Abstract

We give a proof of hardness amplification for a new type of Weakly Verifiable Puzzles that generalize the previous approaches.

An important question in cryptography is whether it is possible to build cryptographic construction that is strongly hard from a one that is only weakly hard. The well known result of Yao states that it is feasible to build a strong one-way function from functions that are only weakly hard.

Weakly Verifiable Puzzles can be seen as problems where a solution cannot be efficiently verified by a party that solves a puzzle. The problem of hardness amplification for various variants of Weakly Verifiable Puzzles has been studied in a series of papers [CHS04, DIJK09, HS10]. Dodis, Impagliazz, Jaiswal, and Kabanets [DIJK09] show that it possible to amplify hardness for Dynamic Weakly Verifiable Puzzles, which covers the case of cryptographic primitives like message authentication codes and signature schemes. Holenstein and Schoenebeck [HS10] prove a similar result for Interactive Weakly Verifiable Puzzles that embraces the case of constructions like bit commitment protocols.

In this thesis we introduce the notion of *Dynamic Interactive Weakly Verifiable Puzzles* that combines the previous definitions of dynamic and interactive puzzles. Furthermore, we prove that if there is no efficient probabilistic algorithm that solves a single puzzle with probability higher than δ, then there is no efficient algorithm that solves k independent instances of puzzles with probability higher than an algorithm that solves each of the k puzzles independently with probability δ and where an arbitrary binary monotone function is used to decide which of the k puzzle should be solved.

Contents

Contents **iii**

1 Introduction **1**
1.1 Hardness Amplification Theorems 1
1.2 Weakly Verifiable Puzzles 2
1.3 Contribution of the Thesis 3
1.4 Organization of the Thesis 3

2 Notation and Definitions **5**

3 Dynamic Interactive Weakly Verifiable Puzzles **9**
3.1 The Definition 9
3.2 The Hardness Amplification Theorem 10

4 Examples of Weakly Verifiable Cryptographic Primitives **13**
4.1 Message Authentication Codes 13
4.2 Public Key Signature Schemes 14
4.3 Bit Commitment Protocols 16
4.4 Automated Turing Tests 18

5 Previous Results **21**
5.1 Weakly Verifiable Puzzles 21
5.1.1 The definition 22
5.1.2 The hardness amplification 22
5.2 Dynamic Weakly Verifiable Puzzles 26
5.2.1 The definition 26
5.2.2 The hardness amplification theorem 28
5.3 Interactive Weakly Verifiable Puzzles 31
5.3.1 The definition 31
5.3.2 The hardness amplification theorem 31

6 Hardness Amplification for Dynamic Interactive Weakly Verifiable Puzzles **33**
6.1 Main Theorem 33
6.1.1 The intuition 35
6.1.2 Domain partitioning 36
6.1.3 The hardness amplification proof for partitioned domains 41
6.1.4 Putting it together 53
6.2 Discussion 54
6.2.1 Optimality of the result 54

A Appendix **55**
A.1 Concentration Bounds 55
A.2 The Proof of Lemma 6.4 Under the Simplifying Assumptions . 56

Bibliography **57**

Chapter 1

Introduction

1.1 Hardness Amplification Theorems

An important question in cryptography is whether it is possible to turn a certain problem that is only mildly hard into a problem that solving is substantially harder. An example is the well known hardness amplification result by Yao [Gol00] which states that it is possible to build a strong one–way function from a function that is only weakly one-way.

In this Thesis we study hardness amplification of *weakly verifiable puzzles*. This notion generalize numerous basic cryptographic primitives like signature schemes, message authentication codes, or bit commitment protocols, just to name a few. The characteristic property of a weakly verifiable puzzle is that one does not require that verifying the correctness of a solution by a puzzle solver is efficient[1]. On the other hand, we assume that an algorithm that generates an instance of a weakly verifiable puzzle has access to secret information which makes the task of verifying correctness of a solution easy. An example is CAPTCHA where a problem of checking whether a solution is correct by a solver is comparably hard to finding a correct solution. Whereas, the poser knows the text from which the CAPTCHA is generated and can trivially verify whether a solution is correct.

There are two main approaches to amplify hardness that base on combining several weakly hard problems into a construction that is strongly hard. First, one can try to use *sequential repetition* where a protocol is repeated in rounds that start one after another. It has been observed that sequential repetition amplifies hardness of weakly verifiable puzzles [VABHL03]. However, this approach may be inefficient as it increases the number of rounds creating additional communication burden.

[1] We note that the word *weakly* in the context of weakly verifiable puzzles is used in a different meaning than for a weak one–way function.

A more useful technique both from the practical and theoretical point is to amplify hardness using *parallel repetition* where several independent weakly hard problems are sent in one round. However, it has been shown that in some settings parallel repetition may fail to amplify hardness [BIN97]. Therefore, some studies are required to prove that parallel repetition amplifies hardness of weakly verifiable puzzles.

Proving hardness amplification requires showing that the following implication holds

$$A \implies B$$

where A is a statement that a problem P is hard, and B denotes that a problem Q is hard. It turns out that it is often sensible to consider a logically equivalent statement

$$\neg B \implies \neg A.$$

Thus, under the assumption that Q is easy, we try to prove that P is easy. This approach is used in the Thesis. More precisely, we assume existence of an algorithm that successfully solves the parallel repetition of weakly verifiable puzzles with substantial probability, and under this assumption we construct an algorithm that solves a single puzzle with substantial probability.

1.2 Weakly Verifiable Puzzles

Breaking security is often defined as a game in which an adversary has to solve a certain problem. It turns out that for some cryptographic primitives a task of winning a game by an adversary is equivalent to solving a weakly verifiable puzzle.

The proof of Yao for amplifying hardness of one–way functions relies to the great extent on the fact that it is possible for an adversary to easily verify correctness of a solution. Therefore, to show hardness amplification for weakly verifiable puzzles a different approach has to be developed.

Weakly verifiable puzzles have been introduced and studied by Cannetti, Halevi, and Steiner [CHS04]. They show how to amplify hardness of the parallel repetition of weakly verifiable puzzles.

In some cryptographic settings it is enough to solve only a fraction of puzzles included in the parallel repetition of weakly verifiable puzzles. A significant example are hard artificial intelligence problems like CAPTCHAs where the goal is to distinguish a human from a computer program. A human has on average an advantage over computer programs in recognizing a distorted text. However, we can not exclude a situation where he or she also makes mistakes.

The proof of hardness amplification where a threshold function is used to decide what fraction of weakly verifiable puzzles has to be solved correctly in order to successfully solve the parallel repetition of weakly verifiable puzzles is given by Impagliazzo, Jaiswal, and Kabanets [IJK07]. A similar proof by Dodis, Impagliazzo, Jaiswal, and Kabanets [DIJK09] additionally takes into account settings where an adversary can ask a limited number of queries that verify correctness of a solution. Furthermore, an adversary can obtain limited number of hints. The puzzles defined in this way captures some standard security definitions of cryptographic primitives like message authentication codes and signature schemes.

Holenstein and Schoenebeck [HS10] give a more natural proof for hardness amplification of weakly verifiable puzzles where only a fraction of puzzles has to be solved correctly. Furthermore, the puzzles considered by them generalize to games such as breaking the binding property of bit commitment protocols where an instance of a puzzle is generated in an interactive phase.

1.3 Contribution of the Thesis

In this Thesis we apply the proof technique presented in [HS10] in the context of weakly verifiable puzzles as in [DIJK09]. As a result we prove that it is possible to amplify hardness of weakly verifiable puzzles where an adversary can ask a limited number of hint and verification queries, an instance of a puzzle is created in an interactive protocol, and a monotone binary function is used to decide which puzzles of the parallel repetition of weakly verifiable puzzles have to be successfully solved (this generalize a case where only a fraction of puzzles has to be solved successfully).

1.4 Organization of the Thesis

In Chapter 2 we lay down notation and terminology used in the Thesis.

Next, in Chapter 3 we define a *dynamic interactive weakly verifiable puzzle* and give an overview of cryptographic primitives that are generalized by this notion. Furthermore, we give an outline of earlier studies of weakly verifiable puzzles and compare it with the results contained in this Thesis.

Finally, in Chapter 6 we formulate and prove the main theorem of this Thesis. Namely, we show that it is possible to amplify hardness of dynamic interactive weakly verifiable puzzles.

Chapter 2

Notation and Definitions

In this chapter we set up the notation and introduce definitions used in the thesis.

We denote Boolean circuits using capital letters from the Greek or English alphabet. Furthermore, we use calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$ to denote sets and small letters to denote set elements. The randomness used by algorithms and circuits is denoted by small letters from the Greek alphabet $\rho, \delta, \ldots$.

Bitstrings and tuples. We denote a n-tuple $(x_1, \ldots, x_n)$ by $x^{(n)}$. The i-th element of $x^{(n)}$ is denote by $x_i^{(n)}$ Furthermore, for tuples $x^{(n)}$, $y^{(k)}$ we use $x^{(n)} \circ y^{(k)}$ to denote the concatenation of $x^{(n)}$ and $y^{(k)}$ which results in a tuple $(x_1, \ldots, x_l, y_1, \ldots, y_k)$. We denote a bitstring $(b_1, \ldots, b_n) \in \{0,1\}^n$ as a n-tuple of bits.

Functions. We write $p(\alpha_1, \ldots, \alpha_n)$ to denote a polynomial on variables $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$. We call a function $f : \mathbb{N} \to \mathbb{R}$ *negligible* if for every $p(n)$ there exists $n_0 \in \mathbb{N}$ such that for all $n \in \mathbb{N} : n > n_0$ we have

$$f(n) < \frac{1}{p(n)}.$$

On the other hand, we say that a function $f : \mathbb{N} \to \mathbb{R}$ is *non-negligible* if there exists $p(n)$ such that for some $n_0 \in \mathbb{N}$ and for all $n \in \mathbb{N} : n > n_0$ we have

$$f(n) \geq \frac{1}{p(n)}.$$

We note that it is possible that a function $f : \mathbb{N} \to \mathbb{R}$ is neither negligible nor non-negligible.

We say that a function f is *efficiently computable* if there exists a polynomial time algorithm computing f.

Definition 2.1 (Binary monotone function.) For $n \in \mathbb{N}$ we say that $g : \{0,1\}^n \to \{0,1\}$ is a *binary monotone function* if for any two bitstrings $(b_1, \ldots, b_n) \in \{0,1\}^n$ and $(b'_1, \ldots, b'_n) \in \{0,1\}^n$ and sets $\mathcal{I} := \{i \in \{1, \ldots, n\} : b_i = 1\}$ and $\mathcal{I}' := \{i \in \{1, \ldots, n\} : b'_i = 1\}$ it holds that if $\mathcal{I} \subseteq \mathcal{I}'$ and $g(b_1, \ldots, b_n) = 1$ then $g(b'_1, \ldots, b'_n) = 1$.

Algorithms and Circuits. We define a *probabilistic circuit* as a Boolean circuit $C_{m,n} : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^*$ and write $C_{m,n}(x; \rho)$ to denote a probabilistic circuit that takes as input $x \in \{0,1\}^m$ and the randomness $\rho \in \{0,1\}^n$. If a probabilistic circuit takes as input only the randomness, we slightly abuse the notation and write $C_n(r)$. We make sure that it is clear from the context that probabilistic circuits that take as input only the randomness are not confused with deterministic Boolean circuits. We use $\{C_n\}_{n \in \mathbb{N}}$ to denote a family of probabilistic circuits. For a (probabilistic) circuit C we write $Size(C)$ to denote the total number of vertices of C. W define a *family of (probabilistic) polynomial size circuits* as a family of (probabilistic) circuits where the size of a circuit is polynomial in the number of input vertices.

Sometimes we talk about interactive protocols consisting of two phases where two probabilistic circuits interact with each other. In this settings we define a *two-phase circuit* $C := (C_1, C_2)$ as a circuit where in the first phase a circuit C_1 is used and in the second phase a circuit C_2. If C_1 and C_2 are probabilistic circuits we write $C(\delta) := (C_1, C_2)(\delta)$ to denote that in both phases C_1 and C_2 use the same randomness $\delta \in \{0,1\}^*$.

We write $Time(A)$ to denote the number of steps it takes to execute an algorithm A as a function of the length of input A takes. We say that A runs in the *polynomial time* if there exists $p(n)$ such that $Time(A)$ is bounded by $p(n)$ where $n \in \mathbb{N}$ denotes the length of the input that A takes.

Similarly as for a probabilistic circuit we often write the randomness used by a probabilistic algorithm explicitly.

Probabilities and distributions. For a finite set $\mathcal{R}$ we write $r \xleftarrow{\$} \mathcal{R}$ to denote that r is chosen from $\mathcal{R}$ uniformly at random. For $\delta \in \mathbb{R} : 0 \leq \delta \leq 1$ we write μ_δ to denote the Bernoulli distribution where outcome 1 occurs with probability δ and 0 with probability $1 - \delta$. Moreover, we use μ_δ^k to denote the probability distribution over k-tuples where each element of a k-tuple is drawn independently according to μ_δ. Finally, let $u \leftarrow \mu_\delta^k$ denote that a k-tuple u is chosen according to μ_δ^k.

Let $(\Omega_n, \mathcal{F}_n, \Pr)$ be a probability space and $n \in \mathbb{N}$. Furthermore, let $E_n \in \mathcal{F}_n$ denote an event whose probability depends on n. We say that E_n happens *almost surely* or with *high probability* if there exists $n_0 \in \mathbb{N}$ such that for all $n \in \mathbb{N} : n > n_0$ there exists $p(n)$ such that $\Pr[E_n] \geq 1 - 2^{-n} p(n)$.

Interactive protocols. We are often interested in situations where two prob-

abilistic circuits interact with each other according to some protocol by means of messages representable by bitstrings. Let $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ be families of circuits such that $A_n : \{0,1\}^* \to \{0,1\}^*$ and $B_n : \{0,1\}^* \to \{0,1\}^*$ where a tuple $(a_1, \ldots, a_l)$ taken as input by A_n is encoded as a bitstring in $\{0,1\}^*$, we use analogously technique for B_n. An *interactive protocol* is defined by $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ where for random bitstrings $\rho_A \in \{0,1\}^n$, $\rho_B \in \{0,1\}^n$ in the first round a message $m_0 := A_n(\rho_A)$ is sent and in the second round a message $m_1 := B_n(\rho_B, m_0)$. In general in the $(2k-1)$-th round we have $m_{2k-2} := A_n(\rho_A, m_1, \cdots, m_{2k-3})$ and in the $2k$-th round $m_{2k-1} := B_n(\rho_B, m_1, \cdots, m_{2k-2})$. The number of rounds of interaction is naturally bounded by the size of circuits A_n and B_n. The protocol execution between two probabilistic circuits A and B is denoted by $\langle A, B\rangle$. The output of A in the protocol execution is denoted by $\langle A, B\rangle_A$ and of B by $\langle A, B\rangle_B$. The sequence of all messages sent by A and B in the protocol execution is called a *communication transcript* and is denoted by $\langle A, B\rangle_{trans}$. The time complexity of the protocol depends on the size of A_n and B_n. For example, we say that a protocol runs in the polynomial time if the size of A_n and B_n are bounded by some $p(n)$.

Oracle algorithms. We use the notion of *oracle circuits* following the standard definition in the literature [Gol04]. If a circuit A has oracle access to a circuit B, we write A^B. If additionally B has oracle access to a circuit C, we write A^{B^C}. However, to shorten the notation, we often write A^B instead and make sure that it is clear from the context which oracle is accessed by B.

In many situations when studying the time complexity of algorithms with oracle access we count each oracle call as a single step. We emphasize this by writing that an algorithm has a certain time complexity *with oracle calls*. On the other hand, in some settings we are interested in giving a more rigorous bound on the running time of an algorithm. In these situations we compute the running time explicitly with regard to the time needed for accessing the oracle.

Definition 2.2 (Polynomial time samplable distribution.) We say that a distribution is polynomial time samplable if it can be approximated by an algorithm running in time $poly(\log|\mathcal{D}|, \log|\mathcal{R}|)$ up to an exponential factor.

Definition 2.3 (Pairwise independent family of efficient hash functions.) Let $\mathcal{D}$ and $\mathcal{R}$ be finite sets and $\mathcal{H}$ be a family of functions mapping values from $\mathcal{D}$ to values in $\mathcal{R}$. We say that $\mathcal{H}$ is a family of pairwise independent efficient hash functions if $\mathcal{H}$ has the following properties.

Pairwise independent. For all $x, y \in \mathcal{D}, x \neq y$ and for all $\alpha, \beta \in \mathcal{R}$, it holds that

$$\Pr_{hash \leftarrow \mathcal{H}}[hash(x) = \alpha \mid hash(y) = \beta] = \frac{1}{|\mathcal{R}|}.$$

Polynomial time samplable. For every $hash \in \mathcal{H}$ the function $hash$ is sampleable in time $poly(\log|\mathcal{D}|, \log|\mathcal{R}|)$.

Efficiently computable. For every $hash \in \mathcal{H}$ there exists an algorithm running in time $poly(\log|\mathcal{D}|, \log|\mathcal{R}|)$ which on input $x \in \mathcal{D}$ outputs $y \in \mathcal{R}$ such that $y = hash(x)$.

We note that the pairwise independence property is equivalent to

$$\Pr_{hash \leftarrow \mathcal{H}}[hash(x) = \alpha \wedge hash(y) = \beta] = \frac{1}{|\mathcal{R}|^2}.$$

It is well known [CW77] that there exists families of functions meeting the criteria stated in Definition 2.3.

Chapter 3

Dynamic Interactive Weakly Verifiable Puzzles

This chapter gives an overview of weakly verifiable cryptographic primitives. First, in Section 3.1 we give the definition of a *dynamic interactive weakly verifiable puzzle*. Then, in Section 3.2 we define the direct product of puzzles and state the hardness amplification theorem for dynamic interactive weakly verifiable puzzles which is the primary focus of this thesis.

3.1 The Definition

We define a *dynamic interactive weakly verifiable puzzle* as follows.

Definition 3.1 (Dynamic Interactive Weakly Verifiable Puzzle.) A *dynamic interactive weakly verifiable puzzle (DIWVP)* is defined by a family of probabilistic circuits $\{P_n\}_{n\in\mathbb{N}}$. A circuit belonging to $\{P_n\}_{n\in\mathbb{N}}$ is called the *problem poser*. A *solver* $C := (C_1, C_2)$ is a probabilistic two-phase circuit. We write $P_n(\pi)$ to denote the execution of P_n with the randomness fixed to $\pi \in \{0,1\}^n$ and $C(\rho) := (C_1, C_2)(\rho)$ to denote the execution of both C_1 and C_2 with the randomness fixed to $\rho \in \{0,1\}^*$.

