# Merkle's Key Agreement Protocol is Optimal: An $O(n^2)$ Attack on any Key Agreement from Random Oracles

Boaz Barak* Mohammad Mahmoody†

**Abstract**

We prove that every key agreement protocol in the random oracle model in which the honest users make at most $n$ queries to the oracle can be broken by an adversary who makes $O(n^2)$ queries to the oracle. This improves on the previous $\tilde{\Omega}(n^6)$ query attack given by Impagliazzo and Rudich (STOC '89) and resolves an open question posed by them.

Our bound is optimal up to a constant factor since Merkle proposed a key agreement protocol in 1974 that can be easily implemented with $n$ queries to a random oracle and cannot be broken by any adversary who asks $o(n^2)$ queries.

**Keywords:** Key Agreement, Random Oracle, Merkle Puzzles.

*Microsoft Research New England and Harvard University, b@boazbarak.org.

†University of Virginia, mohammad@cs.virginia.edu. Supported by NSF CAREER award CCF-1350939.

# Contents

# 1 Introduction

In the 1970's Diffie, Hellman, and Merkle [Mer74, DH76, Mer78] began to challenge the accepted wisdom that two parties cannot communicate confidentially over an open channel without first exchanging a secret key using some secure means. The first such protocol (at least in the open scientific community) was proposed by Merkle in 1974 [Mer74] for a course project in Berkeley. Even though the course's instructor rejected the proposal, Merkle continued working on his ideas and discussing them with Diffie and Hellman, leading to the papers [DH76,Mer78]. Merkle's original key exchange protocol was extremely simple and can be directly formalized and implemented using a random oracle[1] as follows:

**Protocol 1.1** (Merkle's 1974 Protocol using Random Oracles)**.** Let $n$ be the security parameter and $H : [n^2] \mapsto \{0,1\}^n$ be a function chosen at random accessible to all parties as an oracle. Alice and Bob execute the protocol as follows.

1. Alice chooses $10n$ distinct random numbers $x_1, \ldots, x_{10n}$ from $[n^2]$ and sends $a_1, \ldots, a_{10n}$ to Bob where $a_i = H(x_i)$.

2. Similarly, Bob chooses $10n$ random numbers $y_1, \ldots, y_{10n}$ in $[n^2]$ and sends $b_1, \ldots, b_{10n}$ to Alice where $b_j = H(y_j)$. (This step can be executed in parallel with Alice's first step.)

3. If there exists any $a_i = b_i$ among the exchanged strings, Alice and Bob let $(i, j)$ to be the lexicographically first index of such pair; Alice takes $x_i$ as her key and Bob takes $y_j$ as his key. If no such $(i, j)$ pair exits, they both take 0 as the agreed key.

It is easy to see that with probability at least $1 - n^4/2^n$, the random function $H : [n^2] \mapsto \{0,1\}^n$ is injective, and so any $a_i = b_i$ will lead to the same key $x_i = y_j$ used by Alice and Bob. In addition, the probability of *not* finding a "collision" $a_i = b_j$ is at most $(1 - 10/n)^{10n} \leq (1/e)^{100} < 2^{-100}$ for all $n \geq 10$. Moreover, when there is a collision $a_i = b_j$, Eve has to essentially search the whole input space $[n^2]$ to find the preimage $x_i = y_j$ of $a_i = b_j$ (or, more precisely, make $n^2/2$ calls to $H(\cdot)$ on average).

