secrets and their hashes. Indeed, while Lightning uses the same secret x for every HTLC, Fulgor provides a different secret and hash lock for each player.

Fulgor's routing mechanism is illustrated in Figure 7, where player A generates different secrets and hash locks at the beginning. The secrets and the hashes relate in the following way:  $h(x_1) = y_1$ ,  $h(x_1 + x_2) = y_2$ ,  $h(x_1 + x_2 + x_3) = y_3$  and  $h(x_1 + x_2 + x_3 + x_4) = y_4$ . Therefore, a player only gets to know a sum of secrets when the right-hand party unlocks and subtracts the secret value received from A to unlock their HTLC. A also provides a zero-knowledge-proof ZKP<sub>i</sub> for each intermediary [44] to prove that the secrets and hashes constructed this way guarantee successful unlocking of the left HTLC, which is essential to not lose funds.

The game-theoretical (EFG) model of Fulgor  $G_{Ful}$  reported in Figure 6 looks similar to the routing game  $G_{rout}$ , yet, with one significant difference. Consider the history  $(S_H, L, L, L, L, U, S_{S_{E_1}})$  in Figure 6, which enables player  $E_1$  to unlock (action U) the HTLC created by A. In Fulgor, the same history does not enable  $E_1$  to unlock the HTLC. As Figure 7 shows, the secrets that  $E_2$  can share after action 6 are  $x_4$  and  $x_1 + x_2 + x_3 + x_4$ . Further,  $E_1$  only knows  $x_2$ , thus there is no way to compute  $x_1$ . This is however the value needed to unlock the HTLC created by A. Indeed, Fulgor is

not affected by the Wormhole attack. Nevertheless, similarly to Theorem VI.2, the honest behavior of Fulgor is not weak immune, as it is vulnerable to the Griefing attack.

### VII. CONCLUSIONS

Our work advocates the use of Extensive Form Games (EFGs) for the game-theoretic security analysis of off-chain protocols. In particular, we introduce two instances of EFGs to model the closing and the routing of the Lightning Network. By doing so, we take the first step towards closing the gap existing security proof techniques have due to using informal arguments about rationality. We express security properties as formal requirements over joint strategies in EFGs, allowing us to establish optimal strategies for closing off-chain and capture security vulnerabilities amid attacks. Given the theoretical expressiveness of our EFGs, future work includes the definitions of new games to capture a wider range of off-chain protocols. To overcome the burden of tedious manual analysis, we also plan to leverage SMT solving and/or automated theorem proving in order to provide automated security proofs.

Acknowledgments. We thank our anonymous reviewers for their valuable feedback. The work was partially supported by the European Research Council (ERC) under the ERC CoG ARTIST 101002685 and the ERC CoG BROWSEC 71527; by the TU Wien Doctoral College SecInt; by the Austrian Science Fund (FWF) projects PROFET P31621 and LogiCS W1255-N23; by the Austrian Research Promotion Agency (FFG) (COMET K1 SBA, COMET K1 ABC); by the Vienna Business Agency project Vienna Cybersecurity and Privacy Research Center (VISP); by the Austrian Federal Ministry for Digital and Economic Affairs; the National Foundation for Research, Technology and Development; and the Christian Doppler Research Association through CDL-BOT.

#### REFERENCES

- S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," https://bitcoin.org/bitcoin.pdf, 2008.
- [2] G. Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," https://gavwood.com/paper.pdf, 2014.
- [3] A. Hafid, A. Senhaji Hafid, and M. Samih, "Scaling Blockchains: A Comprehensive Survey," *IEEE Access*, pp. 125 244 – 125 262, 2020.
- [4] A. Chauhan, O. Malviya, M. Verma, and T. Singh Mor, "Blockchain and Scalability," in QRS-C, 2018, pp. 122–128.
- [5] L. Gudgeon, P. Moreno-Sanchez, S. Roos, P. McCorry, and A. Gervais, "SoK: Layer-Two Blockchain Protocols," in FC, 2020, pp. 201–226.
- [6] J. Poon and T. Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments," 2016, https://lightning.network/lightning-network-paper.pdf.
- [7] G. Malavolta, P. Moreno-Sanchez, C. Schneidewind, A. Kate, and M. Maffei, "Anonymous Multi-Hop Locks for Blockchain Scalability and Interoperability," in NDSS, 2019.
- [8] L. Aumayr, O. Ersoy, A. Erwig, S. Faust, K. Hostakova, M. Maffei, P. Moreno-Sanchez, and S. Riahi, "Generalized Channels from Limited Blockchain Scripts and Adaptor Signatures," in *AsiaCrypt*, 2021.
- [9] L. Aumayr, M. Maffei, O. Ersoy, A. Erwig, S. Faust, S. Riahi, K. Hostakova, and P. Moreno-Sanchez, "Bitcoin-Compatible Virtual Channels," in SP, 2021, pp. 901–918.
- [10] L. Aumayr, P. Moreno-Sanchez, A. Kate, and M. Maffei, "Blitz: Secure Multi-Hop Payments Without Two-Phase Commits," in *Usenix Security*, 2021, pp. 4043–4060.
- [11] Z. Avarikioti, E. K. Kogias, R. Wattenhofer, and D. Zindros, "Brick: Asynchronous Incentive-Compatible Payment Channels," in *FC*, 2021, pp. 209–230.