In the first phase, the problem poser $P_n(\pi)$ and the solver $C_1(\rho)$ interact. As the result of the interaction $P_n(\pi)$ outputs a *verification circuit* Γ_V and a *hint circuit* Γ_H. The circuit $C_1(\rho)$ produces no output. The circuit Γ_V takes as input $q \in \mathcal{Q}$ (for some set $\mathcal{Q}$ of indices), an answer $y \in \{0,1\}^*$ and outputs a bit. We say that an answer (q, y) is a *correct solution* if and only if $\Gamma_V(q, y) = 1$. The circuit Γ_H on input $q \in \mathcal{Q}$ outputs a hint such that $\Gamma_V(q, \Gamma_H(q)) = 1$.

In the second phase, C_2 takes as input $x := \langle P_n(\pi), C_1(\rho)\rangle_{trans}$ and has oracle access to Γ_V and Γ_H. The execution of C_2 with the input x and the randomness fixed to ρ is denoted by $C_2(x, \rho)$. The queries of C_2 to Γ_V and Γ_H are called *verification queries* and *hint queries*, respectively. We say that the

circuit C_2 *succeeds* if and only if it makes a verification query (q, y) such that $\Gamma_V(q, y) = 1$, and it has not previously asked for a hint query on this q.

The above definition is very general and does not pose any constrains on the size of the circuits or the time complexity of the interactive protocol.

There is no loss of generality in assuming that the problem poser and the solver are defined by probabilistic circuits. Definition 3.1 embraces also a case where the problem poser and the solver are probabilistic polynomial time algorithms. We use the well known fact [Hol13b] that a probabilistic polynomial time algorithm can be transformed into an equivalent family of probabilistic Boolean circuits of the polynomial size[1].

We use the term *weakly verifiable* to emphasize that there is no easy way for the solver to check the correctness of a solution except for asking a verification query.

We call a weakly verifiable puzzle *dynamic* if the number of hint queries is greater than zero. Furthermore, we say that a weakly verifiable puzzle is *interactive* if in the first phase the number of messages exchanged between the problem poser and the solver is greater than one. Finally, we say that a weakly verifiable puzzle is *non-dynamic* if the number of hint queries is zero and *non-interactive* if the number of the messages sent in the first phase is at most one.

Definition 3.1 generalizes and combines previous approaches that study *weakly verifiable puzzles* [CHS04], *dynamic weakly verifiable puzzles* [DIJK09], and *interactive weakly verifiable puzzles* [HS10].

3.2 The Hardness Amplification Theorem

In this section we give the definition of the k-wise direct product of dynamic interactive weakly verifiable puzzles, then we state the hardness amplification theorem for dynamic interactive weakly verifiable puzzles.

Definition 3.2 (k-wise direct-product of DIWVPs.) Let $g : \{0,1\}^k \to \{0,1\}$ be a monotone function and $P_n^{(1)}$ a problem poser as in Definition 3.1. The k-wise direct product of $P_n^{(1)}$ is a DIWVP defined by a circuit $P_{kn}^{(g)}$. We write $P_{kn}^{(g)}(\pi^{(k)})$ to denote the execution of $P_{kn}^{(g)}$ with the randomness fixed to $\pi^{(k)} := (\pi_1, \ldots, \pi_k)$ where $\pi_i \in \{0,1\}^n$ for each $1 \leq i \leq n$. Let $(C_1, C_2)(\rho)$ be a probabilistic two-phase circuit called a *solver*. In the first phase, the problem poser $P_{kn}^{(g)}(\pi^{(k)})$ sequentially interacts in k rounds with the circuit $C_1(\rho)$. In the i-th round $C_1(\rho)$ interacts with $P_n^{(1)}(\pi_i)$, and as the result $P_n^{(1)}(\pi_i)$ generates

[1] Theorem 6.10 from [Hol13b] is stated for the probabilistic oracle programs with a single bit of output, but it can be adopted to the case where an output is longer than a single bit.

circuits Γ_V^i, Γ_H^i. Finally, after k rounds $P_{kn}^{(g)}(\pi^{(k)})$ outputs a verification circuit

$$\Gamma_V^{(g)}(q, y_1, \ldots, y_k) := g(\Gamma_V^1(q, y_1), \ldots, \Gamma_V^k(q, y_k))$$

and a hint circuit

$$\Gamma_H^{(k)}(q) := (\Gamma_H^1(q), \ldots, \Gamma_H^k(q)).$$

If it is clear from the context, we omit the subscript n and write $P(\pi)$ instead of $P_n(\pi)$ where $\pi \in \{0,1\}^n$.

A verification query (q, y) of a solver C for which a hint query on this q has been asked before cannot be a verification query for which C succeeds. Therefore, without loss of generality throughout this chapter, we make the assumption that C does not ask verification queries on q for which a hint query has been asked before. Moreover, we assume that once C asked a verification query that succeeds, it does not ask any further hint or verification queries.

Experiment $Success^{P,C}(\pi, \rho)$

Oracles: A problem poser P, a solver $C = (C_1, C_2)$ for P.
Input: Bitstrings $\pi \in \{0,1\}^n$, $\rho \in \{0,1\}^*$.
Output: A bit $b \in \{0,1\}$.

run $\langle P(\pi), C_1(\rho)\rangle$
  $(\Gamma_V, \Gamma_H) := \langle P(\pi), C_1(\rho)\rangle_P$
  $x := \langle P(\pi), C_1(\rho)\rangle_{trans}$

run $C_2^{\Gamma_V, \Gamma_H}(x, \rho)$
  if $C_2^{\Gamma_V, \Gamma_H}(x, \rho)$ asks a verification query (q, y) s.t. $\Gamma_V(q, y) = 1$ **then**
    return 1
return 0

We define the *success probability* of C in solving a puzzle defined by P as

$$\Pr_{\pi,\rho}[Success^{P,C}(\pi, \rho) = 1]. \tag{3.1}$$

Furthermore, for fixed P we say that C *succeeds* for π, ρ if $Success^{P,C}(\pi, \rho) = 1$.

Theorem 3.3 ***(Hardness amplification for dynamic weakly verifiable puzzles.)*** *Let $P_n^{(1)}$ be a fixed problem poser as in Definition 3.1 and $P_{kn}^{(g)}$ a problem poser for the k-wise direct product of $P_n^{(1)}$. Additionally, let C be a solver for $P_{kn}^{(g)}$ asking at most h hint queries and v verification queries.*

There exists a probabilistic algorithm $\widetilde{Gen}$ with oracle access to C, a monotone function $g : \{0,1\}^k \to \{0,1\}$, and a problem poser $P_n^{(1)}$. The algorithm $\widetilde{Gen}$ takes as input parameters ε, δ, n, k, h, v, and outputs a solver circuit $\widetilde{D}$ for $P_n^{(1)}$ such that:
If C is such that

$$\Pr_{\substack{\pi^{(k)} \in \{0,1\}^{kn} \\ \rho \in \{0,1\}^*}} \left[Success^{P_{kn}^{(g)}, C}(\pi^{(k)}, \rho) = 1 \right] \geq 16(h+v) \Big(\Pr_{u \leftarrow \mu_\delta^k} [g(u) = 1] + \varepsilon \Big),$$

then $\widetilde{D}$ is a two phase probabilistic circuit that almost surely over the randomness of $\widetilde{Gen}$ satisfies

$$\Pr_{\substack{\pi \in \{0,1\}^{n} \\ \rho \in \{0,1\}^*}} \left[Success^{P_n^{(1)}, \widetilde{D}}(\pi, \rho) = 1 \right] \geq \delta + \frac{\varepsilon}{6k}.$$

Additionally, $\widetilde{D}$ requires oracle access to g, $P_n^{(1)}$, C, hint and verification circuits generated by the problem poser $P_n^{(1)}$ after the first phase and asks at most $\frac{6k}{\varepsilon} \log\left(\frac{6k}{\varepsilon}\right) h$ hint queries and one verification query. Finally, $Time(\widetilde{Gen}) = poly(k, \frac{1}{\varepsilon}, n, v, h)$ with oracle calls.

We emphasize that the number of hint queries asked by D is greater than the number of hint queries asked by C whereas the number of verification queries is limited to at most one. In many applications, making such an assumption about the number of hint and verification queries is reasonable. In particular, we cannot assume that a solver for a single puzzle may ask more verification queries than a solver for the k-wise direct product.

The monotone restriction on g is essential. For $g(b) := 1 - b$ a circuit that deliberately gives incorrect answers satisfies g with probability 1 whereas a circuit that solves a puzzle successfully with probability $\gamma > 0$ succeeds only with probability $1 - \gamma$. Clearly, a desirable property for a solver for the k-wise direct product of interactive weakly verifiable puzzles is that a circuit that solves puzzles on all coordinates with the higher probability does not solve the k–wise direct product with probability lower than a circuit that success probability on these coordinates is lower.

Chapter 4

Examples of Weakly Verifiable Cryptographic Primitives

In the previous chapter we defined dynamic interactive weakly verifiable puzzles(DIWVPs), the k-wise direct product of DIWVP, and stated the hardness amplification theorem for DIWVP. In this chapter we give an overview of well known cryptographic primitives that can be seen as different types of weakly verifiable puzzles. First, in Section 4.1 we describe message authentication codes which game based security definition can be seen being a dynamic weakly verifiable puzzle. In Section 4.2 we describe public key signature schemes and give the chosen-message attack security definition of public key signature schemes which is an example of dynamic weakly verifiable puzzle. Then, in Section 4.3 we focus on bit commitment protocols that binding property can be seen as interactive weakly verifiable puzzle. Finally, in Section 4.4 we describe a problem of solving CAPTCHAs that can be considered to be a weakly verifiable puzzle.

4.1 Message Authentication Codes

In this section we describe message authentication codes (MAC) and conclude that the game based definition of MAC security can be seen being a dynamic non-interactive weakly verifiable puzzle.

Let us consider the setting in which two parties, a *sender* and a *receiver*, communicate over an *insecure channel* where messages of the sender may be intercepted, modified, and deleted by a third party called an *adversary*. The receiver needs a way to ensure that the received messages were indeed sent by the sender and were not modified by the adversary. The solution is to use *message authentication codes*.

Loosely speaking, message authentication codes can be explained as follows. Let the sender, receiver, and adversary be polynomial time algorithms and

messages be represented as bitstrings. Furthermore, we assume that the sender and the receiver share a secret key to which the adversary has no access. The sender appends to every message a tag which is computed as a function of the key and the message. The receiver, using the key, has a way to efficiently check whether an appended tag is valid for a received message. The receiver accepts a message if the tag is valid, otherwise it rejects. We require that it is hard for the adversary to find a tag and a message that was not sent before and with the non-negligible probability is accepted by the receiver.

Below we give the formal definition of a *message authentication code* based on [Mau13, Gol04].

Definition 4.1 (Message Authentication Code) Let $\mathcal{M}$ be a set of messages, $\mathcal{K}$ a set of keys and $\mathcal{T}$ a set of tags. A *Message authentication code* (MAC) is defined by an efficiently computable function $f : \mathcal{M} \times \mathcal{K} \to \mathcal{T}$. We say that MAC is *secure* if f satisfies the following condition:

Let k be chosen uniformly at random from $\mathcal{K}$ and H be a polynomial size circuit with hard-coded k that takes as input a message $m \in \mathcal{M}$ and outputs a tag $t \in \mathcal{T}$ such that $f(m, k) = t$. We say that MAC is secure if there is no probabilistic polynomial time algorithm with oracle access to H that with non-negligible probability outputs a message $m \in \mathcal{M}$ as well as a corresponding tag $t \in \mathcal{T}$ such that $f(m, k) = t$, and H has not been queried on m.

We describe how a task of breaking the security of MAC can be seen as solving a dynamic weakly verifiable puzzle where at most one verification query is asked. For fixed f and $n \in \mathbb{N}$ the sender corresponds to the problem poser and the adversary to the solver. Furthermore, the key k_π depends on the randomness $\pi \in \{0, 1\}^n$ used by the problem poser. The set $\mathcal{Q}$ is the set of messages $\mathcal{M}$.

In the first phase, which is non-interactive, neither the problem poser nor the solver send any messages. The problem poser outputs a hint circuit Γ_H and a verification circuit Γ_V where both circuits have hard-coded π. The circuit Γ_H corresponds to the circuit H from Definition 4.1 and takes as input a message m and outputs a tag t such that $f(m, k_\pi) = t$. The circuit Γ_V takes as input $m \in \mathcal{M}$ and $t \in \mathcal{T}$ and outputs 1 if and only if $f(m, k_\pi) = t$. In the second phase, the solver takes no input (x is empty string), is given oracle access to Γ_H and Γ_V and can ask at most one verification query.

Thus, the task of the adversary to find a valid tag $t \in \mathcal{T}$ for some message $m \in \mathcal{M}$ such that the hint query for m has not been asked before corresponds to making a successful verification query by a problem solver to Γ_V.

4.2 Public Key Signature Schemes

In this section we describe *public key signature schemes* which the game base security definition for the chosen message attack can be seen as a dynamic,

non-interactive weakly verifiable puzzle.

First, we give a definition of *public key encryption schemes*, and what it means for such schemes to be secure. The definitions are based on [Gol04].

Definition 4.2 (Public key signature scheme) Let $\mathcal{Q}$ be the set of messages. A *public key signature scheme* is defined by a triple of probabilistic polynomial time algorithms: G – the key generation algorithm, V – the verification algorithm, S – the signing algorithm, such that the following conditions are satisfied:

- $G(1^n, \rho)$ outputs a pair of bitstrings $k_{priv} \in \{0,1\}^*$ and $k_{pub} \in \{0,1\}^*$ where $n \in \mathbb{N}$ is a security parameter and ρ the randomness used by G. We call k_{priv} a *private key* and k_{pub} a *public key.*
- The signing algorithm S takes as input k_{priv}, $q \in \mathcal{Q}$ and outputs a signature $s \in \mathcal{S}$.
- The verification algorithm V takes as input k_{pub}, $q \in \mathcal{Q}$, and $s \in \mathcal{S}$ and outputs $b \in \{0,1\}$.
- For every k_{priv}, k_{pub} output by G and every $q \in \mathcal{Q}$ it holds

  $$\Pr[V(k_{pub}, q, S(k_{priv}, q))] = 1$$

  where the probability is over the randomness S.

We say that $s \in \mathcal{S}$ is a *valid signature* for $q \in \mathcal{Q}$ if and only if $V(k_{pub}, q, s) = 1$.

It seems that a game of breaking security of the public key signature scheme is not weakly verifiable as the adversary can use V to efficiency check the correctness of the solution. However, we will show that this situation sill can be modeled as a dynamic interactive weakly verifiable puzzle where the solver has access to the verification circuit and can ask a polynomial number of verification queries.

Definition 4.3 (Security of a public key signature scheme with respect to a chosen message attack.) Let H be a polynomial size circuit that has hard-coded k_{priv} and takes as input q and outputs $S(k_{priv}, q)$. An *adversary* A is a probabilistic polynomial time algorithm that takes as input k_{pub} and has oracle access to H. We say that A *succeeds* if it finds a valid signature $s \in \mathcal{S}$ for a message $q \in \mathcal{Q}$ and the oracle H has not been queried for a signature of q. The public key encryption scheme is *secure* if there is no polynomial time adversary that succeeds with non-negligible probability.

We will describe how a task of breaking the security of the public key signature scheme can be seen as a dynamic non-interactive weakly verifiable puzzle.

The problem poser correspond to G and the solver to the adversary. The set Q is a set of messages. In the first phase, the problem poser obtains k_{pub}, k_{priv}

and sends k_{pub} to the solver. Then, the problem poser generates a hint circuit Γ_H and a verification circuit Γ_V. The hint circuit Γ_H corresponds to the circuit H from Definition 4.3 and takes as input $q \in \mathcal{Q}$ and outputs a valid signature for q. The verification circuit Γ_V takes as input $s \in \mathcal{S}$ and $q \in \mathcal{Q}$ and outputs 1 if and only if $s \in \mathcal{S}$ is a valid signature for $q \in \mathcal{Q}$. In the second phase, the solver takes as input a transcript of the messages from the first round which consists solely of k_{pub}. Additionally, the solver is given oracle access to Γ_V and Γ_H and can make the polynomial number of queries to Γ_H and Γ_V. An adversary calls to V can be simulated by making a verification query to Γ_V. It is clear that if the solver asks a successful verification query (q, s), then there exists an adversary that also finds a valid signature for q.

Thus, a task of finding a valid signature by the adversary in a public key signature scheme can be seen as a weakly verifiable puzzle that is dynamic but non-interactive as in the first phase only a single message is sent.

4.3 Bit Commitment Protocols

In this section we describe *bit commitment protocols*. A task of breaking the binding property of a bit commitment protocol can be seen as an interactive, non-dynamic weakly verifiable puzzle where $|\mathcal{Q}| = 1$.

Loosely speaking, we consider the following *bit commitment protocol* that involves two parties, a *sender* and a *receiver*. We suppose that the sender and the receiver are polynomial time probabilistic algorithms. The protocol consists of a *commit phase* and a *reveal phase*. In the commit phase the sender and the receiver interact, as the result the sender commits to a value $b \in \{0, 1\}$. We require that after the commit phase it is hard for the receiver to correctly guess b. In the reveal phase the sender opens the commitment by sending to the receiver a pair (b', y) where $y \in \{0, 1\}^*$ should convince the receiver that the sender committed in the commit phase to the value $b' \in \{0, 1\}$. A desirable property of a bit commitment protocol is that in the reveal phase it is hard for the sender to find two bitstrings y_0 and y_1 such that the receiver recognizes both $(0, y_0)$ and $(1, y_1)$ to be valid decommitments.

We base the following formal definition of a *bit commitment protocol* on [Hol13a].