We note that in his 1978 paper [Mer78] Merkle described a different variant of a key agreement protocol by having Alice send to Bob $n$ "puzzles" $a_1, \ldots, a_n$ such that each puzzle $a_i$ takes $\approx n$ "time" to solve (where the times is modeled as the number of oracle queries), and the solver learns some secret $x_i$. The idea is that Bob would choose at random which puzzle $i \in [n]$ to solve, and so spend $\approx n$ time to learn $x_i$ which he can then use as a shared secret with Alice after sending a hash of $x_i$ to Alice so that she knows which secret Bob chose. On the other hand, Eve would need to solve almost all the puzzles to find the secret, thus spending $\approx n^2$ time. These puzzles can indeed be implemented via a random oracle $H\colon [n] \times [n] \mapsto \{0,1\}^n \times \{0,1\}^m$ as follows. The $i$'th puzzle with hidden secret $x \in \{0,1\}^m$ can be obtained by choosing and $k \leftarrow [n]$ at random and getting $a_i = (H_1(i,k), H_2(i,k) \oplus x)$ where $\oplus$ denotes bitwise exclusive OR, $H_1(\cdot,\cdot)$ denotes the first $n$ bits of $H$'s output, and $H_2(\cdot,\cdot)$ denotes the last $m$ bits of $H$'s output. Now, given puzzles $P_1 = (h_1^1, h_2^2), \ldots, P_n = (h_1^n, h_2^n)$, Bob takes a random puzzle $P_j$, solves it by asking $H(j,k)$ for all $k \in [n]$ to get $H(j,k) = (h_1^j, h_2)$ for some $h_2$, and then he retrieves the puzzle solution $x = h_2 \oplus h_2^j$.

[1] In this work, *random oracles* denote any randomized oracle $O\colon \{0,1\}^* \mapsto \{0,1\}^*$ such that $O(x)$ is independent of $O(\{0,1\}^* \setminus \{x\})$ for every $x$ (see Definition 2.2). The two protocols of Merkle we describe here can be implemented using a length-preserving random oracle (by cutting the inputs and the output to the right length). Our *negative* results, on the other hand, apply to any random oracle.

One problem with Merkle's protocol is that its security was only analyzed in the random oracle model which does not necessarily capture security when instantiated with a cryptographic one-way or hash function [CGH04]. Biham, Goren, and Ishai [BGI08] took a step towards resolving this issue by providing a security analysis for Merkle's protocol under concrete complexity assumptions. In particular, they proved that assuming the existence of one-way functions that cannot be inverted with probability more than $2^{-\alpha n}$ by adversaries running in time $2^{\alpha n}$ for $\alpha \geq 1/2 - \delta$, there is a key agreement protocol in which Alice and Bob run in time $n$ but any adversary whose running time is at most $n^{2-10\delta}$ has $o(1)$ chance of finding the secret.

Perhaps a more serious issue with Merkle's protocol is that it only provides a *quadratic* gap between the running time of the honest parties and the adversary. Fortunately, not too long after Merkle's work, Diffie and Hellman [DH76] and later Rivest, Shamir, and Adleman [RSA78] gave constructions for key agreement protocols that are conjectured to have *super-polynomial* (even subexponential) security and are of course widely used to this day. But because these and later protocols are based on certain algebraic computational problems, they could perhaps be vulnerable to unforseen attacks using this algebraic structure. It remained, however, an important open question to show whether there exist key agreement protocols with super-polynomial security that use only a random oracle.[2] The seminal paper of Impagliazzo and Rudich [IR89] answered this question negatively by showing that every key agreement protocol, even in its full general form that is allowed to run in polynomially many rounds, can be broken by an adversary asking $O(n^6 \log n)$ queries if the two parties ask $n$ queries in the random oracle model.[3] A random oracle is in particular a one-way function (with high probability)[4], and thus an important corollary of [IR89]'s result is that there is no construction of key agreement protocols based on one-way functions with a proof of super-polynomial security that is of the standard black-box type (i.e., the implementation of the protocol uses the one-way function as an oracle, and its proof of security uses the one-way function and any adversary breaking the protocol also as oracles).[5]

**Question and Motivation.** Impagliazzo and Rudich [IR89, Section 8] mention as an open question (which they attribute to Merkle) to find out whether their attack can be improved to $O(n^2)$ queries (hence showing the optimality of Merkle's protocol in the random oracle model) or there exist key agreement protocols in the random oracle model with $\omega(n^2)$ security. Beyond just being a natural question, it also has some practical and theoretical motivations. The practical motivation is that protocols with sufficiently large polynomial gap could be secure enough in practice—e.g., a key agreement protocol taking $10^9$ operations to run and $(10^9)^6 = 10^{54}$ operations to break could be good enough for many applications.[6] In fact, as was argued by Merkle himself [Mer74], as

[2] This is not to be confused with some more recent works such as [BR93], that combine the random oracle model with assumptions on the intractability of other problems such as factoring or the RSA problem to obtain more efficient cryptographic constructions.