- [12] C. Decker and R. Wattenhofer, "A Fast and Scalable Payment Network with Bitcoin Duplex Micropayment Channels," in SSS, 2015, pp. 3–18.
- [13] S. Dziembowski, L. Eckey, S. Faust, and D. Malinowski, "Perun: Virtual Payment Hubs over Cryptocurrencies," in SP, 2019, pp. 106–123.
- [14] S. Dziembowski, L. Eckey, S. Faust, J. Hesse, and K. Hostáková, "Multi-Party Virtual State Channels," in *EuroCrypt*, 2019, pp. 625–656.
- [15] P. McCorry, S. Bakshi, I. Bentov, S. Meiklejohn, and A. Miller, "Pisa: Arbitration Outsourcing for State Channels," in AFT, 2019, pp. 16–30.
- [16] Z. Avarikioti, O. S. T. Litos, and R. Wattenhofer, "Cerberus Channels: Incentivizing Watchtowers for Bitcoin," in *FC*, 2020, pp. 346–366.
- [17] S. A. K. Thyagarajan, G. Malavolta, F. Schmidt, and D. Schröder, "PayMo: Payment Channels For Monero," *IACR Cryptol. ePrint Arch.*, 2020.
- [18] R. Canetti, Y. Dodis, R. Pass, and S. Walfish, "Universally Composable Security with Global Setup," in *TCC*, 2007, pp. 61–85.
- [19] P. Zappalà, M. Belotti, M. Potop-Butucaru, and S. Secci, "Game Theoretical Framework for Analyzing Blockchains Robustness," *IACR Cryptol. ePrint Arch.*, 2020.
- [20] A. Khosla, E. Schwartz, and A. Hope-Bailie, "Connector Risk Mitigations–Interledger RFCs, 0018," 2019. [Online]. Available: https://interledger.org/rfcs/0018-connector-risk-mitigations/
- [21] G. Avarikioti, F. Laufenberg, J. Sliwinski, Y. Wang, and R. Wattenhofer, "Towards Secure and Efficient Payment Channels," *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/abs/1811.12740
- [22] Z. Avarikioti, L. Heimbach, Y. Wang, and R. Wattenhofer, "Ride the Lightning: The Game Theory of Payment Channels," in *FC*, 2020, pp. 264–283.
- [23] G. Avarikioti, R. Scheuner, and R. Wattenhofer, "Payment Networks as Creation Games," in CBT, 2019, pp. 195–210.
- [24] O. Ersoy, S. Roos, and Z. Erkin, "How to Profit from Payments Channels," in FC, 2020, pp. 284–303.
- [25] C. Badertscher, J. Garay, U. Maurer, D. Tschudi, and V. Zikas, "But why does it Work? A Rational Protocol Design Treatment of Bitcoin," in *Eurocrypt*, 2018, pp. 34–65.
- [26] I. Eyal and E. G. Sirer, "Majority is not Enough: Bitcoin Mining is Vulnerable," in FC, 2014, pp. 436–454.
- [27] Y. Kwon, D. Kim, Y. Son, E. Vasserman, and Y. Kim, "Be Selfish and avoid Dilemmas: Fork after Withholding (faw) Attacks on Bitcoin," in *CCS*, 2017, pp. 195–209.
- [28] A. Sapirshtein, Y. Sompolinsky, and A. Zohar, "Optimal Selfish Mining Strategies in Bitcoin," in FC, 2016, pp. 515–532.
- [29] A. Kiayias, E. Koutsoupias, M. Kyropoulou, and Y. Tselekounis, "Blockchain Mining Games," in *EC*, 2016, pp. 365–382.
- [30] X. Chen, C. Papadimitriou, and T. Roughgarden, "An Axiomatic Approach to Block Rewards," in *AFT*, 2019, pp. 124–131.
- [31] I. Eyal, "The Miner's Dilemma," in IEEE S&P, 2015, pp. 89–103.
- [32] J. Teutsch, S. Jain, and P. Saxena, "When Cryptocurrencies Mine their own Business," in FC, 2016, pp. 499–514.
- [33] E. Heilman, A. Kendler, A. Zohar, and S. Goldberg, "Eclipse Attacks on Bitcoin's Peer-to-Peer Network," in *{USENIX} Security*, 2015, pp. 129–144.
- [34] K. Nayak, S. Kumar, A. Miller, and E. Shi, "Stubborn Mining: Generalizing Selfish Mining and Combining with an Eclipse Attack," in *EuroS&P*, 2016, pp. 305–320.
- [35] Z. Liu, C. Nguyen, W. Wang, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "A Survey on Blockchain: A Game Theoretical Perspective," *IEEE Access*, pp. 47615–47643, 2019.
- [36] K. Chatterjee, A. K. Goharshady, and Y. Velner, "Quantitative Analysis of Smart Contracts," in ESOP, 2018, pp. 739–767.
- [37] K. Chatterjee, A. Goharshady, and A. Pourdamghani, "Probabilistic Smart Contracts: Secure Randomness on the Blockchain," in *ICBC*, 2019, pp. 403–412.
- [38] G. Malavolta, P. Moreno-Sanchez, A. Kate, M. Maffei, and S. Ravi, "Concurrency and Privacy with Payment-Channel Networks," in CCS, 2017, p. 455–471.
- [39] S. Mazumdar, P. Banerjee, A. Sinha, S. Ruj, and B. Roy, "Strategic Analysis to Defend against Griefing Attack in Lightning Network," 2022. [Online]. Available: https://arxiv.org/abs/2203.10533
- [40] P. Banerjee, S. Mazumdar, and S. Ruj, "Griefing-Penalty: Countermeasure for Griefing Attack in Bitcoin-compatible PCNs," *CoRR*, vol. abs/2005.09327, 2020. [Online]. Available: https://arxiv.org/abs/2005.09327

- [41] M. J. Osborne, *Introduction to Game Theory*. Oxford University Press USA, 2004.
- [42] R. Mckelvey, A. McLennan, and T. Turocy, "Gambit: Software Tools for Game Theory," 2005.
- [43] R. Savani and B. von Stengel, "Game Theory Explorer Software for the Applied Game Theorist," *CoRR*, vol. abs/1403.3969, 2014.
  [44] O. Goldreich and Y. Oren, "Definitions and Properties of
- [44] O. Goldreich and Y. Oren, Definitions and Properties of Zero-Knowledge Proof Systems," J. of Cryptology, vol. 7, no. 1, pp. 1–32, 1994.