Definition 4.4 (Bit Commitment Protocol) A *bit commitment protocol* is defined by families of circuits $\{S_n\}_{n\in\mathbb{N}}$ and $\{R_n\}_{n\in\mathbb{N}}$ where $S_n = (S_1, S_2)$ is a two-phase probabilistic circuit, R_n is a probabilistic circuit, and $n \in \mathbb{N}$ is a security parameter. We call S_n the *sender* and R_n the *receiver*. The circuit S_1 takes as input a pair (b, ρ_S) where $b \in \{0, 1\}$ is interpreted as a bit to which S commits, and $\rho_S \in \{0, 1\}^*$ is the randomness used by S. The receiver R_n uses the randomness $\rho_R \in \{0, 1\}^*$. The protocol consists of two phases. In

the *commit phase*, S_1 and R_n engage in the protocol execution. As the result, S_n commits to b and R_n generates a circuit V. The circuit V takes as input $b' \in \{0,1\}$ and a bitstring $y \in \{0,1\}^*$ and outputs a bit. In the reveal phase, the circuit S_2 takes as input a communication transcript from the commitment phase $\langle S_1, R_n \rangle_{trans}$, the randomness ρ_s and returns (b', y). We require a bit commitment protocol to have the following properties:

Correctness. For a fixed $b \in \{0,1\}$, we have

$$\Pr_{\substack{\rho_S \in \{0,1\}^*, \rho_R \in \{0,1\}^* \\ V := \langle S_1(b,\rho_S), R_n(\rho_R) \rangle_{R_n} \\ (b',y) := S_2(\langle S_1(b,\rho_s), R_n(\rho_R) \rangle_{trans}, \rho_S)}} \Big[V(b', y) = 1 \Big] \geq 1 - \varepsilon(n),$$

where $\varepsilon(n)$ is a negligible function of n.

Hiding. For every $b \in \{0,1\}$ the probability over the random coins of S and R_n that any polynomial size circuit can guess b correctly after the commit phase is at most $\frac{1}{2} + \varepsilon(n)$ where $\varepsilon(n)$ is a negligible function of n.

Binding. For every probabilistic polynomial size two-phase circuit $S^*(\rho_S) := (S_1^*, S_2^*)(\rho_S)$ we have

$$\Pr_{\substack{\rho_S \in \{0,1\}^*, \rho_R \in \{0,1\}^* \\ V := \langle S_1^*(b,\rho_S), R_n(\rho_R) \rangle_{R_n} \\ ((0,y_0),(1,y_1)) := S_2^*(\rho_S)}} [V(0, y_0) = 1 \wedge V(1, y_1) = 1] \leq \varepsilon(k),$$

where $\varepsilon(n)$ is a negligible function in n.

Breaking the binding property of a bit commitment protocol can be seen as a problem of solving an interactive weakly verifiable puzzle. The receiver corresponds the problem poser and the solver to the sender that tries to break the binding property of the bit commitment protocol. The number of hint queries is zero, and the number of the verification queries is at most one. Thus, we conclude that the puzzle is non-dynamic.

In the first phase, the solver uses its randomness to choose a bit to which it commits. Then, it engage in interaction with the problem poser.

When the interaction in the commit phase is completed the problem poser generates circuits Γ_V, Γ_H. The circuit Γ_H takes as input $y_0 \in \{0,1\}^*$, $y_1 \in \{0,1\}^*$ and outputs 1 if and only if $V(0, y_0) = 1 \wedge V(1, y_1) = 1$. The hint circuit Γ_H outputs $\perp$ for any query. Thus, without loss of generality, we assume that the solver does not ask any verification queries. Furthermore, for the bit commitment protocols $|\mathcal{Q}| = 1$, thus we do not write explicitly on which $q \in \mathcal{Q}$ the solver asks a verification query.

In the second phase, the solver is given oracle access to Γ_V and is allowed to ask at most one verification query. We emphasize that the solver has only

oracle access to Γ_V. Therefore, it has no efficient way to check whether the solution is successful except asking a verification query.

Finally, we conclude that for the problem poser and the solver defined as above a task of making a successful verification query corresponds to breaking the binding property of the bit commitment protocol.

4.4 Automated Turing Tests

In this section we describe CAPTCHAs which can be seen as non-dynamic and non-interactive weakly verifiable puzzles.

The goal of *Automated Turing tests* is to distinguish humans from computer programs. An example of such a test is *CAPTCHA*, which is formally defined in [VABHL03]. Loosely speaking, CAPTCHA is a test for which it is hard to write a computer program whose success probability is comparable to or higher than the one achieved by humans. An example is an image depicting a distorted text where the goal is to guess the text used to generate the image.

A task of solving a CAPTCHA based on guessing the distorted text can be seen as a non-dynamic, non-interactive weakly verifiable puzzle. In the first round the problem poser sends the solver an image containing the distorted text. Then, the problem poser generates circuits Γ_V and Γ_H. The circuit Γ_V takes as input a bitstring y and outputs 1 if and only if y correctly encodes the text depicted on the distorted image. A hint circuit Γ_H outputs $\perp$ in response to every query. Therefore, without loss of generality, we assume that the solver makes no hint queries.

In the second phase, the solver takes as input the image with the distorted text, has oracle access to Γ_V and can ask at most a single verification query. In general, for CAPTCHAs checking the correct of a solution is comparably hard to finding a correct solution. Thus we conclude that a task of correctly guessing a text depicted on the image can be seen as a weakly verifiable non-interactive and non-dynamic puzzle.

It is not known how good an algorithm for solving CAPTCHAs can be. Thus, it is likely that a gap between the human performance and a performance of the best computer programs may be small. We are interested in finding a way to amplify this gap. In the paper [HS10] it was shown that this is possible if the parallel repetition of non-dynamic interactive weakly verifiable puzzles is used where a solver is given n independent weakly verifiable puzzles. The verifier for the parallel repetition accepts if the solver succeeds on at least some fraction of these puzzles.

In Chapter 3 we define the k-wise direct product of puzzles and state the similar theorem about hardness amplification for dynamic weakly verifiable puzzles as in [HS10]. If this hardness amplification theorem was true, then

it would be possible to construct a weakly verifiable puzzle for which is very hard for even the best computer programs but is still very easy for humans (under the assumption that humans have a noticeable advantage over computer programs).

Chapter 5

Previous Results

In the last chapter we gave an overview of different types of cryptographic primitives that motivated studies of weakly verifiable puzzles. The focus of this chapter is on giving the outline of the previous research concerning different types of weakly verifiable puzzles. We give a short overview of techniques used in the series of papers [CHS04, DIJK09, HS10] and provide some intuition and insight into the problem of the hardness amplification for weakly verifiable puzzles. First, we describe weakly verifiable puzzles studied in [CHS04] that are neither interactive nor dynamic. Then, in Section 5.2 we bring our focus on the dynamic non-interactive puzzles studied in [DIJK09]. Finally, in Section 5.3 we give an overview of the results of Holenstein and Schoenebeck [HS10] where non-dynamic but interactive weakly verifiable puzzles have been studied.

5.1 Weakly Verifiable Puzzles

The notion of *weakly verifiable puzzles* has been coined by Canetti, Halevi, and Steiner in the paper *Hardness amplification of weakly verifiable puzzles* [CHS04]. In comparison to Definition 3.1 the puzzles considered in [CHS04] are non-dynamic and non-interactive. Moreover, the number of verification queries is limited to one. This constitutes a special case of Definition 3.1. In this section we provide the definition of weakly verifiable puzzles (WVP) that closely follows the one contained in [CHS04] and state the theorem of hardness amplification of weakly verifiable puzzles in a similar vein as in [CHS04]. Finally, we give an intuition behind the proof of this theorem. It is noteworthy that the main proof of this Thesis, contained in Chapter 6, uses many ideas of the work of Canetti, Halevi, and Steiner [CHS04].

5.1.1 The definition

We give the definition of WVP that uses the notation and terminology used in this Thesis.

Definition 5.1 (Weakly Verifiable Puzzles, [CHS04]) A *weakly verifiable puzzle* is defined by a pair of polynomial time algorithms: a probabilistic puzzle–generation algorithm G and a deterministic verification algorithm V. We write $G(1^k;\rho)$ to denote that G takes as input a bitstring 1^k, where k is a security parameter and the randomness $\rho \in \{0,1\}^*$. The algorithm G outputs a bitstring $p \in \{0,1\}^*$ and a check information $c \in \{0,1\}^*$. The verifier V is a deterministic algorithm that takes as input p, c, an answer $a \in \{0,1\}^*$ and outputs $b \in \{0,1\}$.

A solver S for G is a polynomial time probabilistic algorithm that takes as input p and outputs a. We denote the randomness used by S as $\pi \in \{0,1\}^*$ and define the success probability of S in solving a puzzle defined by P as

$$\Pr_{\substack{\rho\in\{0,1\}^*,\pi\in\{0,1\}^* \\ (p,c):=G(1^k;\rho) \\ a:=S(p,\pi)}}\Big[V(p,c,a)=1\Big].$$

We write $P := (G, V)$ to denote a weakly verifiable puzzle P defined by algorithms G and V.

Let us argue that Definition 5.1 is a special case of Definition 3.1. First, we note that for the efficient algorithm G that takes as input 1^k the length of the randomness used by G is bounded by some polynomial $p(k)$. For a fixed k, without loss of generality, we can represent $G(1^k;\rho)$ as a probabilistic circuit of polynomial size that takes as input only the randomness ρ. The security parameters from Definition 5.1 and Definition 3.1 are not equivalent, as in the later definition the security parameter limits the length of the random bistring. Moreover, in Definition 5.1 a verification algorithm takes as input p, c, a. Again, without loss of generality, we can assume that bitstrings p and c are hard-coded in the circuit Γ_V from Definition 3.1. Hence, for fixed p and c the algorithm V corresponds to Γ_V. The puzzles considered in Definition 5.1 are non–dynamic. Thus, there is no element corresponding to the hint circuit Γ_H. Finally, we note that the puzzles described in Definition 5.1 are non–interactive.

5.1.2 The hardness amplification

We will now give the definition of the n-wise direct product of weakly verifiable puzzles along the lines of [CHS04].

Definition 5.2 (n-wise direct product of weakly verifiable puzzles [CHS04].) Let $n \in \mathbb{N}$ and a weakly verifiable puzzle $P = (G, V)$ be fixed. We define

the n-fold repetition of P as a weakly verifiable puzzle where the puzzle–generation algorithm $G^{(n)}$ takes as input 1^{kn}, as an auxiliary input a bit-string $\rho \in \{0,1\}^*$ and outputs tuples $p^{(n)} := (p_1, \ldots, p_n) \in \{0,1\}^*$ and $c^{(n)} := (c_1, \ldots, c_k) \in \{0,1\}^*$ where for each $1 \leq i \leq n$ a pair (p_i, c_i) is an independent instance of a weakly verifiable puzzle defined by G and V with the security parameter k. Finally, the verification algorithm $V^{(n)}$ takes as input $p^{(n)}$, $c^{(n)}$, an answer $a^{(n)}$, and outputs $b \in \{0,1\}$ such that $b = 1$ if and only if for all $1 \leq i \leq n$ we have $V(p_i, c_i, a_i) = 1$.

Let us give some notation and terminology. We write $P^{(n)} := (G^{(n)}, V^{(n)})$ to denote the n–fold repetition of P. For $P^{(n)} := (G^{(n)}, V^{(n)})$ when writing a *puzzle on the i-th coordinate* we refer to the i-th puzzle of the n–fold repetition of P (this puzzle corresponds to the one generated by $G_i^{(n)}$, $V_i^{(xn)}$).

The n-wise direct product of weakly verifiable puzzles is solved successfully if and only if all n puzzles are solved successfully. In contrast, in Definition 3.2 we are interested in a more general situation where a monotone function $g : \{0,1\}^n \to \{0,1\}$ is used to decide which coordinates of the n-wise direct product of puzzles have to be solved correctly. Clearly, we can assume that g is such that puzzles on all coordinates have to be solved successfully which matches the case considered in Definition 5.2.

The main theorem proved in [CHS04] states that it is possible to turn a good solver for $P^{(n)}$ to a good solver for P.

Theorem 5.3 (Hardness amplification for weakly verifiable puzzles [CHS04].) *Let $n, q \in \mathbb{N}$ and $\delta : \mathbb{N} \to (0,1)$ be efficiently computable functions. Moreover, let $P = (G, V)$ be a weakly verifiable puzzle. We denote the running time of the puzzle–generation algorithm G for P by T_G and of the verification algorithm V for P by T_V. If $S^{(n)}$ is a solver for the n–fold repetition of P that success probability is at least δ^n and the running time is T, then there exists a solver S for G with oracle access to $S^{(n)}$ that success probability is at least $\delta(1 - \frac{1}{q})$ and the running time is $O\Big(\frac{nq^3}{\delta^{2n-1}}(T + nT_G + nT_V)\Big)$.*

The following algorithm is used in [CHS04] to prove Theorem 5.3. It turns $S^{(n)}$ for $P^{(n)}$ with success probability at least δ^n into a solver for a single puzzle P with probability at least $\delta(1 - \frac{1}{q})$. The parameter q is introduced as it is not possible to achieve the perfect hardness amplification. We note that in the analysis of the running time of *CHS–solver* we explicitly take into account the time needed for the oracle calls to $S^{(n)}$, V, G.

Let us denote by $p \in \{0,1\}^*$ an output of G, which is taken as input by *CHS–solver*. To make notation shorter in the following code excerpts we do not write randomness used by G explicitly.

Algorithm: $\mathit{CHS\text{–}solver}^{S^{(n)},V,G}(p, n, k, q, \delta)$

Oracle: A solver $S^{(n)}$ for $P^{(n)}$, a verification algorithm V for P, a puzzle–generation algorithm G for P.
Input: A bistring $p \in \{0,1\}^*$, parameters n, k, q, δ.

$\mathit{prefix} := \emptyset$
for $i = 1$ **to** $n-1$ **do:**
  $p^* := \mathit{ExtendPrefix}^{S^{(n)},V,G}(\mathit{prefix}, i, n, k, q, \delta)$
  if $p^* = \perp$ **then return** $\mathit{OnlinePhase}^{S^{(n)},V,G}(\mathit{prefix}, p, i, n, k, q, \delta)$
  else $\mathit{prefix} := \mathit{prefix} \circ p^*$
$a^{(n)} := S^{(n)}(\mathit{prefix} \circ p)$
return a_n

Algorithm: $\mathit{OnlinePhase}^{S^{(n)},V,G}(\mathit{prefix}, p, i, n, k, q, \delta)$

Oracle: A solver algorithm $S^{(n)}$ for $P^{(n)}$, a puzzle–generation algorithm G for P, a verification algorithm V for P.
Input: A $(v-1)$–tuple of bitstrings *prefix*, a bitstring $p \in \{0,1\}^*$, parameters v, n, k, q, δ.

Repeat $\left\lceil \frac{6q \ln(6q)}{\delta^{n-i+1}} \right\rceil$ times
  $((p_{i+1}, \ldots, p_n), (c_{i+1}, \ldots, c_n)) := G^{(n-i-1)}(1^k)$
  $a^{(n)} := S^{(n)}(\mathit{prefix}, p, p_{i+1}, \ldots, p_n)$
  if $\forall_{i+1 \le j \le n} V(p_j, c_j, a_j) = 1$ **then return** a_i
return $\perp$

Algorithm: $\mathit{ExtendPrefix}^{S^{(n)},V,G}(\mathit{prefix}, i, n, k, q, \delta)$

Oracle: A solver algorithm $S^{(n)}$ for $P^{(n)}$, a puzzle–generation algorithm G for P, a verification algorithm V for P.
Input: A $(i-1)$–tuple of puzzles *prefix*, parameters i, n, k, q, δ.

Repeat $\left\lceil \frac{6q}{\delta^{n-v+1}} \ln(\frac{18qn}{\delta}) \right\rceil$ times
  $(p^*, c^*) := G(1^k)$
  $\bar{\nu}_i := \mathit{EstimateResSuccProb}^{G,V}(\mathit{prefix} \circ p^*, i, n, k, q, \delta)$
  if $\bar{\nu}_i \ge \delta^{n-i}$ **then return** p^*
return $\perp$

Algorithm: $EstimateResSuccProb^{S^{(n)},V,G}(prefix, i, n, k, q, \delta)$

Oracle: A solver algorithm for $P^{(n)}$, a verification algorithm V for P, a generation algorithm G for P
Input: A i–tuple of puzzles *prefix*, parameters i, n, k, q, δ.

$successes := 0$
Repeat $M := \left\lceil \frac{84q^2}{\delta^{n-i}} \ln\left(\frac{18qn \cdot N_i}{\delta}\right)\right\rceil$ times
  $((p_{i+1}, \ldots, p_n), (c_{i+1}, \ldots, c_n)) := G^{(n-i)}(1^k)$
  $a^{(n)} := A(prefix, p_{i+1}, \ldots, p_n)$
  if $\forall_{i+1 \leq j \leq n} : V(p_j, c_j, a_j) = 1$ **then** $successes := successes + 1$
return $successes/M$

A detailed proof of Theorem 5.3 is presented in [CHS04]. We limit ourselves to providing the intuition why the CHS–solver transforms a good solver for the n–wise direct product of P to a good solver for P.

Let us consider the n-fold repetition of P, and for simplicity a deterministic solver $S^{(n)}$ for $P^{(n)} := (G^{(n)}, V^{(n)})$. Furthermore, we write $p^{(n)}, c^{(n)}$ to denote the output of $G^{(n)}$. We define a matrix M as follows. The columns of M are labeled with all possible bitstrings p_1 whereas the rows are labeled with all possible tuples $(p_2, \ldots, p_n)$ output by algorithm $G^{(n)}$ executed with different randomness. A cell of M contains a binary n-tuple such that the i-th bit equals 1 if and only if $V_i(p_i, c_i, a_i) = 1$ where $a^{(n)} := S^{(n)}(p^{(n)})$ and $p^{(n)}$ is a tuple of bitstring inferred by a column and a row of the cell. We make the following observation.

Observation 5.4 *For a deterministic polynomial time algorithm $S^{(n)}$ that successfully solves the n–fold repetition of P with probability at least δ^n the matrix M defined as above has either a column with $\delta^{(n-1)}$ fractions of cells that are all one vectors, or a conditional probability that a cell is of the form 1^n given that the last $(n-1)$ bits of the cell are equal 1 is at least δ.*

We show, at least intuitively using Observation 5.4, how the algorithm *CHS–solver* can be used to solve a puzzle defined by P with substantial probability given oracle access to $S^{(n)}$ for $P^{(n)}$. The algorithm starts with the first position and tries to fix a bitstring p^* on this position such that the success probability of $S^{(n)}$ on the remaining $(n-1)$ position is at least $\delta^{(n-1)}$. If it is possible to find p^* such that this condition is satisfied, then we fix p^* on this position and repeat the whole procedure again in the consecutive iteration for the next position. If *CHS–solver* fails to find a bitstring p^*, then we assume that there is no column of M that contains a $\delta^{(n-1)}$ fraction of cells that are of the form 1^n. We use Observation 5.4 to conclude that the conditional probability of solving the first puzzle given that all puzzles on the remaining position are

solved successfully is at least δ. We place the input puzzle p on this position and note that all remaining puzzles are generated by *CHS–solver*. Thus, it is possible to efficiently verify whether these puzzles are successful solved by $S^{(n)}$.