[3] More accurately, [IR89] gave an $O(m^6 \log m)$-query attack where $m$ is the maximum of the number of queries $n$ and the number of communication rounds, though we believe their analysis could be improved to an $O(n^6 \log n)$-query attack. For the sake of simplicity, when discussing [IR89]'s results we will assume that $m = n$, though for our result we do not need this assumption.

[4] The proof of this statement for the case of non-uniform adversaries is quite nontrivial; see [GGKT05] for a proof.

[5] This argument applies to our result as well, and of course extends to any other primitive that is implied by random oracles (e.g., collision-resistant hash functions) in a black-box way.

[6] These numbers are just an example, and in practical applications the constant terms will make an important difference; however we note that these particular constants are not ruled out by [IR89]'s attack but are ruled out by ours by taking number of operations to mean the number of calls to the oracle.

technology improves and honest users can afford to run more operations, such polynomial gaps only become more useful since the ratio between the work required by the attacker and the honest user will grow as well. Thus, if known algebraic key agreement protocols were broken, one might look to polynomial-security protocol such as Merkle's for an alternative. Another motivation is theoretical— Merkle's protocol has very limited interaction (consisting of one round in which both parties simultaneously broadcast a message) and in particular it implies a public key encryption scheme. It is natural to ask whether more interaction can help achieve some polynomial advantage over this simple protocol. Brakerski et al. [BKSY11] show a simple $O(n^2)$-query attack for protocols with *perfect completeness* based on a random oracles,[7] where the probability is over both the oracle and parties' random seeds. In this work we focus on the main question of [IR89] in full fledged form.

### 1.1 Our Results

In this work we answer the above question of [IR89], by showing that every protocol in the random oracle model where Alice and Bob make $n$ oracle queries can be broken with high probability by an adversary making $O(n^2)$ queries. That is, we prove the following:

**Theorem 1.2** (Main Theorem)**.** *Let $\Pi$ be a two-party protocol in the random oracle model such that when executing $\Pi$ the two parties Alice and Bob make at most $n$ queries each, and their outputs are identical with probability at least $\rho$. Then, for every $0 < \delta < \rho$, there is an eavesdropping adversary Eve making $O(n^2/\delta^2)$ queries to the oracle whose output agrees with Bob's output with probability at least $\rho - \delta$.*

To the best of our knowledge, no better bound than the $\widetilde{O}(n^6)$-query attack of [IR89] was previously known even in the case where one does not assume the one-way function is a random oracle (which would have made the task of proving a negative result easier).

In the original publication of this work [BMG09], the following technical result (Theorem 1.3) was implicit in the proof of Theorem 1.2. Since this particular result has found uses in subsequent works to the original publication of this work [BMG09], here we state and prove it explicitly. This theorem, roughly speaking, asserts that by running the attacker of Theorem 1.2 the "correlation" between the "views" of Alice and Bob (conditioned on Eve's knowledge) remains close to zero at all times. The view of a party consists of the information they posses at any moment during the execution of the protocol: their private randomness, the public messages, and their private interaction with the oracle.

**Theorem 1.3** (Making Views almost Independent—Informal)**.** *Let $\Pi$ be a two-party protocol in the random oracle model such that when executing $\Pi$ the two parties Alice and Bob make at most $n$ oracle queries each. Then for any $\alpha, \beta < 1/10$ there is an eavesdropper Eve making $\mathrm{poly}(n/(\alpha\beta))$ queries to the oracle such that with probability at least $1-\alpha$ the following holds at the end of* every *round: the joint distribution of Alice's and Bob's views so far conditioned on Eve's view is $\beta$-close to being independent of each other.*

See Section 4 for the formal statement and proof of Theorem 1.3.

[7] We are not aware of any perfectly complete $n$-query key agreement protocol in the random oracle with $\omega(n)$ security. In other words, it seems conceivable that all such protocols could be broken with a *linear* number of queries.