## APPENDIX

# A. Proof of the Resilience Properties

We restate the results for better readability.

**Lemma III.1** (Resilience Properties). Let  $\sigma \in \mathscr{S}$  be a joint strategy. The following and only the following implications hold.

1) 
$$\sigma$$
 is  $SR_{\subseteq} \Rightarrow \sigma$  is  $SR$ ,  $CR$ ,  $sNE$ .  
2)  $\sigma$  is  $SR \Rightarrow \sigma$  is  $CR$ .  
 $CR$   $SR \subseteq SR_{\subseteq}$ 

*Proof.* We start by showing property (2). Let  $\sigma$  be SR and let  $S = \{s_1, ..., s_j\} \subset N$ ,  $\sigma'_S \in \mathscr{S}_S$  be arbitrary but fixed. Then, for all  $p \in S$  we have  $u_p(\sigma) \ge u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1}, ..., \sigma'_{s_j}/\sigma_{s_j}])$  and thus also  $\sum_{p \in S} u_p(\sigma) \ge \sum_{p \in S} u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1}, ..., \sigma'_{s_j}/\sigma_{s_j}])$ . Hence  $\sigma$  is CR and the implication is proven. For implication (1) we see that SR<sub> $\subseteq$ </sub>  $\Rightarrow$  SR is trivial. If the property is satisfied for every  $S \subseteq N$ , then it is also satisfied for every  $S \subset N$ . By (2) and the transitivity of implication we also get SR<sub> $\subseteq$ </sub>  $\Rightarrow$  CR. For the last implication let  $\sigma$  be SR<sub> $\subseteq$ </sub> and let  $S = \{s_1, ..., s_j\} \subseteq N$ ,  $S \neq \emptyset$  and  $\sigma'_S \in \mathscr{S}_S$  be arbitrary but fixed. Then there exists some  $p \in S$  and by definition all  $p \in S$  satisfy  $u_p(\sigma) \ge u_p(\sigma[\sigma'_{s_1}/\sigma_{s_1}, ..., \sigma'_{s_j}/\sigma_{s_j}])$ . Therefore,  $\sigma$  is SNE.

To prove that no other implication holds between those four concepts, we provide three counterexamples. An overview of which game disproves which implication is given in Table IX.

The three-player NFG  $\Gamma_1$  in Table IV shows a joint strategy  $(H_1, H_2, H_3)$  that is sNE, but not CR (refer to Example III.5). Using the just proven (1) and (2), we get that  $(H_1, H_2, H_3)$  is also not SR nor SR<sub> $\subset$ </sub>.

The three-player game  $\Gamma_2$  (Table V) shows a strategy  $(H_1, H_2, H_3)$  that is CR, but not SR (see Example III.5) and thus also not SR<sub> $\subseteq$ </sub> (property (1)). It is, however, sNE: The only relevant deviation from  $(H_1, H_2, H_3)$  is  $(D_1, H_2, D_3)$ , as it yields a different utility (3, 0, -2) instead of (1, 1, 1). While player  $P_1$  profits in this case, player  $P_3$  does not. One deviating player not profiting suffices for a strong Nash Equilibrium, thus  $(H_1, H_2, H_3)$  is SNE.

To prove the remaining implications incorrect, we consider the two-player game  $\Gamma_3$  in Table VIII. We can easily see that  $(H_1, H_2)$  is not SR<sub> $\subseteq$ </sub>, nor sNE. This is the case, as all players  $\{P_1, P_2\}$  can deviate to play  $(D_1, D_2)$  which yields a strict increase for both. However, since no player profits from deviating alone,  $(H_1, H_2)$  is still SR and CR.

## B. Results of Security Analysis and Their Proofs

In this section all omitted proofs of the results from Section V are provided. Additionally, the results Theorem A.1, Theorem A.2 and Theorem A.3 about edge cases are stated

TABLE	VIII
GAME	Гз.

	$H_2$	$D_2$
$H_1$	(1,1)	(1,1)
$D_1$	(1,1)	(2,2)

$\rightarrow$	SR	$SR_{\subseteq}$	sNE	CR
SR		$\Gamma_3$	$\Gamma_3$	$\checkmark$
$SR_\subseteq$	$\checkmark$		$\checkmark$	$\checkmark$
sNE	$\Gamma_1$	$\Gamma_1$		$\Gamma_1$
CR	$\Gamma_2$	$\Gamma_2$	$\Gamma_3$	

TABLE IX OVERVIEW OF IMPLICATIONS AND COUNTEREXAMPLES.

and proven.

**Theorem V.1** (Weak Immunity of Honest Behavior – (P1)). The terminal histories (H) of honest unilateral closing, and  $(C_h, S)$  of honest collaborative closing of  $G_c(A)$  are weak immune, if the channel balances are higher than the fee required in a revocation transaction, that is if  $a \ge f$  and  $b \ge f$ .

*Proof.* Let  $a, b \ge f$ . For history (*H*), we consider any strategy  $\sigma$ , where *A* chooses *H* after the empty history  $\emptyset$ , *B* chooses *S* after ( $C_h$ ), *P* after (*D*) and *H* after ( $C_c$ ). Such a strategy  $\sigma$  yields terminal history (*H*). If we can show that  $\sigma$  is weak immune, also history (*H*) is weak immune by Definition III.11. Assume, player *A* honestly follows  $\sigma$ , i.e., choosing (*H*), then *B*'s deviation from  $\sigma$  cannot affect the outcome. Thus, *A*'s utility remains non-negative. The other way around, if *B* follows  $\sigma$ , *A* can deviate to any initial action  $C_h$ , *D* or  $C_c$ , player *B*'s utility never drops below 0, by following strategy  $\sigma$ , as  $a \ge f$ . Since honest players cannot get negative utility,  $\sigma$  is weak immune.