Obviously, the algorithm *CHS–solver* can still fail. First, it may happen that it does not find a column with a high fraction of puzzles that are solved successfully, although such a column exists. Secondly, we cannot exclude a situation where no such column exists, but the algorithm fails to find a cell such that last $(n-1)$ bits are 1. Finally, it is also possible that an estimate returned by *EstimateResSuccProb* is incorrect.

It is possible to show that all these events happen with small probability. Therefore, at least intuitively we see that the algorithm *CHS–solver* solves a single WVP puzzle successfully with probability at least $\delta(1-\frac{1}{q})$ almost surely.

In Chapter 6 we study a more general class of puzzles that are not only weakly verifiable but also dynamic and interactive. Furthermore, we allow a more general situation where a solver successfully solves the n-fold repetition of puzzles[1] although it succeeds only on some coordinates of the n-fold repetition of P. It turns out that it is possible to use a similar technique of fixing puzzles on consecutive positions of the n-fold repetition of puzzles to prove the hardness amplification in this more general setting.

5.2 Dynamic Weakly Verifiable Puzzles

Some of the cryptographic constructions presented in Section **??** are not only weakly verifiable but also dynamic (MAC and SIG). This type of puzzles are defined and studied in [DIJK09]. We give a short overview of this work, state the definition of a *dynamic weakly verifiable puzzle* that closely follows the one included in [DIJK09]. Finally, we provide intuition for the proof of the hardness amplification of DWVP included in [DIJK09].

5.2.1 The definition

Definition 5.5 (Dynamic Weakly Verifiable Puzzle.) A dynamic weakly verifiable puzzle (DWVP) is defined by a distribution $\mathcal{D}$ on pairs (x, α) where $\alpha \in \{0,1\}^*$ is an advice used to generate and evaluate responses and $x \in \{0,1\}^*$ is a bitstring taken as input by the solver. Furthermore, we consider a set $\mathcal{Q}$ and a probabilistic polynomial time computable relation R such that $R(\alpha, q, r) = 1$ if and only if r is a correct answer to $q \in \mathcal{Q}$ on the set of puzzle determined by α. Finally, let $H(\alpha, q)$ be a probabilistic polynomial time computable hint relation.

[1] Actually, in Chapter 6 we define the k-wise repetition of puzzles following the terminology used in [DIJK09] which is equivalent to the n-fold repetition of puzzles.

A solver S takes as input x and can ask hint queries on $q \in \mathcal{Q}$ which are answered using $H(\alpha, q)$ and verification queries of the form (q, r) answered by means of $R(\alpha, q, r)$. We say that S succeeds if and only if it makes a verification query on (q, r) such that $R(\alpha, q, r) = 1$ and it has not previously asked for a hint query on this q. We write $P := (\mathcal{D}, R, H)$ to denote a DWVP with a distribution $\mathcal{D}$ of pairs (x, α), and R, H being a verification and hint relations respectively.

We show now how the above definition is generalized by Definition 3.1. First, instead of considering a distribution on pairs (x, α) in Definition 3.1 we use a probabilistic problem poser that outputs circuits Γ_H and Γ_V that corresponds to hint and verification relations respectively. Furthermore, the problem poser may interact in the first phase with the problem solver. In particular, the problem poser can send a bitstring x as a message in the first phase. Thus, Definition 3.1 captures a more general case where the distribution of puzzles is defined by both the problem poser and the problem solver.

We define the *n-wise direct product of DWVPs* which is conceptually similar to the n-fold repetition of WVPs.

Definition 5.6 (n-wise direct product of DWVPs) For a dynamic weakly verifiable puzzle $P := (\mathcal{D}, R, H)$ we define the n-wise direct product of P as a DWVP with a distribution $\mathcal{D}^{(n)}$ on tuples $(x_1, \alpha_1), \ldots, (x_n, \alpha_n)$. Furthermore, the hint relation is defined by $H^{(n)}(q, \alpha_1, \ldots, \alpha_n) := (H(\alpha_1, q), \ldots, H(\alpha_n, q))$ and the verification relation $R^{(n)}(\alpha_1, \ldots, \alpha_n, r_1, \ldots, r_n, q)$ evaluates to 1 if and only if for $1 \leq i \leq n$ at least $n - (1 - \gamma)\delta n$ is such that $R(\alpha_i, q, r_i) = 1$ where $0 \leq \gamma, \delta \leq 1$. We write $P^{(n)} := (D^{(n)}, H^{(n)}, R^{(n)})$ to denote the n-wise direct product of $P := (D, H, R)$.

In contrast to the n-fold repetition of puzzles defined in the previous section, here we require the solver to succeed only on a fraction of puzzles.

Dynamic weakly verifiable puzzles generalize a games of breaking security of message authentication codes and public signature schemes. In case of MAC the adversary takes x which is an empty string. For the public signature schemes x is the public key.

We write $(\mathcal{H}_{hint}, \mathcal{V}_{verif}) \leftarrow S(x; \delta)$ to denote the execution of a solver S with input $x \in \{0, 1\}^*$ and using randomness $\delta \in \{0, 1\}^*$. Furthermore, $\mathcal{H}_{hint}$ is the set of all hint queries asked by S and $\mathcal{V}_{verif}$ is a set of all pairs (q, a) of verification queries asked in the execution of S.

With no loss of generality we make the assumption that the solver does not make hint queries on the successful verification queries. We define the *success*

probability of a solver S for $P := (G, V)$ as

$$\Pr_{\substack{\delta\in\{0,1\}^* \\ (x,\alpha)\leftarrow\mathcal{D} \\ (\mathcal{H}_{hint},\mathcal{V}_{verif})\leftarrow S(x,\delta))}} \left[\exists(q,a) \in \mathcal{V}_{verif} : q \notin \mathcal{H}_{hint} \wedge V(q,a) := 1\right]$$

5.2.2 The hardness amplification theorem

Theorem 5.7 ***(Hardness amplification for dynamic weakly verifiable puzzles).*** *Let $S^{(n)}$ be a probabilistic algorithm for $P^{(n)}$ that succeeds with probability at least ε, where $\varepsilon \geq (800/\gamma\delta) \cdot (h+v) \cdot e^{-\gamma^2\delta n/40}$, and h and v denote the number of hint and verification queries asked by $S^{(n)}$ respectively. Then there exists a probabilistic algorithm S that succeeds in solving P with probability at least $1-\delta$ making $O(h(h+v)/\varepsilon) \cdot \log(1/\gamma\delta)$ hint queries and at most one verification query. Furthermore, the running time is polynomial in $h, v, \frac{1}{\varepsilon}, t, \omega, \log(1/\gamma\delta)$ where ω is the time needed to ask a single hint query.*

It is worth seeing why the approach presented in the previous section that works well for the n-fold repetition of WVP cannot be applied for the k-wise direct product of DWVP (moving aside for a moment the issue of solving only a fraction of puzzle successfully). For DWVP the algorithm *CHS–solver* breaks in the *OnlinePhase* where the solver $S^{(n)}$ can be called multiple times. It is possible that in one of these runs $S^{(n)}$ asks a hint query on q for which in one of the later runs a verification query (q, r) is asked for which algorithm would return an answer for the input puzzle (in other words the condition $\forall_{v+1\leq i\leq n} V(p_i, c_i, a_i) = 1$ is satisfied). However, the fact that a hint query on this q has been asked makes it impossible to ask a successful verification query on this q. Thus, we can not dismiss a situation where the success probability of $S^{(n)}$ decreases with the number of iterations.

The solution proposed in [DIJK09] is to partition the set $\mathcal{Q}$ into a set of *attacking queries* $\mathcal{Q}_{attack}$ and a set of *advice queries* $\mathcal{Q}_{adv}$. The idea is to allow a solver for the n-wise direct product to ask hint queries only on $q \in \mathcal{Q}_{adv}$, and to halt the execution whenever a hint query is asked on $q \in \mathcal{Q}_{attack}$.

It is possible, for a solver S that asks at most h hint queries and v verification queries, to find a function $\mathcal{Q} \to \{0, 1, \ldots 2(h+v)-1\}$ such that the success probability of S with respect to $\mathcal{Q}_{attack}$ and $\mathcal{Q}_{adv}$ is multiplied by $\frac{1}{8(h+v)}$. If h and v are not too big, then the success probability of S can be still substantial. More formally, for a function $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$ we define $\mathcal{Q}_{attack} := \{q \in \mathcal{Q} : hash(q) = 0\}$ and $\mathcal{Q}_{adv} := \{q \in \mathcal{Q} : hash(q) \neq 0\}$

In [DIJK09] the following lemma is proved.

Lemma 5.8 *Let S be a solver for DWVP which success probability is at least δ, the running time is at most t, and the number of hint and verification*

queries is at most h and v respectively. There exists a probabilistic algorithm that runs in time polynomial in $h, v, \frac{1}{\delta}, t$ that outputs a function hash $: \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$ that partitions $\mathcal{Q}$ to $\mathcal{Q}_{attack}$ and $\mathcal{Q}_{adv}$ such that with probability at least $\frac{\delta}{8(h+v)}$ the first successful verification query (q', a) asked by S is such that $q' \in \mathcal{Q}_{attack}$ and all previous hint and verification queries has been asked on $q \in \mathcal{Q}_{adv}$.

A function *hash* can be found by using a natural sampling technique. We follow exactly the same approach of the partitioning the domain in the main proof of this Thesis in Section 6.1.2.

Let $H_\alpha(q)$ denote a polynomial time probabilistic algorithm that takes as input q, has hard-coded α and outputs $H(\alpha, q)$. Similarly, we use $R_\alpha(q, r)$ to denote a polynomial time probabilistic algorithm that computes relations $R(\alpha, q, r)$ and has hard-coded bitstring α. The following algorithm is used in [DIJK09] in the proof of Lemma 5.7. It gains oracle access to R_α and H_α as well as a function *hash* from Lemma 5.8.

Algorithm: $DWVP\text{–}solver^{S^{(n)}, hash, H_\alpha^{(n)}, R_\alpha^{(n)}}(x)$

Oracle: A solver $S^{(n)}$ for $P^{(n)}$, a function $hash : Q \to \{0, 1, \ldots, 2(h+v)-1\}$.
Input: A bistring $x \in \{0,1\}^*$.

Repeat at most $O(\frac{h+v}{\varepsilon} \cdot \log(\frac{1}{\gamma\delta}))$ times

Let $i \xleftarrow{\$} \{1, \ldots, n\}$ be a position for x.
Generate $(x_1, \alpha_1), \ldots, (x_{i-1}, \alpha_{i-1}), (x_{i+1}, \alpha_{i+1}), \ldots, (x_n, \alpha_n)$
using $(n-1)$ calls to P each time with fresh randomness.
run $S^{(n)}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$
  if $S^{(n)}$ asks a hint query on q **then**
    if $hash(q) \neq 0$ **then** abort current run of $S^{(n)}$
    Ask a verification query $r := H(q)$
    Let $(r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n)$ be hints for query q for puzzle sets $(x_1, \ldots, x_{i-1}, x_{i+1}, x_n)$
    Answer the hint query of $S^{(n)}$ using $(r_1, \ldots, r_{i-1}, r, r_{i+1}, r_n)$
  if $S^{(n)}$ asks a verification query $(q, r_1, \ldots, r_n)$ **then**
    if $hash(q) = 0$ **then** answer the query with 0
    Let $m := |j : V(q, r_j) = 1, j \neq i|$
    if $m \geq n - n(1-\gamma)\delta$ **then**
      make a verification query (q, r_i) and halt.
    else with probability $\rho^{m - n(1-\gamma)\delta}$ ask a verification query (q, r_i) and halt.
    Halt the current run of $S^{(n)}$ and go to the next iteration.

return $\perp$

In the above algorithm we execute multiple times a solver $S^{(n)}$ for the k-wise direct product of DWVPs. In each iteration the position for $x \in \{0,1\}^*$ is chosen uniformly at random. The remaining $(n-1)$ puzzles are generated by the algorithm, thus it is possible to answer all hint and verification queries for these puzzles. We use a function *hash* to partition the query domain and assume that *hash* is such that the success probability of $S^{(n)}$ with respect to *hash* is at least $\frac{\delta}{8(h+v)}$. We check on which q the solver $S^{(n)}$ asks hint and verification queries. If a hint query is asked on q such that $hash(q) = 0$ then the execution of $S^{(n)}$ is aborted and we go to the next iteration. This way we make sure that the algorithm never asks a hint query that could prevent a verification query from succeeding.

If a verification query is asked on q such that $hash(q) \neq 0$ we answer such a verification query with 0.

Finally, in case when $S^{(n)}$ asks a verification query using index q such that $hash(q) = 0$, then we use a soft decision system to decide whether to ask a verification query. The idea is that if there are many puzzles among the ones generated by the algorithm that are solved successfully, then it is likely that also the input puzzle is solved successfully. We discount $\gamma\delta n$ to take into account that not all puzzles have to be solved successfully. The detail calculations provided in [DIJK09] show that this approach yields a demanded result. We do not give a more detail description of the proof of [DIJK09] as in Chapter 6, except the domain partitioning, we use a different technique.

In case of weakly verifiable primitives like CAPTCHAs we assume that most people have at least slightly higher probability of solving these kind of puzzles than the best computer programs. Still, it may happen that humans do not solve all puzzles. This is a motivation to introduce a threshold function such that on average solutions of humans are treated as solved successfully but the ones of computer programs on average are classified as not successful. This motivates a study of the situations where only a fraction of puzzles is solved successfully.

In Chapter 6 we consider a weakly verifiable puzzles that are not only dynamic but also interactive. We use a very similar technique to partition domain Q into advice and hint queries as presented in [DIJK09]. Instead of the requirement to succeed only on a fraction of puzzles we consider an arbitrary, monotone function $g : \{0,1\}^n \to \{0,1\}$ that determines on which coordinates the solver has to succeed in order to successfully solve the n-wise direct product of puzzles.

To show the hardness amplification for the n-wise direct product with the domain partitioned we use the approach similar to the one presented in Section 5.1. Namely, we try to find a good position for the input puzzle instead of choosing the position uniformly on random as in [DIJK09].

5.3 Interactive Weakly Verifiable Puzzles

The hardness amplification of interactive weakly verifiable puzzles has been studied by T.Holenstein and G.Schoenebeck in [HS10]. We will give now an overview of this work and compare it with our approach.

5.3.1 The definition

The following definition of an *interactive weakly verifiable puzzle* closely follow the one from [HS10].

Definition 5.9 An *interactive weakly verifiable puzzle* is defined by a protocol given by two probabilistic algorithms P and S. The algorithm P is called the problem poser and produces as output a verification circuit Γ. The algorithm S called the problem solver produces no output. Furthermore, the success probability of the algorithm S in solving an interactive weakly verifiable puzzle defined by (P, S) is:

$$\Pr_{\substack{\rho,\pi \\ \Gamma^{(g)}:=\langle P(\rho),S(\pi)\rangle_P}} \left[\Gamma^{(g)}(\langle P(\rho), S(\pi)\rangle_S) = 1\right].$$

We are intrested in the hardness amplification of interactive weakly verifiable puzzles. Thus, similarly as in the previous sections we define the k-wise direct product of puzzles.

Definition 5.10 (k-wise direct product of interactive weakly verifiable puzzles) Let $g : \{0,1\}^k \to \{0,1\}$ be a monotone function and (P, S) be a fixed interactive weakly verifiable puzzle. The k-wise direct product of (P, S) defined by $(P^{(g)}, S^{(g)})$ is an interactive weakly verifiable puzzles in which the sender and the receiver sequentially interact in k rounds where in each round (P, S) is used to generate an instance of interactive weakly verifiable puzzle. As the result circuits $\Gamma^{(1)}, \ldots, \Gamma^{(k)}$ for P are generated. Finally, $P^{(g)}$ outputs the circuit $\Gamma^{(g)}(y_1, \ldots, y_k) := g(\Gamma^{(1)}(y_1), \ldots, \Gamma^{(k)}(y_k))$.

Similarly as in Definition 3.1 the puzzles considered in [HS10] are interactive. Furthermore, a monotone binary function is used to determine whether the k-wise repetition has been successfully solved. Unlike, puzzles in Definition 3.1, the puzzles studied by T.Holenstein and G.Schoenebeck are non-dynamic. Thus, only a verification circuit Γ is generated and no hint circuit is ever used.

5.3.2 The hardness amplification theorem

The following hardness amplification theorem is proved in [HS10].

Theorem 5.11 *There exists an algorithm $Gen(C, g, \varepsilon, \delta, n)$ which takes as input a solver circuit C for the k-wise direct product of P, a monotone function*

$g : \{0,1\}^* \to \{0,1\}$, and parameters ε, δ, n. The algorithm Gen outputs a solver circuit D for P such that the following holds. If C is such that

$$\Pr\left[\Gamma^{(g)}(\langle P^{(g)}, C\rangle_C) = 1\right] \geq \Pr_{u \leftarrow \mu_\delta^{(k)}}\left[g(u) = 1\right] + \varepsilon,$$

then, D satisfies almost surely,

$$\Pr\left[\Gamma(\langle P, D\rangle_D) = 1\right] \geq \delta + \frac{\varepsilon}{6k}.$$

Additionally, Gen and D only require oracle access to g and C. Furthermore, Size$(D) \leq$ Size$(C) \cdot \frac{6k}{\varepsilon} \log(\frac{6k}{\varepsilon})$, and the running time of Gen is polynomial in $k, \frac{1}{\varepsilon}, n$ with oracle calls.

First, we notice that the above definition does not impose any restrictions on the time complexity of the poser and the solver. We consider a general approach where *Gen* is used to define a polynomial time reduction between a solver for the k-wise direct product of puzzles to a solver for a single puzzle. Furthermore, in the previous sections we considered solvers for the k-wise direct product that were compared with the algorithms that either solves all puzzles ([CHS04]) or allowed a fraction of puzzles to be solved incorrectly ([DIJK09]). In the above definition a more general case is considered where we use a binary monotone function g. More precisely we are interested functions that are binary monotonously non-decreasing.

The proof technique used by T.Holenstein and G.Schoenebeck is similar to the one presented in Section 5.1. In chapter 6 we use very similar approach and fix puzzles on consecutive coordinates of the n-wise direct product.