## 1.2 Related Work

**Quantum-Resilient Key Agreement.** In one central scenario in which some algebraic key agreement protocols will be broken—the construction of practical quantum computers— Merkle's protocol will also be broken with linear oracle queries using Grover's search algorithm [Gro96]. In the original publication of this work we asked whether our $O(n^2)$-query classical attack could lead to an $O(n)$ quantum attack against any classical protocol (where Eve accesses the random oracle in a superposition). We note that using quantum communication there is an *information theoretically* secure key agreement protocol [BBE92]. Brassard and Salvail [BS08] (independently observed by [BGI08]) gave a quantum version of Merkle's protocol, showing that Alice and Bob can use quantum computation (but classical communication) to obtain a key agreement protocol with super-linear $n^{3/2}$ security in the random oracle model against quantum adversaries. Finally, Brassard et al. [BHK+11] resolved our question negatively by presenting a *classical* protocol in the random oracle model with super linear security $\Omega(n^{3/2-\varepsilon})$ for arbitrary small constant $\varepsilon$.

**Attacks in Small Parallel Time.** Mahmoody, Moran, and Vadhan [MMV11] showed how to improve the *round complexity* of the attacker of Theorem 1.2 to $n$ (which is optimal) for the case of one-message protocols, where a round here refers to a set of queries that are asked to the oracle in parallel.[8] Their result rules out constructions of "time-lock puzzles" in the *parallel* random oracle model in which the polynomial-query solver needs more *parallel time* (i.e., rounds of parallel queries to the random oracle) than the puzzle generator to solve the puzzle. As an application back to our setting, [MMV11] used the above result and showed that every $n$-query (even multi-round) key agreement protocol can be broken by $O(n^3)$ queries in only $n$ rounds of oracle queries, improving the $\Omega(n^2)$-round attack of our work by a factor of $n$. Whether an $O(n)$-round $O(n^2)$-query attack is possible remains as an intriguing open question.

**Black-Box Separations and the Power of Random Oracle.** The work of Impagliazzo and Rudich [IR89] laid down the framework for the field of *black-box separations*. A black-box separation of a primitive $\mathcal{Q}$ from another primitive $\mathcal{P}$ rules out any construction of $\mathcal{Q}$ from $\mathcal{P}$ as long as it treats the primitive $\mathcal{P}$ and the adversary (in the security proof) as oracles. We refer the reader to the excellent survey by Reingold et al. [RTV04] for the formal definition and its variants. Due to the abundance of black-box techniques in cryptography, a black-box separation indicates a major disparity between how hard it is to achieve $\mathcal{P}$ vs. $\mathcal{Q}$, at least with respect to black-box techniques. The work of [IR89] employed the so called "oracle separation" method to derive their black-box separation. In particular, they showed that relative to the oracle $O = (R, \mathbf{PSPACE})$ in which $R$ is a random oracle one-way functions exist (with high probability) but secure key agreement does not. This existence of such an oracle implies a black-box separation.

The main technical step in the proof of [IR89] is to show that relative to a random oracle $R$, any key agreement protocol could be broken by an adversary who is computationally unbounded and asks at most $S = \text{poly}(n)$ number of queries (where $n$ is the security parameter). The smallest such polynomial $S$ for any construction $\mathcal{C}$ could be considered as a quantitative black-box security for $\mathcal{C}$ in the random oracle model. This is indeed the setting of our paper, and we study the optimal black-box security of key agreement in the random oracle model. Our Theorem 1.2 proves that

[8] For example, a *non-adaptive* attacker who prepares all of its oracle queries and then asks them in one shot, has round complexity one.

$\Theta(n^2)$ is the optimal security one can achieve for an $n$-query key agreement protocol in the random oracle model. The techniques used in the proof of Theorem 1.2 have found applications in the contexts of black-box separations and black-box security in the random oracle model (see, e.g., [KSY11, BKSY11, MP12]). In the following we describe some of the works that focus on the power of random oracles in secure two-party computation.

Dachman-Soled et al. [DSLMM11] were the first to point out that results implicit in our proof of Theorem 1.2 in the original publication of this work [BMG09] could be used to show the existence of eavesdropping attacks that gather enough information from the oracle in a way that conditioned on this information the views of Alice and Bob become "close" to being independent (see Lemma 5 of [DSLMM11]). Such results were used in [DSLMM11], [MMP14], and [HHRS07] to explore the power of random oracles in secure two-party computation. Dachman-Soled et al. showed that "optimally-fair" coin tossing protocols [Cle86] cannot be based on one-way functions with $n$ input and $n$ output bits in a black-box way if the protocol has $o(n/\log n)$ rounds.