Similarly, for  $(C_h, S)$ , we consider any strategy  $\sigma'$ , where A chooses  $C_h$  initially, player B chooses S after  $(C_h)$ , P after (D) and H after  $(C_c)$ . Further, player A takes P after  $(C_h, D)$  and H after  $(C_h, \Im)$ ,  $(C_h, U^+)$  and  $(C_h, U^-)$ . This strategy  $\sigma'$  yields terminal history  $(C_h, S)$ . Deviation of A has the same effects as before, never causing the honest B, who follows  $\sigma$ , negative utility. If B deviates now to one of  $U^+$ ,  $U^-$ ,  $\Im$ , D, or H, honest A, following  $\sigma$ , also never gets negative utility, since  $b \ge f$ . Therefore,  $\sigma'$  and hence history  $(C_h, S)$  are weak immune.

**Theorem V.2** (Incentive-Compatibility – (P2)). If  $a - p_B + d_A \ge f$  and  $b - p_A + d_B \ge f$ , then

- 1) honest unilateral closing (H) is CR, but not practical.
- 2) honest collaborative closing  $(C_h, S)$  is CR. It is practical iff  $c \neq p_A$ .

**Proof.** Let us first prove collusion resilience CR of (H) and  $(C_h, S)$ . As in the previous proof, we only have to find a strategy  $\sigma$  that yields history (H) and another strategy  $\sigma'$  yielding history  $(C_h, S)$ , that are CR, to prove (H) and  $(C_h, S)$  CR, as defined in Definition III.11. Additionally, collusion resilience is defined on strict subsets of players. Thus, in a two-player game, it considers only deviations of single players and since the summation over one value is the value itself,

CR is equivalent to being a Nash Equilibrium in this case. We therefore only have to check whether  $\sigma$  and  $\sigma'$  are Nash Equilibria.

For (*H*), we consider a strategy  $\sigma$ , where player *A* chooses *H* initially, player *B* chooses  $\mathfrak{I}$  after  $(C_h)$ , and  $\mathfrak{I}$  after  $(C_c)$ . Additionally, player B always chooses P after a history (...,D), where the last action was D. Player A takes action H after  $(C_h, \mathfrak{I})$  and  $(C_c, \mathfrak{I})$ . Further, B takes action  $\mathfrak{I}$  after  $(C_{h/c}, \mathfrak{I}, U^{+/-})$  (subgames  $S_1, S_2, S'_1, S'_2$  in Figure 10). For A, we finally assume she takes action H after  $(C_{h/c}, \mathfrak{I}, U^{+/-}, \mathfrak{I})$ . This strategy yields history (*H*). Deviations from  $\sigma$  of player B cannot change the utility, hence in particular cannot increase his utility. Let us consider deviations of player A. A deviation to D at any point in the game, leads to A losing all her funds a, which is a strict decrease in utility. This is the case because in  $\sigma$  player B always chooses P after D. Therefore this option is not a threat. If A deviates to  $C_h$  or  $C_c$  initially, we end up in  $(C_h, \mathfrak{I}), (C_c, \mathfrak{I})$  respectively. Closing honestly (action H) here leads to the same utility as not deviating. Also a deviation to  $\mathfrak{I}$  does not lead to a better utility. The options she has left is taking  $U^+$  or  $U^-$ . Either way, B takes  $\Im$  and leaves A similar choices to before: action H or action  $\Im$ , both of which do not yield a better utility for her. Since no player can increase their utility by deviating from  $\sigma$ , it is a Nash Equilibrium, and hence (H) is too.

To show that  $(C_h, S)$  is a Nash Equilibrium, we consider a strategy  $\sigma'$ , where *A* picks  $C_h$  initially, *B* chooses *S* after  $(C_h)$ , *P* after (D) and *H* after  $(C_c)$ . Further, let *A* pick *P* after  $(C_h, D)$ , *H* after  $(C_h, \mathfrak{I})$  and  $(C_h, U^{+/-})$  (subgames  $S_3, S_4$  in Figure 11). This strategy  $\sigma'$  has terminal history  $(C_h, S)$ . A deviation of player *B*, results in either the same utility (choosing action  $\mathfrak{I}$ ,  $U^+$ , or  $U^-$  after  $(C_h)$  and having *A* taking *H*) or in strictly worse utility (choosing *H* or *D*, where *A* takes *P*). Every other deviation has no impact on the resulting history. Similarly, player *A* cannot profit from deviating. Choosing action *H* or *D* initially, leads to a strict loss, as *B* plays *P*, whereas taking action  $C_c$  yields the same utility for *A* (as *B* will take action *H*). Every other deviation has no impact on the history. Hence, no player can increase their utility by deviating, which makes  $\sigma'$  and therefore  $(C_h, S)$  a Nash Equilibrium.