Chapter 6

Hardness Amplification for Dynamic Interactive Weakly Verifiable Puzzles

In the previous chapter we gave an overview of the former studies of different types of weakly verifiable puzzles. The focus of this chapter is on giving a constructive proof of hardness amplification for dynamic interactive weakly verifiable puzzles. In Section **??** we formulate the main theorem which is then proved in the succeeding sections. We begin, in Section 6.1.2, by constructing an algorithm that finds an efficiently computable function used to partition the domain of hint and verification queries. Then, in Section 6.1.3 we prove hardness amplification of dynamic interactive weakly verifiable puzzles under the assumption that for partitioned domains the success probability of a solver for the product of puzzles is still substantial. Finally, in Section 6.1.4 we complete the proof by combining the previous steps.

6.1 Main Theorem

A verification query (q, y) of a solver C for which a hint query on this q has been asked before cannot be a verification query for which C succeeds. Therefore, without loss of generality, we make the assumption that C does not ask verification queries on q for which a hint query has been asked before. Moreover, we assume that once C asked a verification query that succeeds, it does not ask any further hint or verification queries.

Experiment $\mathit{Success}^{P,C}(\pi, \rho)$

Oracle: A problem poser P, a solver $C = (C_1, C_2)$ for P.
Input: Bitstrings $\pi \in \{0,1\}^n$, $\rho \in \{0,1\}^*$.

Output: A bit $b \in \{0,1\}$.

---

run $\langle P(\pi), C_1(\rho)\rangle$
  $(\Gamma_V, \Gamma_H) := \langle P(\pi), C_1(\rho)\rangle_P$
  $x := \langle P(\pi), C_1(\rho)\rangle_{trans}$

run $C_2^{\Gamma_V,\Gamma_H}(x,\rho)$
  if $C_2^{\Gamma_V,\Gamma_H}(x,\rho)$ asks a verification query (q,y) s.t. $\Gamma_V(q,y) = 1$ **then**
    return 1
return 0

We define the *success probability* of C in solving a puzzle defined by P as

$$\Pr_{\pi,\rho}[Success^{P,C}(\pi,\rho) = 1]. \tag{6.1}$$

Furthermore, we say that C *succeeds* for π, ρ if $Success^{P,C}(\pi,\rho) = 1$.

Theorem 6.1 ***(Hardness amplification for dynamic weakly verifiable puzzles.)*** *Let $P_n^{(1)}$ be a fixed problem poser as in Definition 3.1 and $P_{kn}^{(g)}$ a problem poser for the k-wise direct product of $P_n^{(1)}$. Additionally, let C be a solver for $P_{kn}^{(g)}$ asking at most h hint queries and v verification queries. There exists a probabilistic algorithm $\widetilde{Gen}$ with oracle access to C, a monotone function $g : \{0,1\}^k \to \{0,1\}$, and problem posers $P_n^{(1)}$, $P_{kn}^{(g)}$. Furthermore, $\widetilde{Gen}$ takes as input parameters ε, δ, n, k, h, v, and outputs a solver circuit $\widetilde{D}$ for $P_n^{(1)}$ such that the following holds:*
If C is such that

$$\Pr_{\substack{\pi^{(k)}\in\{0,1\}^{kn}\\ \rho\in\{0,1\}^*}}\left[Success^{P_{kn}^{(g)},C}(\pi^{(k)},\rho) = 1\right] \geq 16(h+v)\Big(\Pr_{u\leftarrow\mu_\delta^k}[g(u) = 1] + \varepsilon\Big),$$

then $\widetilde{D}$ is a two phase probabilistic circuit that with high probability satisfies

$$\Pr_{\substack{\pi\in\{0,1\}^n\\ \rho\in\{0,1\}^*}}\left[Success^{P_n^{(1)},\widetilde{D}}(\pi,\rho) = 1\right] \geq \delta + \frac{\varepsilon}{6k}.$$

Additionally, $\widetilde{D}$ requires oracle access to g, $P_n^{(1)}$, C, hint and verification circuits generated by the problem poser $P_n^{(1)}$ after the first phase and asks at most $\frac{6k}{\varepsilon}\log\left(\frac{6k}{\varepsilon}\right)h$ hint queries and one verification query. Finally, $Time(\widetilde{Gen}) = poly(k, \frac{1}{\varepsilon}, n, v, h)$ with oracle calls.

We emphasize that the number of hint queries asked by D is greater than the number of hint queries asked by C whereas the number of verification queries is limited to at most one. In many applications, making such an assumption about the number of hint and verification queries is reasonable. In particular, we cannot assume that a solver for a single puzzle may ask more verification queries than a solver for the k-wise direct product.

6.1.1 The intuition

We refer to the puzzle solved by a circuit $\widetilde{D}$ as an *input puzzle*. The idea is to use a solver C for the k-wise direct product of $P^{(1)}$ with the input puzzle placed on one of the k coordinates. The puzzles on the remaining $k-1$ coordinates are generated by $\widetilde{\mathrm{Gen}}$ and $\widetilde{D}$ and chosen in such a way that the input puzzle is solved with substantial probability. The approach used in this Thesis is similar to the one in [CHS04, HS10] and briefly described in Section 5.

We fix randomness used to generate puzzles on the consecutive coordinates of the k-wise direct product of $P^{(1)}$ as long as the success probability of C on the remaining puzzles is still substantial. More precisely, the success probability of C with the first puzzle fixed for the remaining $k-1$ puzzles should satisfy the assumptions of Theorem 6.1 for the $(k-1)$-wise direct product of $P^{(1)}$.

If it is possible to find such randomness for the puzzle on the first position, then $\widetilde{\mathrm{Gen}}$ recursively solves the problem for the $(k-1)$–wise direct product of DIWVP. If $\widetilde{\mathrm{Gen}}$ reaches $k = 1$, then from the assumptions of Theorem 6.1 for the 1-wise direct product of puzzles the solver C can be easily used to successfully solve the input puzzle.

Unfortunately, we cannot exclude a situation where $\widetilde{\mathrm{Gen}}$ in one of the recursive calls does not find randomness that could be used to generate a puzzle on the first coordinate such that the success probability of C on the remaining coordinates is substantial. However, we make the following important observation which is very much in the same vein as Observation 5.4. If $\widetilde{\mathrm{Gen}}$ fails to find good randomness for the puzzle on the first position, then we suspect that the success probability of C on the remaining coordinates is no longer substantial. Furthermore, we also know that when all coordinates are considered, then C has substantial success probability. Intuitively, this means that the first coordinate is somehow important in the sense that C solves a puzzle on the first coordinate unusually often. Therefore, it is reasonable we place the input puzzle on the first position.

What is left is to find randomness used to generate puzzles on the remaining $(k-1)$ coordinates such that the puzzle on the first position is solved often. In Section 5.1 where we made Observation 5.4, we have seen that in the context of weakly verifiable puzzles it makes sense to consider a situation where all the remaining $k-1$ puzzles are correctly solved. Here, we generalize this approach. We fix randomness for the remaining $k-1$ puzzles such that when a puzzle on the first position was solved successfully, then the k–wise direct product would be solved successfully, and if a puzzle on the first position was unsuccessfully solved, then the k–wise direct product would be also solved unsuccessfully. It turns out that in case of a function g which requires all puzzles to be solved correctly, the above described approach corresponds exactly to the one presented in Section 5.1. Later in the proof of Theorem 6.1, we show that this

approach is indeed correct.

All puzzles except the input puzzle are generated by $\widetilde{\mathrm{Gen}}$ and $\widetilde{D}$. Therefore, it is possible to answer all hint and verification queries concerning these puzzles. For the input puzzle to answer hint and verification queries we have to use the hint and verification oracle respectively.

In Section **??** we have already described that the approach of fixing the puzzles on the consecutive coordinates may fail for dynamic weakly verifiable puzzles. The very similar arguments are also valid for the approach described above. More precisely, we would like to be sure that we never ask a hint query that prevents one of the (later) verification queries to succeed. To satisfy this requirement we partition the set Q into a set of advice queries and a set of attacking queries in the similar way as described in Section **??** and [DIJK09].

6.1.2 Domain partitioning

Let $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$, the idea is to partition $\mathcal{Q}$ such that the set of preimages of 0 for $hash$ contains $q \in \mathcal{Q}$ on which C is not allowed to ask hint queries, and the first successful verification query (q, y) of C is such that $hash(q) = 0$. Therefore, if C makes a verification query (q, y) such that $hash(q) = 0$, then we know that no hint query is ever asked on this q.

We denote the i-th query of C by q_i if it is a hint query and by (q_i, y_i) if it is a verification query. Let us define the following experiment *CanonicalSuccess* in which $\mathcal{Q}$ is partitioned using a function $hash$. We say that a solver circuit *succeeds* in the experiment *CanonicalSuccess* if it asks a successful verification query (q_j, y_j) such that $hash(q_j) = 0$, and no hint query q_i has been asked before (q_j, y_j) such that $hash(q_i) = 0$.

Experiment $CanonicalSuccess^{P,C,hash}(\pi, \rho)$

Oracle: A problem poser P, a solver circuit $C = (C_1, C_2)$ for P,
a function $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h + v) - 1\}$.
Input: Bitstrings $\pi \in \{0, 1\}^n$, $\rho \in \{0, 1\}^*$.
Output: A bit $b \in \{0, 1\}$.

run $\langle P(\pi), C_1(\rho)\rangle$
  $(\Gamma_V, \Gamma_H) := \langle P(\pi), C_1(\rho)\rangle_P$
  $x := \langle P(\pi), C_1(\rho)\rangle_{trans}$

run $C_2^{\Gamma_V,\Gamma_H}(x, \rho)$
  if $C_2^{\Gamma_V,\Gamma_H}(x, \rho)$ does not succeed for any verification query **then**
    return 0
  Let (q_j, y_j) be the first verification query such that $\Gamma_V(q_j, y_j) = 1$.

```
if (∀i < j : hash(q_i) ≠ 0) and (hash(q_j) = 0) then
    return 1
else
    return 0
```

We define the *canonical success probability* of a solver circuit C for P with respect to a function *hash* as

$$\Pr_{\pi,\rho}[\mathit{CanonicalSuccess}^{P,C,hash}(\pi,\rho) = 1]. \tag{6.2}$$

For fixed *hash* and P a *canonical success* of C for bistrings π, ρ is a situation where $\mathit{CanonicalSuccess}^{P,C,hash}(\pi,\rho) = 1$.

We show now that if a solver circuit C for P often succeeds in the experiment *Success*, then there exists a function *hash* such that C also often succeeds in the experiment *CanonicalSuccess* provided that the number of hint and verification queries is not too large.

Lemma 6.2 ***(Success probability in solving a dynamic interactive weakly verifiable puzzle with respect to a function hash.)*** *For a fixed problem poser P_n let C be a solver for P_n with the success probability at least γ, asking at most h hint queries and v verification queries. Moreover, let $\mathcal{H}$ be a family of pairwise independent efficient hash functions $\mathcal{Q} \to \{0,1,\ldots,2(h+v)-1\}$. There exists a probabilistic algorithm FindHash that takes as input parameters γ, n, h, v, and has oracle access to C and P_n. Furthermore, FindHash runs in time polynomial in $(h, v, \frac{1}{\gamma}, n)$ and with high probability outputs a function $hash \in \mathcal{H}$ such that the canonical success probability of C with respect to hash is at least $\frac{\gamma}{16(h+v)}$.*

Proof (of Lemma 6.2). We fix a problem poser P and a solver C for P in the whole proof of Lemma 6.2. For $k, l \in \{1,\ldots,(h+v)\}$ and $\alpha, \beta \in \{0,1,\ldots,2(h+v)-1\}$ by the pairwise independence property, we have that for every $q_k, q_l \in \mathcal{Q}$ such that $q_k \neq q_l$ the following holds

$$\begin{aligned}\Pr_{hash \leftarrow \mathcal{H}}[hash(q_k) = \alpha \mid hash(q_l) = \beta] &= \Pr_{hash \leftarrow \mathcal{H}}[hash(q_k) = \alpha] \\ &= \frac{1}{2(h+v)}.\end{aligned} \tag{6.3}$$

We write $\mathcal{P}_{Success}$ to denote a set containing all (π,ρ) for which $\mathit{Success}^{P,C}(\pi,\rho) = 1$. Let us fix $(\pi^*,\rho^*) \in \mathcal{P}_{Success}$. We are interested in the probability over a choice of *hash* of the event $\mathit{CanonicalSuccess}^{P,C,hash}(\pi^*,\rho^*) = 1$. Let (q_j, y_j) denote

the first verification query such that $\Gamma_V(q_j, y_j) = 1$, we have

$$\begin{aligned}
&\Pr_{hash \leftarrow \mathcal{H}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi^*, \rho^*) = 1\Big] \\
&\quad = \Pr_{hash \leftarrow \mathcal{H}}\big[\, hash(q_j) = 0 \wedge (\forall i < j : hash(q_i) \neq 0)\big] \\
&\quad = \Pr_{hash \leftarrow \mathcal{H}}\big[\forall i < j : hash(q_i) \neq 0 \mid hash(q_j) = 0\big] \Pr_{hash \leftarrow \mathcal{H}}[hash(q_j) = 0] \\
&\quad \overset{(6.3)}{=} \frac{1}{2(h+v)}\left(1 - \Pr_{hash \leftarrow \mathcal{H}}[\exists i < j : hash(q_i) = 0 \mid hash(q_j) = 0]\right) \\
&\quad \overset{(*)}{\geq} \frac{1}{2(h+v)}\left(1 - \sum_{i<j} \Pr_{hash \leftarrow \mathcal{H}}[hash(q_i) = 0 \mid hash(q_j) = 0]\right) \\
&\quad \overset{(6.3)}{=} \frac{1}{2(h+v)}\left(1 - \sum_{i<j} \Pr_{hash \leftarrow \mathcal{H}}[hash(q_i) = 0]\right) \\
&\quad \overset{(6.3)}{\geq} \frac{1}{4(h+v)}, \qquad (6.4)
\end{aligned}$$

where in $(*)$ we used the union bound. Let us denote the set of those (π, ρ) for which $\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1$ by $\mathcal{P}_{Canonical}$. If for π, ρ the circuit C succeeds canonically, then for the same π, ρ we also have $\mathit{Success}^{P,C}(\pi, \rho) = 1$. Hence, $\mathcal{P}_{Canonical} \subseteq \mathcal{P}_{Success}$, and we conclude

$$\begin{aligned}
&\Pr_{\substack{hash \leftarrow \mathcal{H} \\ \pi, \rho}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1\Big] \\
&\quad = \Pr_{\substack{hash \leftarrow \mathcal{H} \\ \pi, \rho}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1 \mid (\pi, \rho) \in \mathcal{P}_{Success}\Big] \Pr_{\pi, \rho}[(\pi, \rho) \in \mathcal{P}_{Success}] \\
&\qquad + \Pr_{\substack{hash \leftarrow \mathcal{H} \\ \pi, \rho}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1 \mid (\pi, \rho) \notin \mathcal{P}_{Success}\Big] \Pr_{\pi, \rho}[(\pi, \rho) \notin \mathcal{P}_{Success}] \\
&\quad = \Pr_{\substack{hash \leftarrow \mathcal{H} \\ \pi, \rho}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1 \mid (\pi, \rho) \in \mathcal{P}_{Success}\Big] \Pr_{\pi, \rho}[(\pi, \rho) \in \mathcal{P}_{Success}] \\
&\quad \geq \Pr_{\substack{hash \leftarrow \mathcal{H} \\ \pi, \rho}}\Big[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1 \mid (\pi, \rho) \in \mathcal{P}_{Success}\Big] \cdot \gamma \\
&\quad = \mathop{\mathbb{E}}_{(\pi, \rho) \in \mathcal{P}_{Success}}\left[\Pr_{hash \leftarrow \mathcal{H}}[\mathit{CanonicalSuccess}^{P,C,hash}(\pi, \rho) = 1]\right] \cdot \gamma \\
&\quad \overset{(6.4)}{\geq} \frac{\gamma}{4(h+v)}. \qquad (6.5)
\end{aligned}$$

Algorithm FindHash$^{P,C}(\gamma, n, h, v)$

Oracle: A problem poser P, a solver circuit C for P.
Input: Parameters γ, n. The number of hint queries h and of verification queries v.
Output: A function $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$.

```
for i := 1 to 32n(h+v)^2/γ^2 do:
    hash ← H
    count := 0
    for j := 1 to 32n(h+v)^2/γ^2 do:
        π ← {0,1}^n
        ρ ← {0,1}*
        if CanonicalSuccess^{P,C,hash}(π, ρ) = 1 then
            count := count + 1
    if count ≥ γ/(12(h+v)) · 32(h+v)^2/γ^2 · n then
        return hash
return ⊥
```

We show that FindHash chooses $hash \in \mathcal{H}$ such that the canonical success probability of C with respect to $hash$ is at least $\frac{\gamma}{16(h+v)}$ almost surely. Let $\mathcal{H}_{Good}$ denote a family of functions $hash \in \mathcal{H}$ for which

$$\Pr_{\pi,\rho}\left[CanonicalSuccess^{P,C,hash}(\pi,\rho) = 1\right] \geq \frac{\gamma}{8(h+v)}, \tag{6.6}$$

and $\mathcal{H}_{Bad}$ be a family of functions $hash \in \mathcal{H}$ such that

$$\Pr_{\pi,\rho}\left[CanonicalSuccess^{P,C,hash}(\pi,\rho) = 1\right] \leq \frac{\gamma}{16(h+v)}. \tag{6.7}$$

Let N denote the number of iterations of the inner loop of FindHash. We consider a single iteration of the outer loop of FindHash in which $hash$ is fixed. Let us define independent and identically distributed binary random variables $X_1, \ldots, X_N$ such that

$$X_i = \begin{cases} 1 & \text{if in the } i\text{-th iteration of the inner loop } count \text{ is increased} \\ 0 & \text{otherwise.} \end{cases}$$

We turn now to the case where $hash \in \mathcal{H}_{Bad}$ and show that it is unlikely that $hash \in \mathcal{H}_{Bad}$ is returned by FindHash. From (6.7) it follows that $\mathbb{E}_{\pi,\rho}[X_i] \leq \frac{\gamma}{16(h+v)}$. In the following inequalities (6.8) and (6.9) in steps denoted with $(*)$ we use the trivial facts $h + v \geq 1$ and $\gamma \leq 1$. For any fixed $hash \in \mathcal{H}_{Bad}$ using the

Chernoff bound we get

$$\Pr_{\pi,\rho}\left[\frac{1}{N}\sum_{i=1}^{N} X_i \geq \frac{\gamma}{12(h+v)}\right] \leq \Pr_{\pi,\rho}\left[\frac{1}{N}\sum_{i=1}^{N} X_i \geq \big(1+\frac{1}{3}\big)\mathbb{E}[X_i]\right]$$
$$\leq e^{-\frac{\gamma}{16(h+v)}N/27} \leq e^{-\frac{2}{27}\frac{(h+v)}{\gamma}n} \overset{(*)}{\leq} e^{-\frac{2}{27}n}. \quad (6.8)$$