Mahmoody, Maji, and Prabhakaran [MMP14] proved that random oracles are useful for secure two-party computation of finite (or at most polynomial-size domain) deterministic functions only as the commitment functionality. Their results showed that "non-trivial" functions can not be computed securely by a black-box use of one-way functions.

Haitner, Omri, and Zarosim [HOZ13] studied input-less randomized functionalities and showed that a random oracle[9] is, to a large extent, useless for such functionalities as well. In particular, it was shown that for every protocol $\Pi$ in the random oracle model, and every polynomial $p(\cdot)$, there is a protocol in the no-oracle model that is "$1/p(\cdot)$-close" to $\Pi$. [HOZ13] proved this result by using the machinery developed in the original publication of this work (e.g., the *graph characterization* of Section 3.3.2) and simplified some of the steps of the original proof. [HOZ13] showed how to use such lower-bounds for the input-less setting to prove black-box separations from one-way functions for "differentially private" two-party functionalities for the *with-input* setting.

### 1.3 Our Techniques

The main technical challenge in proving our main result is the issue of *dependence* between the executions of the two parties Alice and Bob in a key agreement protocol. At first sight, it may seem that a computationally unbounded attacker that monitors all communication between Alice and Bob will trivially be able to find out their shared key. But the presence of the random oracle allows Alice and Bob to correlate their executions even without communicating (which is indeed the reason that Merkle's protocol achieves nontrivial security). Dealing with such correlations is the cause of the technical complexity in both our work and the previous work of Impagliazzo and Rudich [IR89]. We handle this issue in a different way than [IR89]. On a very high level our approach can be viewed as using more information about the structure of these correlations than [IR89] did. This allows us to analyze a more efficient attacking algorithm that is more frugal with the number of queries it uses than the attacker of [IR89]. Below we provide a more detailed (though still high level) exposition of our technique and its relation to [IR89]'s technique.

We now review [IR89]'s attack (and its analysis) and particularly discuss the subtle issue of *dependence* between Alice and Bob that arises in both their work and ours. However, no result of this section is used in the later sections, and so the reader should feel free at any time to skip ahead to the next sections that contain our actual attack and its analysis.

[9] [HOZ13] proved this result for a larger class of oracles, see [HOZ13] for more details.

### 1.3.1 The Approach of [IR89]

Consider a protocol that consists of $n$ rounds of interaction, where each party makes exactly one oracle query before sending its message. [IR89] called protocols of this type "normal-form protocols" and gave an $\widetilde{O}(n^3)$ attack against them (their final result was obtained by transforming every protocol into a normal-form protocol with a quadratic loss of efficiency). Even though without loss of generality the attacker Eve of a key agreement protocol can defer all of her computation till after the interaction between Alice and Bob is finished, it is conceptually simpler in both [IR89]'s case and ours to think of the attacker Eve as running concurrently with Alice and Bob. In particular, the attacker Eve of [IR89] performed the following operations after each round $i$ of the protocol:

- If the round $i$ is one in which Bob sent a message, then at this point Eve samples $1000n \log n$ random executions of Bob from the distribution $\mathcal{D}$ of Bob's executions that are consistent with the information that Eve has at that moment (which consists of the communication transcript and previous oracle answers). That is, Eve samples a uniformly random tape for Bob and uniformly random query answers subject to being consistent with Eve's information. After each time she samples an execution, Eve asks the oracle all the queries that are asked during this execution and records the answers. (Generally, the true answers will be different from Eve's guessed answers when sampling the execution.) If the round $i$ is one in which Alice sent a message, then Eve does similarly by changing the role of Alice and Bob.

Overall Eve will sample $\widetilde{O}(n^2)$ executions making a total of $\widetilde{O}(n^3)$ queries. It's not hard to see that as long as Eve learns all of the *intersection queries* (queries asked by both Alice and Bob during the execution) then she can recover the shared secret with high probability. Thus the bulk of [IR89]'s analysis was devoted to showing the following claim.