To prove the practicality properties, we compute all subgame perfect equilibria of  $G_c(A)$ . We compute subgame perfect equilibria bottom-up. That is, we start comparing the utility of subtrees with leaves only. In  $G_c(A)$ , these are for example the subgames after history  $(C_h, \Im, D)$  or  $(C_c, D)$ . For the latter, A is the player to choose the action. To compute the subgame perfect equilibrium, we have to compare all possible utilities for A after  $(C_c, D)$ . We then replace this internal node labelled A, by the utility that yields the best value for A and proceed until we reach the root. If there is no single best choice for a player, then all actions resulting in best utility have to be considered. Applying this procedure to the subgames  $S_1$ - $S_4$ and  $S'_1$ - $S'_4$  we get subgame perfect terminal history  $(\mathfrak{A}, H)$  with utility  $(\rho + \alpha - \varepsilon, \rho + \alpha)$  for  $S_1$ . For  $S_2$  we get terminal history (S) yielding  $(\alpha, \alpha)$  and  $(\mathfrak{I}, H)$ , yielding  $(\alpha - \varepsilon, \alpha)$ . For  $S_3$  and  $S_4$  it is  $(\mathfrak{I}, S)$  with  $(\alpha, \alpha)$ . The subgame  $S'_1$  has practical history

 $(\mathfrak{I}, H)$ , with  $(\alpha - \varepsilon, \alpha)$  if  $c > p_A$ ,  $(\mathfrak{A}, \mathfrak{I}, S)$  with  $(\rho + \alpha, \rho + \alpha)$ if  $c = p_A$  and  $(\mathfrak{A}, H)$  with  $(\rho + \alpha - \varepsilon, \rho + \alpha)$  if c < p. The subgame  $S'_2$  has practical history  $(\mathfrak{I}, H)$ , yielding  $(\alpha - \varepsilon, \alpha)$ . For  $S'_3$  and  $S'_4$  in Figure 11 we get  $(\mathfrak{I}, H)$  with  $(\alpha, \alpha - \varepsilon)$ and additionally for  $S'_3$ , if c = p, we also have  $(\mathfrak{A}, S)$  yielding  $(\rho + \alpha, \rho + \alpha)$ . All of these results are based on the facts  $a - p_B + d_A \ge f$  and  $b - p_A + d_B \ge f$ , since this causes the revocation transaction always to be better than ignoring the dishonest unilateral closing attempt.

Based on these preliminary results, we can now compute the subgame perfect equilibria for  $G_c(A)$  considering multiple practical histories and case splits as stated: If  $c = p_A$ , then  $(C_c, U^+, \mathfrak{A}, S)$  and  $(C_c, \mathfrak{I}, U^+, \mathfrak{A}, \mathfrak{I}, S)$  are practical, both yielding  $(\rho + \alpha, \rho + \alpha)$ . If  $c > p_A$ , then the histories  $(C_h, S), (C_h, U^+, \mathfrak{I}, S), (C_h, U^-, \mathfrak{I}, S)$  and  $(C_h, \mathfrak{I}, U^-, S)$  all leading to  $(\alpha, \alpha)$  are practical, as well as terminal history  $(C_h, \mathfrak{I}, U^+, \mathfrak{I}, H)$ , yielding  $(\rho + \alpha - \varepsilon, \rho + \alpha)$ . For  $c < p_A$ , all the histories and their utilities from  $c > p_A$  are practical. Additionally  $(C_c, \mathfrak{I}, U^+, \mathfrak{A}, H)$  is subgame perfect in this case and also results in utility  $(\rho + \alpha - \varepsilon, \rho + \alpha)$ .

This shows, that (*H*) is never practical and ( $C_h$ , *S*) is practical if and only if  $c \neq p_A$ .

### 1) Results without Updates:

**Corollary 1.** If there exists an old channel state  $(a+d_A, b-d_A)$ , with  $a+d_A < f$ , then neither history (H) nor  $(C_h, S)$  is weak immune nor practical, but CR.

*Proof.* We fix the old distribution state such that the difference  $d_A$  to the latest state is the value of *A*'s dishonest closing attempt in the closing game. As  $a + d_A < f$  implies a < f, Theorem V.5 applies. Therefore, neither (*H*) nor ( $C_h$ , *S*) are weak immune.

In order to show that they are also not practical, we prove instead, that the only practical history is  $(D, \Im)$ . Since  $a + d_A < f$ ,  $\Im$  is the best choice for *B* after (D),  $(C_h, \Im, D)$  and  $(C_c, \Im, D)$ . Consequently, *A* will choose *D* after  $(C_h, \Im)$  and  $(C_c, \Im)$ . If now  $b + d_B \ge f$ , then *A*'s best choice is *P* after  $(C_h, D)$  and  $(C_c, D)$ . Thus, *B* will take *S* after  $(C_h)$  and *H* after  $(C_c)$ . In the other case,  $b + d_b < f$ , *A*'s best option is  $\Im$  after  $(C_h, D)$  and  $(C_c, D)$ , thus *B*'s best choice after  $(C_h)$  and  $(C_c)$  is *D*, which yields a negative utility for *A*. Therefore, in both cases *A*'s only subgame perfect action is *D*. Hence,  $(D, \Im)$  is the unique subgame perfect history.

For CR, we show instead that there exist extensions (Definition III.11)  $\sigma$  of (*H*) and  $\sigma'$  of (*C<sub>h</sub>*, *S*) that are Nash Equilibria. Let  $\sigma$  be the strategy, where *A* chooses *H*, everyone chooses *P* after a dishonest closing attempt *D*, *B* chooses  $\Im$  after (*C<sub>h</sub>*) and (*C<sub>c</sub>*) and *A* chooses *H* after (*C<sub>h</sub>*,  $\Im$ ) and (*C<sub>c</sub>*,  $\Im$ ). Then, player *B*'s deviations have no impact, thus cannot not increase his utility, and player *A*'s deviations either lead to the same utility as  $\sigma$ , or to the strictly worse utility –*a*. Anyway, no player can deviate to increase their utility and therefore  $\sigma$ and thus (*H*) is CR. To prove (*C<sub>h</sub>*, *S*) is CR, we consider the strategy  $\sigma'$ , which is the same as  $\sigma$ , except *A* initially chooses *C<sub>h</sub>* and *B* chooses *S* after (*C<sub>h</sub>*). A deviation of *A* either leads to utility  $\alpha - \varepsilon$  for her, which is worse than  $\sigma$ 's utility, or to utility -a, which is even worse. For *B*, a deviation either leads to the same utility  $\alpha$  (taking  $\Im$  after  $(C_h)$ ), to a slightly worse  $\alpha - \varepsilon$  (choosing *H* after  $(C_h)$ ) or to the way worse -b(*D* after  $(C_h)$ ). Every other deviation has no impact on the history. Hence, as nobody profits from deviating,  $\sigma'$  is also a Nash Equilibrium.