The probability that $hash \in \mathcal{H}_{Good}$, when picked, is not returned amounts

$$\Pr_{\pi,\rho}\left[\frac{1}{N}\sum_{i=1}^{N} X_i \leq \frac{\gamma}{12(h+v)}\right] \leq \Pr_{\pi,\rho}\left[\frac{1}{N}\sum_{i=1}^{N} X_i \leq \big(1-\frac{1}{3}\big)\mathbb{E}[X_i]\right]$$
$$\leq e^{-\frac{\gamma}{8(h+v)}N/18} \leq e^{-\frac{2}{9}\frac{(h+v)}{\gamma}n} \overset{(*)}{\leq} e^{-\frac{2}{9}n} \quad (6.9)$$

where we once more used the Chernoff bound. We show now that the probability of picking $hash \in \mathcal{H}_{Good}$ is at least $\frac{\gamma}{8(h+v)}$. To obtain a contradiction suppose that

$$\Pr_{hash \leftarrow \mathcal{H}}[hash \in \mathcal{H}_{Good}] < \frac{\gamma}{8(h+v)}. \quad (6.10)$$

Using (6.10) we conclude that it is possible to bound the probability of the canonical success as follows

$$\begin{aligned}
&\Pr_{\substack{hash \leftarrow \mathcal{H}\\ \pi,\rho}}[CanonicalSuccess^{P,C,hash}(\pi,\rho)=1] \\
&\quad= \Pr_{\substack{hash \leftarrow \mathcal{H}\\ \pi,\rho}}[CanonicalSuccess^{P,C,hash}(\pi,\rho)=1 \mid hash \in \mathcal{H}_{Good}] \Pr_{hash \leftarrow \mathcal{H}}[hash \in \mathcal{H}_{Good}] \\
&\qquad + \Pr_{\substack{hash \leftarrow \mathcal{H}\\ \pi,\rho}}[CanonicalSuccess^{P,C,hash}(\pi,\rho)=1 \mid hash \notin \mathcal{H}_{Good}] \Pr_{hash \leftarrow \mathcal{H}}[hash \notin \mathcal{H}_{Good}] \\
&\quad\leq \Pr_{hash \leftarrow \mathcal{H}}[hash \in \mathcal{H}_{Good}] + \Pr_{\substack{hash \leftarrow \mathcal{H}\\ \pi,\rho}}[CanonicalSuccess^{P,C,hash}(\pi,\rho)=1 \mid hash \notin \mathcal{H}_{Good}] \\
&\quad\overset{\substack{(6.6)\\(6.10)}}{<} \frac{\gamma}{8(h+v)} + \frac{\gamma}{8(h+v)} = \frac{\gamma}{4(h+v)},
\end{aligned}$$

which contradicts (6.5). Hence, we conclude that the probability of choosing $hash \in \mathcal{H}_{Good}$ amounts at least $\frac{\gamma}{8(h+v)}$.

We show that FindHash picks in one of its iteration $hash \in \mathcal{H}_{Good}$ almost surely. Let K denote the random variable that takes value equal to the number of iterations of the outer loop of FindHash and Y_i be a random variable for the event that in the i-th iteration of the outer loop $hash \notin \mathcal{H}_{Good}$ is picked.

We use $\Pr_{hash \leftarrow \mathcal{H}}[hash \in \mathcal{H}_{Good}] \geq \frac{\gamma}{8(h+v)}$ and $K \leq \frac{32(h+v)^2}{\gamma^2} n$ to conclude

$$\Pr_{hash \leftarrow \mathcal{H}}\Big[\bigcap_{1 \leq i \leq K} Y_i\Big] \leq \left(1 - \frac{\gamma}{8(h+v)}\right)^{\frac{32(h+v)^2}{\gamma^2} n} \leq e^{-\frac{\gamma}{8(h+v)} \frac{32(h+v)^2}{\gamma^2} n}$$
$$\leq e^{-\frac{4(h+v)}{\gamma} n} \leq e^{-n}.$$

It is clear that the running time of FindHash is $poly(n, h, v, \gamma)$ with oracle calls. This finishes the proof of Lemma 6.2. □

6.1.3 The hardness amplification proof for partitioned domains

Let $C = (C_1, C_2)$ be a solver circuit for a dynamic interactive weakly verifiable puzzle as in Definition 3.1. We write $C_2^{(\cdot,\cdot)}$ to emphasize that C_2 does not obtain direct access to hint and verification circuits. Instead, whenever C_2 asks a hint or verification query, it is answered explicitly as in the following code excerpt of the circuit $\widetilde{C}_2$.

Given $C = (C_1, C_2)$ we define the circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$. Every hint query q asked by $\widetilde{C}$ is such that $hash(q) \neq 0$. Furthermore, $\widetilde{C}$ asks no verification queries. Instead, it returns (q, y) such that $hash(q) = 0$ or $\perp$.

Circuit $\widetilde{C}_2^{\Gamma_H, C_2, hash}(x, \rho)$

Oracle: A hint circuit Γ_H, a circuit C_2,
a function $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v) - 1\}$.
Input: Bitstrings $x \in \{0,1\}^*$, $\rho \in \{0,1\}^*$.
Output: A pair (q, y) where $q \in \mathcal{Q}$ and $y \in \{0,1\}^*$.

```
run C_2^(·,·)(x, ρ)
    if C_2^(·,·)(x, ρ) asks a hint query on q then
        if hash(q) = 0 then
            return ⊥
        else
            answer the query of C_2^(·,·)(x, ρ) using Γ_H(q)

    if C_2^(·,·)(x, ρ) asks a verification query (q, y) then
        if hash(q) = 0 then
            return (q, y)
        else
            answer the verification query of C_2^(·,·)(x, ρ) with 0
return ⊥
```

For fixed π, ρ, and *hash* we say that the circuit $\widetilde{C}$ *succeeds* if for $x := \langle P(\pi), C_1(\rho)\rangle_{trans}$, $(\Gamma_V, \Gamma_H) := \langle P(\pi), C_1(\rho)\rangle_P$, we have

$$\Gamma_V(\widetilde{C}_2^{\Gamma_H, C_2, hash}(x, \rho)) = 1.$$

Lemma 6.3 *For fixed P, C, and hash the following statement is true*

$$\Pr_{\pi,\rho}[CanonicalSuccess^{P,C,hash}(\pi,\rho) = 1] \leq \Pr_{\substack{\pi,\rho \\ x:=\langle P(\pi),C_1(\rho)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P(\pi),C_1(\rho)\rangle_P}}[\Gamma_V(\widetilde{C}_2^{\Gamma_H,C_2,hash}(x,\rho)) = 1].$$

Proof. If for some π, ρ, and *hash* a circuit $C = (C_1, C_2)$ succeeds canonically, then for the same π, ρ, and *hash* a circuit $\widetilde{C} := (C_1, \widetilde{C}_2)$ also succeeds. Using this observation, we conclude that

$$\begin{aligned}
&\Pr_{\pi,\rho}\Big[CanonicalSuccess^{P,C,hash}(\pi,\rho) = 1\Big] \\
&\qquad \leq \mathop{\mathbb{E}}_{\substack{\pi,\rho \\ x:=\langle P(\pi),C_1(\rho)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P(\pi),C_1(\rho)\rangle_P}}\big[\Gamma_V(\widetilde{C}_2^{\Gamma_H,C_2,hash}(x,\rho)) = 1\big] \\
&\qquad = \Pr_{\substack{\pi,\rho \\ x:=\langle P(\pi),C_1(\rho)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P(\pi),C_1(\rho)\rangle_P}}\big[\Gamma_V(\widetilde{C}_2^{\Gamma_H,C_2,hash}(x,\rho)) = 1\big]
\end{aligned}$$

□

Lemma 6.4 ***(Hardness amplification for a dynamic interactive weakly verifiable puzzle with respect to hash.)*** *Let $g : \{0,1\}^k \to \{0,1\}$ be a monotone function, $P_n^{(1)}$ a fixed problem poser and $\widetilde{C} := (C_1, \widetilde{C}_2)$ a circuit with oracle access to a function hash : $\mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$ and a solver $C := (C_1, C_2)$ for $P_{kn}^{(g)}$ which asks at most h hint queries and v verification queries. There exists an algorithm Gen that takes as input parameters ε, δ, n, k, has oracle access to $P_n^{(1)}$, $\widetilde{C}$, hash, g, and outputs a circuit $D := (D_1, D_2)$ such that the following holds:*
If $\widetilde{C}$ is such that

$$\Pr_{\substack{\pi^{(k)}\in\{0,1\}^{kn},\rho\in\{0,1\}^* \\ x:=\langle P^{(g)}(\pi^{(k)}),C_1(\rho)\rangle_{trans} \\ (\Gamma_H^{(k)},\Gamma_V^{(g)}):=\langle P^{(g)}(\pi^{(k)}),C_1(\rho)\rangle_{P^{(g)}}}}[\Gamma_V^{(g)}(\widetilde{C}_2^{\Gamma_H^{(k)},C_2,hash}(x,\rho)) = 1] \geq \Pr_{u\leftarrow\mu_\delta^k}[g(u) = 1] + \varepsilon,$$

then D satisfies almost surely

$$\Pr_{\substack{\pi\in\{0,1\}^n,\rho\in\{0,1\}^* \\ x:=\langle P^{(1)}(\pi),D_1^{\widetilde{C}}(\rho)\rangle_{trans} \\ (\Gamma_H,\Gamma_V):=\langle P^{(1)}(\pi),D_1^{\widetilde{C}}(\rho)\rangle_{P^{(1)}}}}[\Gamma_V(D_2^{P^{(1)},\widetilde{C},hash,g,\Gamma_H}(x,\rho)) = 1] \geq \delta + \frac{\varepsilon}{6k}.$$

Furthermore, D asks at most $\frac{6k}{\varepsilon}\log\left(\frac{6k}{\varepsilon}\right)h$ hint queries and no verification queries. Finally, the running time of Gen is polynomial in $k, \frac{1}{\varepsilon}, n$ with oracle calls.

We note that the circuit D from Lemma 6.4 instead of asking verification queries outputs a pair (q, y) such that $hash(q) = 0$ or $\perp$.

Before we give the proof of Lemma 6.4 we define additional algorithms. First, in the following code listing the algorithm Gen from Lemma 6.4 is defined. The procedures used by Gen are presented on the succeeding code listings.

Algorithm $\mathrm{Gen}^{P^{(1)},\widetilde{C},g,hash}(\varepsilon,\delta,n,k)$

Oracle: A poser $P^{(1)}$, a solver $\widetilde{C}$ for $P^{(g)}$, functions $g : \{0,1\}^k \to \{0,1\}$, $hash : \mathcal{Q} \to \{0,1,\ldots,2(h+v)-1\}$.
Input: Parameters ε, δ, n, k.
Output: A circuit D.

---

for $i := 1$ **to** $\frac{6k}{\varepsilon}n$ **do:**
  $\pi^* \xleftarrow{\$} \{0,1\}^n$
  $\widetilde{S}_{\pi^*,0} := \mathrm{EstimateSurplus}^{P^{(1)},\widetilde{C},g,hash}(\pi^*,0,k,\varepsilon,\delta,n)$
  $\widetilde{S}_{\pi^*,1} := \mathrm{EstimateSurplus}^{P^{(1)},\widetilde{C},g,hash}(\pi^*,1,k,\varepsilon,\delta,n)$
  if $\exists b \in \{0,1\} : \widetilde{S}_{\pi^*,b} \geq (1-\frac{3}{4k})\varepsilon$ **then**
    Let C_1' have oracle access to $\widetilde{C}$, and have hard-coded π^*.
    Let $\widetilde{C}_2'$ have oracle access to $\widetilde{C}$, and have hard-coded π^*.
    $\widetilde{C}' := (C_1', \widetilde{C}_2')$
    $g'(b_2,\ldots,b_k) := g(b,b_2,\ldots,b_k)$
    return $Gen^{P^{(1)},\widetilde{C}',g',hash}(\varepsilon,\delta,n,k-1)$
// all estimates are lower than $(1-\frac{3}{4k})\varepsilon$
return $D^{P^{(1)},\widetilde{C},hash,g}$

We are interested in the probability that for $u \leftarrow \mu_\delta^k$ and a bit b we have $g(b,u_2,\ldots,u_k) = 1$. The estimate of this probability is calculated by the algorithm EstimateFunctionProbability.

Algorithm $\mathrm{EstimateFunctionProbability}^g(b,k,\varepsilon,\delta,n)$

Oracle: A function $g : \{0,1\}^k \to \{0,1\}$.
Input: A bit $b \in \{0,1\}$, parameters k, ε, δ, n.
Output: An estimate $\widetilde{g}_b$ of $\Pr_{u\leftarrow\mu_\delta^k}[g(b,u_2,\ldots,u_k)=1]$.

---

for $i := 1$ **to** $N := \frac{64k^2}{\varepsilon^2}n$ **do:**

$u \leftarrow \mu_\delta^k$
$g_i := g(b, u_2, \ldots, u_k)$
return $\frac{1}{N}\sum_{i=1}^{N} g_i$

For fixed $\pi^{(k)}$, ρ, and *hash* we say that the circuit $\widetilde{C} := (C_1, \widetilde{C}_2)$ *succeeds on the i-th coordinate* if for $x := \langle P^{(g)}(\pi^{(k)}), C_1(\rho)\rangle_{trans}$, $(\Gamma_V^{(g)}, \Gamma_H^{(k)}) := \langle P^{(g)}(\pi), C_1(\rho)\rangle_{P^{(g)}}$ and $(q, y^{(k)}) := \widetilde{C}_2(x, \rho)$ we have

$$\Gamma_V^i(q, y_i) = 1.$$

Lemma 6.5 *The algorithm EstimateFunctionProbability$^g(b, k, \varepsilon, \delta, n)$ outputs an estimate $\widetilde{g}_b$ such that $|\widetilde{g}_b - \Pr_{u\leftarrow\mu_\delta^k}[g(b, u_2, \ldots, u_k) = 1]| \leq \frac{\varepsilon}{8k}$ almost surely.*

Proof. We fix notation as in the code excerpt of the algorithm EstimateFunctionProbability. Let us define independent and identically distributed binary random variables $K_1, K_2, \ldots, K_N$ such that for each $1 \leq i \leq N$ the random variable K_i takes value g_i. We use the Chernoff bound to obtain

$$\Pr\Big[\Big|\widetilde{g}_b - \Pr_{u\leftarrow\mu_\delta^k}[g(b, u_2, \ldots, u_k) = 1]\Big| \geq \frac{\varepsilon}{8k}\Big]$$
$$= \Pr\Big[\Big|\Big(\frac{1}{N}\sum_{i=1}^{N} K_i\Big) - \mathbb{E}_{u\leftarrow\mu_\delta^k}[g(b, u_2, \ldots, u_k)]\Big| \geq \frac{\varepsilon}{8k}\Big] \leq 2 \cdot e^{-n/3}. \square$$

The algorithm EvalutePuzzles$^{P^{(1)}, \widetilde{C}, hash}(\pi^{(k)}, \rho, n, k)$ evaluates which of the k puzzles of the k-wise direct product of $P^{(1)}$ are solved successfully by $\widetilde{C}(\rho) := (C_1, \widetilde{C}_2)(\rho)$. To decide whether the i-th puzzle of the k-wise direct product is solved successfully we need to gain access to the verification circuit for the puzzle generated in the i-th round of the interaction between $P^{(g)}$ and $\widetilde{C}$. Therefore, the algorithm EvalutePuzzles runs k times $P^{(1)}(\pi_i)$ to simulate the interaction with $C_1(\rho)$ where in each round of interaction a fresh random bitstring $\pi_i \in \{0,1\}^n$ is used.

Let us introduce some additional notation. We denote by $\langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i$ the execution of the i-th round of the sequential interaction. We use $\langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i_{P^{(1)}}$ to denote the output of $P^{(1)}(\pi_i)$ in the i-th round. Finally, we write $\langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i_{trans}$ to denote the transcript of communication in the i-th round. We note that the i-th round of the interaction between $P^{(1)}$ and C_1 is well defined only if all previous rounds have been executed before.

For simplicity of notation in the code excerpts of circuits C_2, D_2, and EvalutePuzzles we omit superscripts of some oracles. Exemplary, we write $\widetilde{C}_2^{\Gamma_H^{(k)}, hash}$ instead of $\widetilde{C}_2^{\Gamma_H^{(k)}, C, hash}$ where the superscript of the oracle circuit C is omitted. We make sure that it is clear from the context which oracles are used.

Algorithm EvaluatePuzzles$^{P^{(1)},\widetilde{C},hash}(\pi^{(k)},\rho,n,k)$

Oracle: A problem poser $P^{(1)}$, a solver circuit $\widetilde{C}=(C_1,\widetilde{C}_2)$ for $P^{(g)}$, a function $hash:\mathcal{Q}\to\{0,1,\ldots,2(h+v)-1\}$.
Input: Bitstrings $\pi^{(k)}\in\{0,1\}^{kn}$, $\rho\in\{0,1\}^*$, parameters n, k.
Output: A tuple $(c_1,\ldots,c_k)\in\{0,1\}^k$.

for $i:=1$ **to** k **do:** *//simulate k rounds of interaction*
  $(\Gamma_V^i,\Gamma_H^i):=\langle P^{(1)}(\pi_i),C_1(\rho)\rangle^i_{P^{(1)}}$
  $x_i:=\langle P^{(1)}(\pi_i),C_1(\rho)\rangle^i_{trans}$
$x:=(x_1,\ldots,x_k)$
$\Gamma_H^{(k)}:=(\Gamma_H^1,\ldots,\Gamma_H^k)$
$(q,y_1,\ldots,y_k):=\widetilde{C}_2^{\Gamma_H^{(k)},hash}(x,\rho)$
if $(q,y_1,\ldots,y_k)=\perp$ **then**
  return $(0,\ldots,0)$
$(c_1,\ldots,c_k):=(\Gamma_V^1(q,y_1),\ldots,\Gamma_V^k(q,y_k))$
return $(c_1,\ldots,c_k)$

All puzzles used by EvalutePuzzles are generated internally. Thus, the algorithm can answer all queries of $\widetilde{C}_2$ itself.