**Claim 1.4.** *With probability at least* 0.9 *Eve never fails, where we say that Eve* fails *at round $i$ if the query made in this round by, say, Alice was asked previously by Bob but not by Eve.*

At first look, it may seem that one could easily prove Claim 1.4. Indeed, Claim 1.4 will follow by showing that at any round $i$, the probability that Eve fails in round $i$ *for the first time* is at most $1/(10n)$. Now all the communication between Alice and Bob is observed by Eve, and if no failure has yet happened then Eve has also observed all the intersection queries so far. Because the answers for non-intersection queries are completely random and independent from one another it seems that Alice has no more information about Bob than Eve does, and hence if the probability that Alice's query $q$ was asked before by Bob is more than $1/(10n)$ then this query $q$ has probability at least $1/(10n)$ to appear in each one of Eve's sampled executions of Bob. Since Eve makes $1000n \log n$ such samples, the probability that Eve misses $q$ would be bounded by $(1-\frac{1}{10n})^{1000n \log n} \ll 1/(10n)$.

**The Dependency Issue.** When trying to turn the above intuition into a proof, the assumption that Eve has as much information about Bob as Alice does translates to the following statement: conditioned on Eve's information, the distributions of Alice's view and Bob's view are *independent* from one another.[10] Indeed, if this statement were true then the above paragraph could have been easily translated into a proof that [IR89]'s attacker is successful, and it wouldn't have been hard to

[10] Readers familiar with the setting of communication complexity may note that this is analogous to the well known fact that conditioning on any transcript of a 2-party communication protocol results in a product distribution (i.e., combinatorial rectangle) over the inputs. However, things are different in the presence of a random oracle.

optimize this attacker to achieve $O(n^2)$ queries. Alas, this statement is false. Intuitively the reason is the following: even the fact that Eve has not missed any intersection queries is some nontrivial information that Alice and Bob share and creates dependence between them.[11]

Impagliazzo and Rudich [IR89] dealt with this issue by a "charging argument", where they showed that such dependence can be charged in a certain way to one of the executions sampled by Eve, in a way that at most $n$ samples can be charged at each round (and the rest of Eve's samples are distributed correctly as if the independence assumption was true). This argument inherently required sampling at least $n$ executions (each of $n$ queries) per round, resulting in an $\Omega(n^3)$ attack.

### 1.3.2 Our Approach

We now describe our approach and how it differs from the previous proof of [IR89]. The discussion below is somewhat high level and vague, and glosses over some important details. Again, the reader is welcome to skip ahead at any time to Section 3 that contains the full description of our attack and does not depend on this section in any way. Our attacking algorithm follows the same general outline as that of [IR89] but has two important differences:

1. One *quantitative* difference is that while our attacker Eve also computes a distribution $\mathcal{D}$ of possible executions of Alice and Bob conditioned on her knowledge, she does *not* sample full executions from $\mathcal{D}$; rather, she computes whether there is any query $q \in \{0,1\}^*$ that has probability more than, say, $1/(100n)$ of being in $\mathcal{D}$ and makes only such *heavy* queries.

   Intuitively, since Alice and Bob make at most $2n$ queries, the total expected number of heavy queries (and hence the query complexity of Eve) is bounded by $O(n^2)$. The actual analysis is more involved since the distribution $\mathcal{D}$ keeps changing as Eve learns more information through the messages she observes and oracle answers she receives.

2. The *qualitative* difference is that here we do not consider the same distribution $\mathcal{D}$ that was considered by [IR89]. Their attacker to some extent "pretended" that the conditional distributions of Alice and Bob are independent from one another and only considered one party in each round. In contrast, we define our distribution $\mathcal{D}$ to be the *joint* distribution of Alice and Bob, where there could be dependencies between them. Thus, to sample from our distribution $\mathcal{D}$ one would need to sample a *pair* of executions of Alice and Bob (random tapes and oracle answers) that are *consistent* with one another and Eve's current knowledge.