We present an additional theorem, discussing the case where player B has little funds left in the channel. Since the roles of player A and B are arbitrary, it is of little importance because the results give stronger security guarantees as for the case where A has a low balance. Nevertheless, we state it for the sake of completeness.

**Theorem A.1.** If there exists an old state with  $b + d_B < f$ , but  $a \ge f$ , then

1) (H) is secure.

2)  $(C_h, S)$  is not practical, not weak immune, but CR.

*Proof.* To prove (1), we start by showing weak immunity for a strategy  $\sigma$  with history (*H*). Consider  $\sigma$ , where *A* takes action *H* initially, player *B* chooses *P* after (*D*), *S* after (*C<sub>h</sub>*) and *H* after (*C<sub>c</sub>*). Then  $\sigma$  and thus (*H*) is weak immune, because *B*'s deviations have no impact on the history and *A*'s deviations can never bring *B*'s utility below zero.

Next, we prove the practicality of (H) by computing all subgame perfect equilibria. Since  $a \ge f$ , the subgame perfect choice after (D),  $(C_h, \Im, D)$  and  $(C_c, \Im, D)$  is P. Thus, A chooses H after  $(C_h, \Im)$  and  $(C_c, \Im)$ . Due to  $b + d_B < f$ , A's best option after  $(C_h, D)$  and  $(C_c, D)$  is  $\Im$ . Hence B's unique subgame perfect choice after  $(C_c)$  and  $(C_h)$  is D. Thus, A's only best response is H. Therefore, (H) is the only practical history. As practicality implies CR in our case, (H) is secure.

For (2), we just showed that  $(C_h, S)$  cannot be practical. Additionally,  $(C_h, S)$  is not weak immune, since *B* could deviate to *D* after  $(C_h)$ , in which case *A* gets negative utility for sure, because of  $b + d_B < f$ .

Finally, we consider the strategy  $\sigma'$ , with history  $(C_h, S)$ , where A initially chooses  $C_h$ , B chooses S after  $(C_h)$ , both take P in case of a dishonest unilateral closing attempt D, B takes H after  $(C_c)$ , similarly A takes H after  $(C_h, \mathfrak{I})$  and  $(C_c, \mathfrak{I})$ . Using similar argumentation as before, we conclude that any deviation of a player leads a utility as most as good as  $\sigma'$  for them, but never better. Hence,  $\sigma'$  is a Nash Equilibrium yielding terminal history  $(C_h, S)$ .

2) Results for Edge Cases: So far, we only considered cases where both balances a and b were strictly greater than zero. This is not necessarily the case. Therefore, we consider these cases here. In the first case, a = 0, B cannot close dishonestly, as there is no old state that increases his balance. The corresponding simplified game is presented in Figure 8.

If b = 0 (Figure 9), player A cannot close dishonestly, as she cannot take any money from B. Thus, both dishonest unilateral closing D and proposing an unfair split in a collaborative closing attempt  $C_c$  are not possible.

Finally, we present results about the two edge cases.

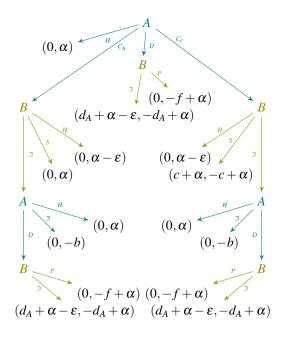


Fig. 8. Closing game  $G_c(A)$  with a = 0.

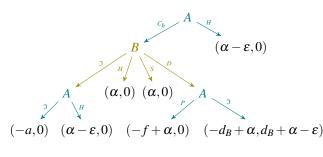


Fig. 9. Closing game  $G_c(A)$  with b = 0.

**Theorem A.2.** If a = 0 and b > 0 (Figure 8), then only histories that involve an explicit cheating attempt are weak immune. Additionally, (H) and ( $C_h$ , S) are practical if and only if  $d_A \ge f$  in every previous state  $(d_A, b - d_A)$ . In any case they are CR.

*Proof.* We first show that only histories that involve an explicit cheating attempt can be weak immune. Let us consider a history h without a D or  $C_c$  action, then A does not initially choose D in h. However, if A deviates to D, then B's utility is negative. Thus, any such history h is not weak immune.

To show both (H) and  $(C_h, S)$  are CR, it suffices to show they are Nash Equilibria as before. We therefore consider any strategy  $\sigma$ , where A initially chooses H, player B chooses P after (D), S after  $(C_h)$  and H after  $(C_c)$ . Further, player A takes action H after  $(C_h, \Im)$ . The strategy  $\sigma$  yields history (H). No matter how player A deviates, she always gets utility 0, as she does in  $\sigma$ . Thus, she has no incentive to deviate. Since player B's deviations cannot change the history, also he has no incentive to do so. Therefore,  $\sigma$  and hence (H) is a Nash Equilibrium. Adapting  $\sigma$ , by making A first choice  $C_h$  we get strategy  $\sigma'$  which leads to history  $(C_h, S)$ . As before, A's utility stays 0 no matter how she deviates from  $\sigma'$ . Also player *B* cannot improve his utility by changing strategy. Hence, also  $\sigma'$  and therefore  $(C_h, S)$  is a Nash Equilibrium.