We are interested in the success probability of $\widetilde{C}$ with the bitstring π_1 fixed to π^* where the fact whether $\widetilde{C}$ succeeds in solving the input puzzle defined by $P^{(1)}(\pi_1)$ placed on the first position is neglected, and instead a bit b is used. More formally, we define the surplus $S_{\pi^*,b}$ as

$$S_{\pi^*,b}=\Pr_{\pi^{(k)},\rho}\left[g(b,c_2,\ldots,c_k)=1\mid\pi_1=\pi^*\right]-\Pr_{u\leftarrow\mu_\delta^k}\left[g(b,u_2,\ldots,u_k)=1\right],\tag{6.11}$$

where $(c_2,c_3,\ldots,c_k)$ is obtained as in EvalutePuzzles.

The algorithm EstimateSurplus returns an estimate $\widetilde{S}_{\pi^*,b}$ for $S_{\pi^*,b}$.

Algorithm EstimateSurplus$^{P^{(1)},\widetilde{C},g,hash}(\pi^*,b,k,\varepsilon,\delta,n)$

Oracle: A problem poser $P^{(1)}$, a circuit $\widetilde{C}$ for $P^{(g)}$, functions $g:\{0,1\}^k\to\{0,1\}$ and $hash:\mathcal{Q}\to\{0,1,\ldots,2(h+v)-1\}$.
Input: A bistring $\pi^*\in\{0,1\}^n$, a bit $b\in\{0,1\}$, parameters k, ε, δ, n.
Output: An estimate $\widetilde{S}_{\pi^*,b}$ for $S_{\pi^*,b}$.

for $i:=1$ **to** $N:=\frac{64k^2}{\varepsilon^2}n$ **do:**
  $(\pi_2,\ldots,\pi_k)\xleftarrow{\$}\{0,1\}^{(k-1)n}$

$\rho \xleftarrow{\$} \{0,1\}^*$
$(c_1, \ldots, c_k) := \text{EvalutePuzzles}^{P^{(1)}, \widetilde{C}, hash}((\pi^*, \pi_2, \ldots, \pi_k), \rho, n, k)$
$\widetilde{s}^i_{\pi^*,b} := g(b, c_2, \ldots, c_k)$
$\widetilde{g}_b := \text{EstimateFunctionProbability}^g(b, k, \varepsilon, \delta, n)$
return $\left(\frac{1}{N}\sum_{i=1}^N \widetilde{s}^i_{\pi^*,b}\right) - \widetilde{g}_b$

Lemma 6.6 *The estimate $\widetilde{S}_{\pi^*,b}$ returned by EstimateSurplus differs from $S_{\pi^*,b}$ by at most $\frac{\varepsilon}{4k}$ almost surely.*

Proof. We use the union bound and similar argument as in Lemma 6.5 which yields that $\frac{1}{N}\sum_{i=1}^N \widetilde{s}^i_{\pi^*,b}$ differs from $\mathbb{E}[g(b, c_2, \ldots, c_k)]$ by at most $\frac{\varepsilon}{8k}$ almost surely. Together, with Lemma 6.5 we conclude that the surplus estimate returned by EstimateSurplus differs from $S_{\pi^*,b}$ by at most $\frac{\varepsilon}{4k}$ with probability at least $1 - 2e^{-n}$. □

We define the following solver circuit $C' = (C'_1, C'_2)$ for the $(k-1)$–wise direct product of $P^{(1)}$.

Circuit $C_1'^{\widetilde{C}, P^{(1)}}(\rho)$

Oracle: A solver circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$ for $P^{(g)}$, a poser $P^{(1)}$.
Input: A bitstring $\rho \in \{0,1\}^*$.
Hard-coded: A bitstring $\pi^* \in \{0,1\}^n$.

Simulate $\langle P^{(1)}(\pi^*), C_1(\rho)\rangle^1$
Use $C_1(\rho)$ for the remaining $k-1$ rounds of interaction.

Circuit $\widetilde{C}_2'^{\Gamma_H^{(k-1)}, \widetilde{C}, hash}(x^{(k-1)}, \rho)$

Oracle: A hint oracle $\Gamma_H^{(k-1)} := (\Gamma_H^2, \ldots, \Gamma_H^k)$,
a solver circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$ for $P^{(g)}$,
a function $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$.
Input: A transcript of $k-1$ rounds of interaction
$x^{(k-1)} := (x_2, \ldots, x_k) \in \{0,1\}^*$, a bitstring $\rho \in \{0,1\}^*$.
Hard-coded: A bitstring $\pi^* \in \{0,1\}^n$.

Simulate $\langle P^{(1)}(\pi^*), C_1(\rho)\rangle^1$
$(\Gamma_H^*, \Gamma_V^*) := \langle P^{(1)}(\pi^*), C_1(\rho)\rangle^1_{P^{(1)}}$
$x^* := \langle P^{(1)}(\pi^*), C_1(\rho)\rangle^1_{trans}$
$\Gamma_H^{(k)} := (\Gamma_H^*, \Gamma_H^2, \ldots, \Gamma_H^k)$
$x^{(k)} := (x^*, x_2, \ldots, x_k)$

$(q, y_1, \ldots, y_k) := \widetilde{C}_2^{\Gamma_H^{(k)}, hash}(x^{(k)}, \rho)$
return $(q, y_2, \ldots, y_k)$

We are ready to define the solver circuit $D = (D_1, D_2)$ for $P^{(1)}$ output by Gen.

Circuit $D_1^{\widetilde{C}}(r)$

Oracle: A solver circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$ for $P^{(g)}$.
Input: A pair $r := (\rho, \sigma)$ where $\rho \in \{0,1\}^*$ and $\sigma \in \{0,1\}^*$.

Interact with the problem poser $\langle P^{(1)}, C_1(\rho)\rangle^1$.
Let $x^* := \langle P^{(1)}, C_1(\rho)\rangle^1_{trans}$.

Circuit $D_2^{P^{(1)}, \widetilde{C}, hash, g, \Gamma_H}(x^*, r)$

Oracle: A poser $P^{(1)}$, a solver circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$ for $P^{(g)}$,
functions $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$, $g : \{0,1\}^k \to \{0,1\}$,
a hint circuit Γ_H for $P^{(1)}$.
Input: A communiation transcript $x^* \in \{0,1\}^*$, a bitstring $r := (\rho, \sigma)$
where $\rho \in \{0,1\}^*$ and $\sigma \in \{0,1\}^*$
Output: A pair (q, y^*).

---

for at most $\frac{6k}{\varepsilon}\log(\frac{6k}{\varepsilon})$ iterations **do:**
  $(\pi_2, \ldots, \pi_k) \leftarrow$ read next $(k-1) \cdot n$ bits from σ
  Use x^* to simulate the first round of interaction of $C_1(\rho)$
  with the problem poser $P^{(1)}$.
  for $i := 2$ **to** k **do:**
    run $\langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i$
      $(\Gamma_V^i, \Gamma_H^i) := \langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i_{P^{(1)}}$
      $x_i := \langle P^{(1)}(\pi_i), C_1(\rho)\rangle^i_{trans}$
  $\Gamma_H^{(k)}(q) := (\Gamma_H(q), \Gamma_H^2(q), \ldots, \Gamma_H^k(q))$
  $(q, y^*, y_2, \ldots, y_k) := \widetilde{C}_2^{\Gamma_H^{(k)}, hash}((x^*, x_2, \ldots, x_k), \rho)$
  $(c_2, \ldots, c_k) := (\Gamma_V^2(q, y_2), \ldots, \Gamma_V^k(q, y_k))$
  if $g(1, c_2, \ldots, c_k) = 1$ **and** $g(0, c_2, \ldots, c_k) = 0$ **then**
    return (q, y^*)
return $\perp$

Proof (of Lemma 6.4). First, let us consider the case where $k = 1$. The function $g : \{0,1\} \to \{0,1\}$ is either the identity or a constant function. In the

latter case, when g is a constant function, Lemma 6.4 is vacuously true. If g is the identity function, then the circuit D returned by Gen directly uses $\widetilde{C}$ to find a solution. From the assumptions of Lemma 6.4 it follows that $\widetilde{C}$ succeeds with probability at least $\delta + \varepsilon$. Hence, D trivially satisfies Lemma 6.4.

For the general case, we consider two possibilities. Namely, either Gen in one of the iterations finds an estimate with high surplus such that $\widetilde{S}_{\pi,b} \geq (1 - \frac{3}{4k})\varepsilon$ and recurses, or in all iterations it fails and outputs the circuit D.

If it is possible to find an estimate with high surplus, then we introduce a new monotone function $g' : \{0,1\}^{k-1} \to \{0,1\}$ such that $g'(b_2, \ldots, b_k) := g(b, b_2, \ldots, b_k)$ and a new circuit $\widetilde{C}' = (C_1', \widetilde{C}_2')$ with oracle access to $\widetilde{C} := (C_1, \widetilde{C}_2)$. W apply Lemma 6.6 and conclude that almost surely it holds

$$S_{\pi^*,b} \geq \widetilde{S}_{\pi^*,b} - \frac{\varepsilon}{4k} \geq \Big(1 - \frac{1}{k}\Big)\varepsilon.$$

It follows that $\widetilde{C}'$ succeeds in solving the $(k-1)$–wise direct product of puzzles with probability at least

$$\Pr_{u \leftarrow \mu_\delta^{(k-1)}}[g'(u_1, \ldots, u_{k-1})] + \Big(1 - \frac{1}{k}\Big)\varepsilon.$$

We see that in this case $\widetilde{C}'$ satisfies the conditions of Lemma 6.4 for the $(k-1)$–wise direct product of puzzles. Therefore, the recursive call to Gen with access to g' and $\widetilde{C}$ returns $D = (D_1, D_2)$ that with high probability satisfies

$$\Pr_{\substack{\pi,\rho \\ x:=\langle P^{(1)}(\pi), D_1^{\widetilde{C}}(\rho)\rangle_{trans} \\ (\Gamma_H,\Gamma_V):=\langle P^{(1)}(\pi), D_1^{\widetilde{C}}(\rho)\rangle_{P^{(1)}}}}\Big[\Gamma_V\big(D_2^{P^{(1)},\widetilde{C},hash,g,\Gamma_H}(x,\rho)\big) = 1\Big] \geq \delta + \Big(1 - \frac{1}{k}\Big)\frac{\varepsilon}{6(k-1)}$$
$$= \delta + \frac{\varepsilon}{6k}. \tag{6.12}$$

Let us bring our attention to the case where none of the estimates is greater than $(1 - \frac{3}{4k})\varepsilon$. If all surpluses $S_{\pi,0}$ and $S_{\pi,1}$ were lower than $(1 - \frac{1}{k})\varepsilon$, then it would mean that $\widetilde{C}$ does not succeed on the remaining $k-1$ puzzles with much higher probability than an algorithm that solves each puzzle independently with success probability δ. However, from the assumptions of Lemma 6.4 we know that on all k puzzles the success probability of $\widetilde{C}$ is higher. Hence, we suspect that the first puzzle is correctly solved unusually often. It remains to show that the fact that Gen fails to find a surplus estimate that is large implies that with high probability there are only few surpluses greater than $(1 - \frac{1}{k})\varepsilon$ and their influence is can be neglected. Additionally, we have to show that the circuit D finds outputs an answer almost surely.

We fix notation as in the code listing of the circuit D_2. Let us consider a single iteration of the outer loop of D_2 where values $\pi_2, \ldots, \pi_k$ are fixed. We write

π_1 to denote randomness used by the problem poser to generate the input puzzle. Furthermore, we define $c_1 := \Gamma_V(q, y_1)$ where Γ_V is the verification circuit generated by $P^{(1)}(\pi_1)$ in the first phase when interacting with $D_1(r)$. We write $c := (c_1, c_2, \ldots, c_k)$, and for $b \in \{0, 1\}$ we define a set

$$\mathcal{G}_b \;:=\; \big\{(b_1, b_2, \ldots, b_k) : g(b, b_2, \ldots, b_k) = 1\big\}.$$

Using this notation we express

$$\begin{aligned}\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_b] &= \Pr_{u \leftarrow \mu_\delta^k}[g(b, u_2, \ldots, u_k) = 1] \\ \Pr_{\pi^{(k)}, \rho}[c \in \mathcal{G}_b] &= \Pr_{\pi^{(k)}, \rho}[g(b, c_2, \ldots, c_k) = 1]. \end{aligned} \tag{6.13}$$

Let us fix randomness π_1 used by the problem poser to generate the input puzzle to π^*. We use (6.11) and (6.13) to obtain

$$\begin{aligned}&\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1] - \Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_0] \\ &= \Pr_{\pi^{(k)}, \rho}[c \in \mathcal{G}_1 \mid \pi_1 = \pi^*] - \Pr_{\pi^{(k)}, \rho}[c \in \mathcal{G}_0 \mid \pi_1 = \pi^*] - (S_{\pi^*,1} - S_{\pi^*,0}) \end{aligned} \tag{6.14}$$

By monotonicity of g it holds $\mathcal{G}_0 \subseteq \mathcal{G}_1$, and we write (6.14) as

$$\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0] = \Pr_{\pi^{(k)}, \rho}[c \in \mathcal{G}_1 \setminus \mathcal{G}_0 \mid \pi_1 = \pi^*] - (S_{\pi^*,1} - S_{\pi^*,0}). \tag{6.15}$$

Let us multiply both sides of (6.15) by

$$\Pr_{\substack{r \\ x^* := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{trans} \\ (\Gamma_V, \Gamma_H) := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*, r)) = 1] \;\Big/\; \Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0],$$

which yields

$$\begin{aligned}&\Pr_{\substack{r \\ x^* := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{trans} \\ (\Gamma_V, \Gamma_H) := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*, r)) = 1] \\ &\qquad = \Pr_{\substack{r \\ x^* := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{trans} \\ (\Gamma_V, \Gamma_H) := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*, r)) = 1] \Pr_{\pi^{(k)}, \rho}[c \in \mathcal{G}_1 \setminus \mathcal{G}_0 \mid \pi_1 = \pi^*] \frac{1}{\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]} \\ &\qquad\quad - \Pr_{\substack{r \\ x^* := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{trans} \\ (\Gamma_V, \Gamma_H) := \langle P^{(1)}(\pi^*), D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*, r)) = 1](S_{\pi^*,1} - S_{\pi^*,0}) \frac{1}{\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]}. \end{aligned} \tag{6.16}$$

Let us study the first summand of (6.16). First, we have

$$\begin{aligned}
&\Pr_{\substack{r \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r))=1] \\
&\quad = \Pr_{\substack{r \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r))=1 | D_2(x^*,r)\neq\perp]\Pr_{\substack{r \\ x^*=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}}}[D_2(x^*,r)\neq\perp] \\
&\quad \overset{(*)}{=} \Pr_{\pi^{(k)},\rho}[c_1=1 \mid c\in\mathcal{G}_1\setminus\mathcal{G}_0,\pi_1=\pi^*]\Pr_{\substack{r \\ x^*=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}}}[D_2(x^*,r)\neq\perp] \qquad (6.17)
\end{aligned}$$

where in $(*)$ we use the observation that $D_2(x^*,r)\neq\perp$ occurs if and only if $D_2(x^*,r)$ finds $\pi^{(k)}$ such that $c\in\mathcal{G}_1\setminus\mathcal{G}_0$. Furthermore, conditioned on $c\in\mathcal{G}_1\setminus\mathcal{G}_0$ we have that $\Gamma_V(D_2(x^*,r))=1$ happens if and only if $c_1=1$. Inserting (6.17) to the numerator of the first summand of (6.16) yields

$$\begin{aligned}
&\Pr_{\substack{r \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r))=1]\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*] \\
&\quad = \Pr_{\substack{r \\ x^*=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}}}[D_2(x^*,r)\neq\perp]\Pr_{\pi^{(k)},\rho}[c_1=1 \mid c\in\mathcal{G}_1\setminus\mathcal{G}_0,\pi_1=\pi^*]\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]. \qquad (6.18)
\end{aligned}$$

We consider the following two cases. First, if $\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]\leq\frac{\varepsilon}{6k}$ then

$$\Pr_{\pi^{(k)},\rho}[c_1=1 \mid c\in\mathcal{G}_1\setminus\mathcal{G}_0,\pi_1=\pi^*]\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]\leq\frac{\varepsilon}{6k}. \qquad (6.19)$$

In case when $\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]>\frac{\varepsilon}{6k}$ the circuit D_2 outputs $\perp$ if and only if it fails in all $\frac{6k}{\varepsilon}\log(\frac{6k}{\varepsilon})$ iterations to find $\pi^{(k)}$ such that $c\in\mathcal{G}_1\setminus\mathcal{G}_0$ which happens with probability

$$\Pr_{\substack{r \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}}}[D_2(x^*,r)=\perp]\leq\left(1-\frac{\varepsilon}{6k}\right)^{\frac{6k}{\varepsilon}\log(\frac{6k}{\varepsilon})}\leq\frac{\varepsilon}{6k}. \qquad (6.20)$$

We conclude that in both aforementioned cases using (6.18), (6.19) and (6.20) the following holds

$$\begin{aligned}
&\Pr_{\substack{r \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}}}[D_2(x^*,r)\neq\perp]\Pr_{\pi^{(k)},\rho}[c_1=1 \mid c\in\mathcal{G}_1\setminus\mathcal{G}_0,\pi_1=\pi^*]\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*] \\
&\quad \geq \Pr_{\pi^{(k)},\rho}[c_1=1 \mid c\in\mathcal{G}_1\setminus\mathcal{G}_0,\pi_1=\pi^*]\Pr_{\pi^{(k)},\rho}[c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]-\frac{\varepsilon}{6k} \\
&\quad = \Pr_{\pi^{(k)},\rho}[c_1=1\wedge c\in\mathcal{G}_1\setminus\mathcal{G}_0 \mid \pi_1=\pi^*]-\frac{\varepsilon}{6k}
\end{aligned}$$

$$
\begin{aligned}
&= \Pr_{\pi^{(k)},\rho}[g(c)=1 \mid \pi_1=\pi^*] - \Pr_{\pi^{(k)},\rho}[c \in \mathcal{G}_0 \mid \pi_1=\pi^*] - \frac{\varepsilon}{6k} \\
&\overset{(6.11)}{=} \Pr_{\pi^{(k)},\rho}[g(c)=1 \mid \pi_1=\pi^*] - \Pr_{u \leftarrow \mu_\delta^{(k)}}[u \in \mathcal{G}_0] - S_{\pi^*,0} - \frac{\varepsilon}{6k}.
\end{aligned}
\tag{6.21}
$$