The main challenge in the analysis is to prove that the attack is *successful* (i.e., that Claim 1.4 above holds) and in particular that the probability of failure at each round (or more generally, at each query of Alice or Bob) is bounded by, say, $1/(10n)$. Once again, things would have been easy if we knew that the distribution $\mathcal{D}$ of the possible executions of Alice and Bob conditioned on Eve's knowledge is a *product distribution*, and hence Alice has no more information on Bob than Eve has. While this is not generally true, we show that in our attack this distribution is *close to being a product distribution*, in a precise sense.

[11] As a simple example for such dependence consider a protocol where in the first round Alice chooses $x$ (which is going to be the shared key) to be either the string $0^n$ or $1^n$ at random, queries the oracle $H$ at $x$ and sends $y = H(x)$ to Bob. Bob then makes the query $1^n$ and gets $y' = H(1^n)$. Now even if Alice chose $x = 0^n$ and hence Alice and Bob have no intersection queries, Bob can find out the value of $x$ just by observing that $y' \neq y$. Still, an attacker must ask a non-intersection query such as $1^n$ to know if $x = 0^n$ or $x = 1^n$.

At any point in the execution, fix Eve's current information about the system and define a bipartite graph $G$ whose left-side vertices correspond to possible executions of Alice that are consistent with Eve's information and right-side vertices correspond to possible executions of Bob consistent with Eve's information. We put an edge between two executions $A$ and $B$ if they are consistent with one another and moreover if they do not represent an execution in which Eve has already *failed* (i.e., there is no intersection query that is asked in both executions $A$ and $B$ but not by Eve). Roughly speaking, the distribution $\mathcal{D}$ that our attacker Eve considers can be thought of as choosing a uniformly random edge in the graph $G$. (Note that the graph $G$ and the distribution $\mathcal{D}$ change at each point that Eve learns some new information about the system.) If $G$ were the complete bipartite clique then $\mathcal{D}$ would have been a product distribution. Although $G$ can rarely be the complete graph, what we show is that $G$ is still *dense* in the sense that each vertex is connected to most of the vertices on the other side. Relying on the density of this graph, we show that Alice's probability of hitting a query that Bob asked before is at most twice the probability that Eve does so if she chooses the most likely query based on her knowledge.

The bound on the degree is obtained by showing that $G$ can be represented as a *disjointness graph*, where each vertex $u$ is associated with a set $S(u)$ (from an arbitrarily large universe) and there is an edge between a left-side vertex $u$ and a right-side vertex $v$ if and only if $S(u) \cap S(v) = \varnothing$. The set $S(u)$ corresponds to the queries made in the execution corresponding to $u$ that are *not* asked by Eve. The definition of the graph $G$ implies that $|S(u)| \leq n$ for all vertices $u$. The definition of our attacking algorithm implies that the distribution obtained by picking a random edge $e = (u, v)$ and outputting $S(u) \cup S(v)$ is *light* in the sense that there is no element $q$ in the universe that has probability more than $1/(10n)$ of being in a set chosen from this distribution. We show that these conditions imply that each vertex is connected to most of the vertices on the other side.

## 2 Preliminaries

We use bold fonts to denote random variables. By $Q \leftarrow \mathbf{Q}$ we indicate that $Q$ is sampled from the distribution of the random variable $\mathbf{Q}$. By $(\mathbf{x}, \mathbf{y})$ we denote a *joint* distribution over random variables $\mathbf{x}, \mathbf{y}$. By $\mathbf{x} \equiv \mathbf{y}$ we denote that $\mathbf{x}$ and $\mathbf{y}$ are identically distributed. For jointly distributed $(\mathbf{x}, \mathbf{y})$, by $(\mathbf{x} \mid \mathbf{y} = y)$ we denote the distribution of $\mathbf{x}$ conditioned on $\mathbf{y} = y$. When it is clear from the context we might simply write $(\mathbf{x} \mid y)$ instead of $(\mathbf{x} \mid \mathbf{y} = y)$. By $(\mathbf{x} \times \mathbf{y})$ we denote a product distribution in which $\mathbf{x}$ and $\mathbf{y}$ are sampled independently. For a finite set $S$, by $x \leftarrow S$ we denote that $x$ is sampled from $S$ uniformly at random. By $\mathrm{Supp}(\mathbf{x})$ we denote the *support set* of the random variable $\mathbf{x}$ defined as $\mathrm{Supp}(\mathbf{x}) = \{x \mid \Pr[\mathbf{x} = x] > 0\}$. For any event $E$, by $\neg E$ we denote the complement of the event $E$.