Towards practicality, we now compute all subgame perfect equilibria. Let  $d_A \ge f$ . In which case *P* is the subgame best choice for *B* after (*D*), ( $C_h, \Im, D$ ) and ( $C_c, \Im, D$ ). Further, after history ( $C_c, \Im$ ), *S* it is never a best option for *B*, because it is strictly dominated by *H*. Therefore, *A* will get utility zero in any case. This makes (*H*) a practical history. Similarly for ( $C_h, S$ ), since *S* is subgame perfect for *B* after ( $C_h$ ).

If now  $d_A < f$ , then  $\mathfrak{I}$  is subgame perfect for B after D. Thus, with similar argumentation as before,  $(D, \mathfrak{I})$  is the only practical history.

## **Theorem A.3.** If a > 0 and b = 0 (Figure 9), then

- 1) (H) is secure.
- 2)  $(C_h, S)$  is not weak immune, but CR. It is practical iff  $d_B \ge f$  in every previous state  $(a d_B, d_B)$ .

*Proof.* We prove (1.) first. The history (H) is weak immune, as *B*'s strategy does not effect the history, since *A*'s initial choice has to be *H*. Further, *A*'s deviation is irrelevant for *B*, as he can never get negative utility in this game.

Practicality of (*H*). We compute subgame perfect equilibria. After history  $(C_h, D)$  the subgame perfect choice of *A* depends on whether  $d_B \ge f$ . In any case, *D* is subgame perfect for *B* after history  $(C_h)$ . If *A* chose *P*, then it is as good as any other choice, yielding 0, otherwise it is the only best option resulting in a positive utility. Thus, *A* either gets  $-f + \alpha$  or  $-d_B + \alpha$  if she chooses  $C_h$ , both of which is negative. Hence *A*'s subgame perfect and therefore practical choice is *H*, yielding the history (*H*).

The fact that (H) is CR follows from practicality. This shows that (H) is secure, if b = 0.

For (2), we start showing  $(C_h, S)$  is not weak immune. We consider any strategy  $\sigma'$  yielding the history  $(C_h, S)$ . Assume now, *B* deviates to *D* after  $(C_h)$ , then no matter what *A*'s choice is, she will get a negative utility, thus  $(C_h, S)$  is not weak immune.

The collusion resilience of  $(C_h, S)$ , can be shown by considering a strategy  $\sigma'$  with history  $(C_h, S)$ , where we additionally fix that *A* chooses *P* after  $(C_h, D)$ . Then *B* has no incentive to deviate as he always gets utility 0, and *A* has no incentive as  $\alpha$ , which is her utility in  $\sigma'$ , is the best possible outcome for her.

To finally show that  $(C_h, S)$  is practical iff  $d_B \ge f$ , we consider *A*'s choice after  $(C_h, D)$ . The option *P* is subgame perfect iff  $d_B \ge f$ . Thus, *S* is subgame perfect for *B* iff  $d_B \ge f$ . For  $d_B < f$ , *D* is the better option for *B*, yielding  $(-d_B + \alpha, d_B + \alpha - \varepsilon)$ . Therefore  $C_h$  is subgame perfect for *A* iff  $d_B \ge f$ , in which case the resulting history is  $(C_h, S)$ .

The weak immunity result of (H) might be misleading, as *B* can actually close dishonestly immediately (before *A* takes action). This is not represented here, but in  $G_c(B)$ , which is analog to  $G_c(A)$  but with swapped roles.

$$(\alpha, \alpha - \varepsilon) \xrightarrow{H} B$$

$$(-a, a - f + \alpha) \xrightarrow{P} B \xrightarrow{D} A \xrightarrow{S} p$$

$$(-a, -b) (\alpha - \varepsilon, \alpha)$$

$$(-a, -b) (\alpha - \varepsilon, \alpha)$$

$$(-d_B + \alpha, d_B + \alpha - \varepsilon) \xrightarrow{P} A$$

$$(b - f + \alpha, -b) \xrightarrow{H} (\rho + \alpha - \varepsilon, \rho + \alpha)$$

$$(\rho + \alpha - \varepsilon, \rho + \alpha) \xrightarrow{P} (-a - x + \rho, a + x - f + \rho + \alpha)$$

$$(y - x + \rho + \alpha, -y + x + \rho + \alpha)$$

$$(p + \alpha, \rho + \alpha - \varepsilon)$$

$$(b - x - f + \rho + \alpha, -b + x + \rho)$$

$$(-d_B + \rho + \alpha, d_B + \rho + \alpha - \varepsilon)$$

Fig. 10. Subgames  $S_{1,2}$ ,  $S'_{1,2}$  with Update  $(a,b) \mapsto (a+x,b-x)$ .

#### C. Subgames of the Closing Game

In the following all the subgames needed for the closing game  $G_c(A)$  are defined. The subgames  $S_{1,2}$  and  $S'_{1,2}$  in Figure 10 cover the case where a channel update is proposed by A, although A has already signed a collaborative closing attempt. In  $S_1$  the closing attempt was honest, hence y = 0and the update is from channel state (a,b) to  $(a+p_A,b-p_A)$ , hence  $x = p_A$ . In  $S_2$  also y = 0 the suggested update is  $(a - p_B, b + p_B)$ , thus  $x = -p_B$ . In  $S'_{1,2}$  the closing attempt was dishonest, therefore y = c. The channel updates are as before, thus  $x = p_A$  for  $S'_1$  and  $x = -p_B$  for  $S'_2$ . Similarly, subgames  $S_{3,4}$  and  $S'_{3,4}$  in Figure 11 cover the case where a channel update is proposed by B, although A has already signed a collaborative closing attempt. As in the first case, we have y = 0 for the honest closing attempt in  $S_{3,4}$  and y = c for dishonest collaborative closing in  $S'_{3,4}$ . Further in  $S_3$  and  $S'_3$ , the proposed update is  $(a + p_A, b - p_A)$ , hence  $x = p_A$ , whereas in  $S_4$  and  $S'_4$  it is  $(a - p_B, b + p_B)$ , thus  $x = -p_B$ .