We insert (6.21) into (6.16) and calculate the expected value over π^* which yields

$$
\begin{aligned}
\Pr_{\substack{\pi,r \\ x:=\langle P^{(1)}(\pi),D_1(r)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x,r))=1] &\geq \mathbb{E}_{\pi^*}\left[\frac{\Pr_{\pi^{(k)},\rho}[g(c)=1|\pi_1=\pi^*] - \Pr_{u\leftarrow\mu_\delta^{(k)}}[u\in\mathcal{G}_0] - \frac{\varepsilon}{6k}}{\Pr_{u\leftarrow\mu_\delta^{(k)}}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]}\right] \\
&- \mathbb{E}_{\substack{\pi^* \\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}\left[\left(\Pr_r[\Gamma_V(D_2(x^*,r))=1](S_{\pi^*,1}-S_{\pi^*,0})+S_{\pi^*,0}\right)\frac{1}{\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]}\right].
\end{aligned}
\tag{6.22}
$$

We show now that if Gen does not recurse, then the majority of estimates is low almost surely. Let us assume that

$$
\Pr_{\pi}\left[\left(S_{\pi,0} \leq (1-\frac{1}{2k})\varepsilon\right) \wedge \left(S_{\pi,1} \leq (1-\frac{1}{2k})\varepsilon\right)\right] < 1-\frac{\varepsilon}{6k}, \tag{6.23}
$$

then Gen recurses almost surely, because the probability that Gen does not find $\widetilde{S}_{\pi,b} \geq (1-\frac{3}{4k})\varepsilon$ in all of the $\frac{6k}{\varepsilon}n$ iterations is at most

$$
\left(1-\frac{\varepsilon}{6k}\right)^{\frac{6k}{\varepsilon}n} \leq e^{-n}
$$

almost surely, where we used Lemma 6.6. Therefore, under the assumption that Gen does not recurse with high probability it holds

$$
\Pr_{\pi,\rho}\left[\left(S_{\pi,0} \leq (1-\frac{1}{2k})\varepsilon\right) \wedge \left(S_{\pi,1} \leq (1-\frac{1}{2k})\varepsilon\right)\right] \geq 1-\frac{\varepsilon}{6k}. \tag{6.24}
$$

Let us define a set

$$
\mathcal{W} = \left\{\pi : \left(S_{\pi,0} \leq (1-\frac{1}{2k})\varepsilon\right) \wedge \left(S_{\pi,1} \leq (1-\frac{1}{2k})\varepsilon\right)\right\}. \tag{6.25}
$$

Additionally, let $\overline{\mathcal{W}}$ denote the complement of $\mathcal{W}$. We bound the numerator

of the second summand in (6.22)

$$\begin{aligned}
&\mathbb{E}_{\pi^*}\Big[S_{\pi^*,0} + \Pr_{\substack{r\\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r))=1](S_{\pi^*,1}-S_{\pi^*,0})\Big]\\
&\quad= \mathbb{E}_{\pi^*}\Big[S_{\pi^*,0} + \Pr_{\substack{r\\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r)=1](S_{\pi^*,1}-S_{\pi^*,0}) \mid \pi^*\in\overline{\mathcal{W}}\Big]\Pr_{\pi^*}[\pi^*\in\overline{\mathcal{W}}]\\
&\qquad+\mathbb{E}_{\pi^*}\Big[S_{\pi^*,0} + \Pr_{\substack{r\\ x^*:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,r))=1](S_{\pi^*,1}-S_{\pi^*,0}) \mid \pi^*\in\mathcal{W}\Big]\Pr_{\pi^*}[\pi^*\in\mathcal{W}]\\
&\quad\le \frac{\varepsilon}{6k}+\mathbb{E}_{\pi^*}\Big[S_{\pi^*,0} + \Pr_{\substack{r\\ x:=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{trans}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi^*),D_1(r)\rangle_{P^{(1)}}}}\big[\Gamma_V(D_2^{\widetilde{C}}(x^*,r))=1\big]\big((1-\frac{1}{2k})\varepsilon-S_{\pi^*,0}\big) \mid \pi^*\in\mathcal{W}\Big]\\
&\quad\le \frac{\varepsilon}{6k}+(1-\frac{1}{2k})\varepsilon=(1-\frac{1}{3k})\varepsilon. \qquad (6.26)
\end{aligned}$$

Finally, we insert (6.26) into (6.22) which yields

$$\Pr_{\substack{\pi,\rho\\ x:=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{\text{trans}}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{P^{(1)}}}}\big[\Gamma_V(D_2(x,\rho))=1\big]\ge \mathop{\mathbb{E}}_{\pi^*}\left[\frac{\Pr_{\pi^{(k)},\rho}[g(c)=1\mid\pi_1=\pi^*]-\Pr_{u\leftarrow\mu_\delta^k}[u\in G_0]-(1-\frac{1}{6k})\varepsilon}{\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]}\right].$$

From the assumptions of Lemma 6.4 it follows that

$$\Pr_{\pi^{(k)},\rho}[g(c)=1]\ge\Pr_{u\leftarrow\mu_\delta^{(k)}}[g(u)=1]+\varepsilon. \qquad (6.27)$$

We observe that

$$\begin{aligned}
\Pr_{u\leftarrow\mu_\delta^k}[g(u)=1]&=\Pr[u\in\mathcal{G}_0\vee(u\in\mathcal{G}_1\setminus\mathcal{G}_0\wedge u_1=1)]\\
&=\Pr[u\in\mathcal{G}_0]+\delta\Pr[u\in\mathcal{G}_1\setminus\mathcal{G}_0]. \qquad (6.28)
\end{aligned}$$

Using (6.28) and (6.27) we obtain

$$\begin{aligned}
\Pr_{\substack{\pi,\rho\\ x:=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{\text{trans}}\\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x,\rho))=1] &\ge \frac{\Pr_{u\leftarrow\mu_\delta^k}[g(u)=1]+\varepsilon-\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_0]-(1-\frac{1}{6k})\varepsilon}{\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]}\\
&\overset{(6.28)}{\ge}\frac{\varepsilon+\delta\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]-(1-\frac{1}{6k})\varepsilon}{\Pr_{u\leftarrow\mu_\delta^k}[u\in\mathcal{G}_1\setminus\mathcal{G}_0]}\ge\delta+\frac{\varepsilon}{6k}. \qquad (6.29)
\end{aligned}$$

Clearly, the running time of Gen is bounded by some polynomial $p(k,\frac{1}{\varepsilon},n)$ with oracle calls. Furthermore, the algorithm Gen outputs a circuit D such that it satisfies with probability at least $1-p(k,\frac{1}{\varepsilon},n)\cdot 2^n$ the statement of Lemma 6.4. This concludes the proof of Lemma 6.4. □

6.1.4 Putting it together

In the previous sections we have shown that it is possible to partition the domain Q such that if the number of hint and verification queries is small, then success probability of a solver for the k-wise direct product is still substantial. As shown in Lemma 6.4 we can build a circuit that returns a solution that is likely to be good. It remains to use these building blocks and prove Theorem 6.1.

Proof (of Theorem 6.1). Let $\widetilde{\mathrm{Gen}}$ be the algorithm from Theorem 6.1 that outputs the solver circuit $\widetilde{D}$ for P. First, in the following code listing we define $\widetilde{\mathrm{Gen}}$.

Algorithm $\widetilde{\mathrm{Gen}}^{P^{(1)},g,C}(n,\varepsilon,\delta,k,h,v)$

Oracle: A problem poser $P^{(1)}$, a function $g:\{0,1\}^k \to \{0,1\}$, a solver circuit C for $P^{(g)}$.
Input: Parameters n, ε, δ, k, h, v.
Return: A circuit $\widetilde{D} = (D_1, \widetilde{D}_2)$.

$hash := \mathrm{FindHash}((h+v)\varepsilon, n, h, v)$
Let $\widetilde{C} := (C_1, \widetilde{C}_2)$ be as in Lemma 6.3 with oracle access to C, $hash$.
$D := Gen^{P^{(1)},\widetilde{C},g,hash}(\varepsilon,\delta,n,k)$
return $\widetilde{D} := (D_1, \widetilde{D}_2)$

Circuit $\widetilde{D}_2^{D,P^{(1)},hash,g,\Gamma_V,\Gamma_H}(x,\rho)$

Oracle: A circuit $D := (D_1, \widetilde{D}_2)$ from Lemma 6.4, a problem poser $P^{(1)}$, functions $hash : Q \to \{0,1,\ldots,2(h+v)-1\}$, $g:\{0,1\}^k \to \{0,1\}$ a verification oracle Γ_V, a hint oracle Γ_H.
Input: Bitstrings $x \in \{0,1\}^*$, $\rho \in \{0,1\}^*$.

$(q,y) := D_2^{P^{(1)},\widetilde{C},hash,g,\Gamma_H}(x,\rho)$
Ask verification query (q,y) to Γ_V.

We show that Theorem 6.1 follows from Lemma 6.2 and Lemma 6.4. We fix $P^{(1)}$, g, $P^{(g)}$ in the whole proof and consider a solver circuit $C = (C_1, C_2)$, asking at most h hint queries and v verification queries such that

$$\Pr_{\pi^{(k)},\rho}\Big[Success^{P^{(g)},C}(\pi^{(k)},\rho)=1\Big] \geq 16(h+v)\Big(\Pr_{u\leftarrow\mu_\delta^k}[g(u)=1]+\varepsilon\Big).$$

First, we note that C meets the requirements of Lemma 6.2. Furthermore, trivially success probability of C is at least $(h+v)\varepsilon$. Thus, the algorithm $\widetilde{\text{Gen}}$ can call FindHash to obtain $hash : \mathcal{Q} \to \{0, 1, \ldots, 2(h+v)-1\}$ such that with high probability it holds

$$\Pr_{\pi^{(k)},\rho}\left[CanonicalSuccess^{P^{(g)},C,hash}(\pi^{(k)},\rho) = 1\right] \geq \Pr_{u \leftarrow \mu_\delta^k}[g(u) = 1] + \varepsilon.$$

Additionally, the running time of FindHash is $poly(h, v, \frac{1}{\varepsilon}, n)$ with oracle calls. Applying Lemma 6.3 for $C = (C_1, C_2)$ we obtain a circuit $\widetilde{C} = (C_1, \widetilde{C}_2)$ that satisfies

$$\Pr_{\substack{\pi^{(k)},\rho \\ x:=\langle P^{(g)}(\pi^{(k)}),C_1(\rho)\rangle_{trans} \\ (\Gamma_V^{(g)},\Gamma_H^{(k)}):=\langle P^{(g)}(\pi^{(k)}),C_1(\rho)\rangle_{P^{(g)}}}}[\Gamma_V^{(g)}(\widetilde{C}_2^{\Gamma_H^{(k)},C_2,hash}(x,\rho)) = 1] \geq \Pr_{u \leftarrow \mu_\delta^k}[g(u) = 1] + \varepsilon.$$

Now, we use the algorithm Gen as in Lemma 6.4 that yields a circuit $D = (D_1, D_2)$ which with high probability satisfies

$$\Pr_{\substack{\pi,\rho \\ x:=\langle P^{(1)}(\pi),D_1^{\widetilde{C}}(\rho)\rangle_{trans} \\ (\Gamma_H,\Gamma_V):=\langle P^{(1)}(\pi),D_1^{\widetilde{C}}(\rho)\rangle_{P^{(1)}}}}[\Gamma_V(D_2^{P^{(1)},\widetilde{C},hash,g,\Gamma_H}(x,\rho)) = 1] \geq (\delta + \frac{\varepsilon}{6k}). \tag{6.30}$$

Finally, $\widetilde{\text{Gen}}$ outputs $\widetilde{D} = (D_1, \widetilde{D}_2)$ with oracle access to D, $P^{(1)}$, $hash$, g such that with high probability it holds

$$\Pr_{\pi,\rho}\left[Success^{P^{(1)},\widetilde{D}}(\pi,\rho) = 1\right] \geq (\delta + \frac{\varepsilon}{6k}).$$

The running time of Gen is polynomial in $k, \frac{1}{\varepsilon}, n)$ with oracle calls. Thus, the overall running time of $\widetilde{Gen}$ is polynomial in $k, \frac{1}{\varepsilon}, h, v, n$ with oracle access. Furthermore, the circuit $\widetilde{D}$ asks at most $\frac{6k}{\varepsilon}\log(\frac{6k}{\varepsilon})h$ hint queries and one verification query. Finally, we have $Size(\widetilde{D}) \leq Size(C) \cdot \frac{6k}{\varepsilon}$. This finishes the proof of Theorem 6.1. □

6.2 Discussion

6.2.1 Optimality of the result

Appendix A

Appendix

A.1 Concentration Bounds

Lemma A.1 (Chernoff Bounds) *For independent and identically distributed Bernoulli random variables* $X_1, \ldots, X_n$ *with* $X := \sum_{i=1}^{n} X_i$ *and* $\Pr[X_i = 1] = p_i$ *for all* $1 \leq i \leq n$ *the following inequalities hold*

$$\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/3} \tag{A.1}$$

$$\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/2} \tag{A.2}$$

$$\Pr[|X - \mathbb{E}[X]| \geq \delta\mathbb{E}[X]] \leq 2e^{-\mathbb{E}[X]\delta^2/3}, \tag{A.3}$$

where $0 \leq \delta \leq 1$.

A.2 The Proof of Lemma 6.4 Under the Simplifying Assumptions

We give a proof of Lemma 6.4 where for the sake of simplicity we make the following assumptions

$$\Pr_{\substack{\pi,\rho \\ x:=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{trans}}}[D_2(x,\rho) \neq \perp] = 1 \tag{A.4}$$

$$\forall \pi \in \{0,1\}^n : S_{\pi,b} \leq (1 - \frac{1}{k}). \tag{A.5}$$

Loosely speaking, in (A.4) we assume that for every π circuit D outputs an answer, and in (A.5) we make the assumption that all surpluses are low. In the complete proof of Lemma 6.4 this assumptions fail only slightly such that it is still possible to obtain the desired result. However, the calculations are fairly lengthy. The following simplified calculations are intended to give the Reader better intuition about the full proof. We have

$$\begin{aligned}
\Pr_{\substack{\pi,\rho \\ x:=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x,\rho)) = 1] &\overset{(A.4)}{=} \Pr_{\substack{\pi,\rho \\ x:=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{trans} \\ (\Gamma_V,\Gamma_H):=\langle P^{(1)}(\pi),D_1(\rho)\rangle_{P^{(1)}}}}[\Gamma_V(D_2(x^*,\rho)) = 1 \mid D_2(x,\rho) \neq \perp] \\
&\overset{(*)}{=} \Pr_{\pi^{(k)}}[c_1 = 1 \mid c \in \mathcal{G}_1 \setminus \mathcal{G}_0] \\
&= \frac{\Pr_{\pi^{(k)}}[c_1 = 1 \wedge c \in \mathcal{G}_1 \setminus \mathcal{G}_0]}{\Pr_{\pi^{(k)}}[c \in \mathcal{G}_1 \setminus \mathcal{G}_0]} \\
&\overset{(6.15)}{=} \mathop{\mathbb{E}}_{\pi^*}\left[\frac{\Pr_{\pi^{(k-1)}}[c_1^* = 1 \wedge c \in \mathcal{G}_1 \setminus \mathcal{G}_0]\big(\Pr_{\pi^{(k-1)}}[c \in \mathcal{G}_0 \setminus \mathcal{G}_1] - (S_{\pi^*,1} - S_{\pi^*,0})\big)}{\Pr_{\pi^{(k-1)}}[c \in \mathcal{G}_0 \setminus \mathcal{G}_1]\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]}\right] \\
&\geq \mathop{\mathbb{E}}_{\pi^*}\left[\frac{\Pr_{\pi^{(k-1)}}[c_1^* = 1 \wedge c \in \mathcal{G}_1 \setminus \mathcal{G}_0]\big(\Pr_{\pi^{(k-1)}}[c \in \mathcal{G}_0 \setminus \mathcal{G}_1] - (1 - \frac{1}{k})\varepsilon\big)}{\Pr_{\pi^{(k-1)}}[c \in \mathcal{G}_0 \setminus \mathcal{G}_1]\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]}\right] \\
&\geq \frac{\Pr_{\pi^{(k)}}[c_1 = 1 \wedge c \in \mathcal{G}_1 \setminus \mathcal{G}_0] - (1 - \frac{1}{k})\varepsilon}{\Pr_{u \leftarrow \mu_\delta^k}[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]} \\
&\overset{(6.28)}{\geq} \frac{\delta \Pr[u \in \mathcal{G}_1 \setminus \mathcal{G}_0] + \varepsilon - (1 - \frac{1}{k})\varepsilon}{\Pr[u \in \mathcal{G}_1 \setminus \mathcal{G}_0]} \\
&\geq \delta + \frac{\varepsilon}{k},
\end{aligned}$$

where in $(*)$ we use the facts that $D_2(x,\rho) \neq \perp$ if and only if D_2 finds $c \in \mathcal{G}_1 \setminus \mathcal{G}_0$ and conditioned on $D_2(x,\rho) \neq \perp$ we have that $\Gamma_V(D_2(x,r)) = 1$ if and only if $c_1 = 1$.

Bibliography

[BIN97] Mihir Bellare, Russell Impagliazzo, and Moni Naor. Does parallel repetition lower the error in computationally sound protocols? In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 374–383. IEEE, 1997.

[CHS04] Ran Canetti, Shai Halevi, and Michael Steiner. Hardness amplification of weakly verifiable puzzles. 2004.

[CW77] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions (extended abstract). In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, STOC '77, pages 106–112, New York, NY, USA, 1977. ACM.

[DIJK09] Yevgeniy Dodis, Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Security amplification for interactive cryptographic primitives. In *Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography*, TCC '09, pages 128–145, Berlin, Heidelberg, 2009. Springer-Verlag.

[Gol00] Oded Goldreich. *Foundations of Cryptography: Basic Tools*. Cambridge University Press, New York, NY, USA, 2000.

[Gol04] Oded Goldreich. *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press, New York, NY, USA, 2004.

[Hol13a] Thomas Holenstein. Lecture notes in complexity theoretic cryptography, Spring 2013.

[Hol13b] Thomas Holenstein. Lecture notes in complexity theory, Spring 2013.

[HS10] Thomas Holenstein and Grant Schoenebeck. General hardness amplification of predicates and puzzles. *CoRR*, abs/1002.3534, 2010.

[IJK07] Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Chernoff-type direct product theorems. In *Advances in Cryptology-CRYPTO 2007*, pages 500–516. Springer, 2007.

[Mau13] Ueli Maurer. Lecture notes in cryptography, Spring 2013.

[VABHL03] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *Advances in Cryptology—EUROCRYPT 2003*, pages 294–311. Springer, 2003.