## D. Subgames of the Routing Game

In this section, one subgame of each type is detailed. First, subgame  $S_1$  in Figure 12 describes the case where player *A* locks an amount of money in the HTLC which deviates from the expected m+3f. The action  $L_w$  means that the subsequent player follow along and forward the deviation of -w to player *B*. Subgame  $S_2$  in Figure 13 illustrates the case that player  $E_1$  creates her own secret and uses its hash *z* as the lock of her

$$(\alpha - \varepsilon, \alpha) \xleftarrow{H} A$$

$$(b - f + \alpha, -b) \xleftarrow{5} A \xleftarrow{5} B$$

$$(-d_{B} + \alpha, d_{B} + \alpha - \varepsilon) \xrightarrow{H} (y + \alpha, -y + \alpha)$$

$$(-a, -b) (\alpha, \alpha - \varepsilon)$$

$$(d_{A} + \alpha - \varepsilon, -d_{A} + \alpha) \xrightarrow{5} P$$

$$(-a, a - f + \alpha) \xrightarrow{5} P$$

$$(p + \alpha, \rho + \alpha - \varepsilon) \xrightarrow{F} A$$

$$(p + \alpha, \rho + \alpha - \varepsilon) \xrightarrow{F} A$$

$$(-d_{B} + \rho + \alpha, d_{B} + \rho + \alpha - \varepsilon) \xrightarrow{F} A$$

$$(-a - x + \rho, -b + x + \rho) \xrightarrow{F} A$$

$$(-a - x + \rho, -b + x + \rho) \xrightarrow{F} A$$

$$(-a - x + \rho, -b + x + \rho) \xrightarrow{F} B$$

$$(-a - x + \rho, -b + x + \rho) \xrightarrow{F} B$$

Fig. 11. Subgames  $S_{3,4}$ ,  $S'_{3,4}$  with Update  $(a,b) \mapsto (a+x,b-x)$ .

HTLC. Action  $L_z$  describes the reusing of hash lock z in the next HTLCs. Lastly, subgame  $S_3$  in Appendix D handles the case, where player I uses a time-out  $t_3$  of the HTLC which is later than the previous ones  $t_1$  and  $t_2$ , thereby neglecting the decreasing ordering of time outs. In Appendix D, the action  $U_{t_2}$  means unlocking the HTLC before  $t_2$  times out, thus enabling the other players to unlock too. The action  $U_{>t_1}$  stands for unlocking after both  $t_1$  and  $t_2$  timed out, therefore I and  $E_1$  cannot unlock their respective HLTCs any more. Finally, action  $U_{[t_2,t_1]}$  means unlocking after  $t_2$  has timed out, but the HTLC with time-out  $t_1$  can still be unlocked.

$$(w+\rho,f,m-w+2f-\varepsilon,-m+w,-w+\rho) \quad (m+3f+\rho-\varepsilon,-m+w-2f,m-w+2f-\varepsilon,-m+w,-w+\rho)$$

$$S_{5} \quad S_{5} \quad (w+\rho,m-w+3f-\varepsilon,-\varepsilon,-m+w,-w+\rho)$$

$$E_{1} \quad (w+\rho,m-w+3f-\varepsilon,-\varepsilon,-m+w,-w+\rho)$$

$$E_{1} \quad (w+\rho,m-w+3f-\varepsilon,-\varepsilon,-\varepsilon,-m+w,-w+\rho)$$

$$(w+\rho,m-w+3f-\varepsilon,-\varepsilon,-\varepsilon,-m+w,-w+\rho)$$

$$(-\varepsilon,-\varepsilon,0,0,0) \quad (s_{7} \quad (m+3f+\rho-\varepsilon,-\varepsilon,-\varepsilon,-m+w,-w+\rho))$$

$$(-\varepsilon,-\varepsilon,-\varepsilon,-\varepsilon,0,0) \quad (s_{7} \quad (m+3f+\rho-\varepsilon,-\varepsilon,-\varepsilon,-m+w,-w+\rho))$$

$$(-\varepsilon,-\varepsilon,-\varepsilon,-\varepsilon,-\varepsilon,-m+w,-f,f,-w+\rho)$$

$$(w+\rho,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

$$(w+\gamma,m-w+3f-\varepsilon,-m+w-f,f,-w+\rho)$$

Fig. 12. Subgame 
$$S_1$$
 with locked amount  $m - w + 3f$ .

$$S_{8} S_{8}$$

$$V_{L_{T}}/L_{H} S_{9} S_{9} (m+3f+\rho-\varepsilon,-\varepsilon,-\varepsilon,-m,\rho)$$

$$I V_{L_{A}} E_{2} V_{L_{A}} E_{2} V_{L_{A}} E_{1} S_{S_{B}} S_{B} V_{A} S_{B} V_{A$$

Fig. 13. Subgame  $S_2$  with used hash lock z.

$$S_{12} S_{10} \mathfrak{S}_{22} \mathfrak{S}_{\overline{s}_{E_1}} E_2 \xrightarrow{\tau} (m+3f+\rho-\varepsilon,-\varepsilon,-m-f,f,\rho)$$

$$E_2 \mathfrak{S}_{11} \mathfrak{S}_{21} \mathfrak{S}_{21} \mathfrak{S}_{22} \mathfrak{S}_{21} \mathfrak{S}$$

Fig. 14. Subgame  $S_3$  with time-out ordering  $t_3 > t_1 > t_2 > t_4$ .