**Definition 2.1.** A *partial function* $F$ is a function $F\colon D \mapsto \{0,1\}^*$ defined over some domain $D \subseteq \{0,1\}^*$. We call two partial functions $F_1, F_2$ with domains $D_1, D_2$ *consistent* if $F_1(x) = F_2(x)$ for every $x \in D_1 \cap D_2$. (In particular, $F_1$ and $F_2$ are consistent if $D_1 \cap D_2 = \varnothing$.)

In previous work random oracles are defined either as Boolean functions [IR89] or length-preserving functions [BR93]. In this work we use a general definition that captures both cases by only requiring the oracle answers to be independent. Since our goal is to give *attacks* in this model, using this definition makes our results more general and applicable to both scenarios.

**Definition 2.2** (Random Oracles)**.** A *random oracle* $\mathbf{H}(\cdot)$ is a random variable whose values are functions $H\colon \{0,1\}^* \mapsto \{0,1\}^*$ such that $\mathbf{H}(x)$ is distributed independently of $\mathbf{H}(\{0,1\}^* \setminus \{x\})$ for

all $x \in \{0,1\}^*$ and that $\Pr[\mathbf{H}(x) = y]$ is a rational number for every pair $(x, y)$.[12] For any finite partial function $F$, by $\Pr_{\mathbf{H}}[F]$ we denote the probability that the random oracle $\mathbf{H}$ is consistent with $F$. Namely, $\Pr_{\mathbf{H}}[F] = \Pr_{H \leftarrow \mathbf{H}}[F \subseteq H]$ and $\Pr_{\mathbf{H}}[\varnothing] = 1$ where $F \subseteq H$ means that the partial function $F$ is consistent with $H$.

**Remark 2.3** (Infinite vs. Finite Random Oracles)**.** In this work, we will always work with *finite* random oracles which are only queried on inputs of length $n \leq \text{poly}(\kappa)$ where $\kappa$ is a (security) parameter given to parties. Thus, we only need a finite variant of Definition 2.2. However, in case of infinite random oracles (as in Definition 2.2) we need a measure space over the space of full infinite oracles that is consistent with the finite probability distributions of $\mathbf{H}(\cdot)$ restricted to inputs $\{0,1\}^n$ for all $n = 1, 2, \ldots$. By Caratheodory's extension theorem, such measure space exists and is unique (see Theorem 4.6 of [Hol15]).

Since for every random oracle $\mathbf{H}(\cdot)$ and fixed $x$ the random variable $\mathbf{H}(x)$ is independent of $\mathbf{H}(x')$ for all $x' \neq x$, we can use the following characterization of $\Pr_{\mathbf{H}}[F]$ for every $F \subseteq \{0,1\}^* \times \{0,1\}^*$. Here we only state and use this lemma for finite sets.

**Proposition 2.4.** *For every random oracle $\mathbf{H}(\cdot)$ and every finite set $F \subset \{0,1\}^* \times \{0,1\}^*$ we have*

$$\Pr_{\mathbf{H}}[F] = \prod_{(x,y)\in F} \Pr[\mathbf{H}(x) = y].$$

Now we derive the following lemma from the above proposition.

**Lemma 2.5.** *For consistent finite partial functions $F_1, F_2$ and random oracle $\mathbf{H}$ it holds that*

$$\Pr_{\mathbf{H}}[F_1 \cup F_2] = \frac{\Pr_{\mathbf{H}}[F_1] \cdot \Pr_{\mathbf{H}}[F_2]}{\Pr_{\mathbf{H}}[F_1 \cap F_2]}.$$

*Proof.* Since $F_1$ and $F_2$ are consistent, we can think of $F = F_1 \cup F_2$ as a partial function. Therefore, by Proposition 2.4 and the inclusion-exclusion principle we have:

$$\begin{aligned}
\Pr_{\mathbf{H}}[F_1 \cup F_2] &= \prod_{(x,y)\in F_1 \cup F_2} \Pr[\mathbf{H}(x) = y] \\
&= \frac{\prod_{(x,y)\in F_1} \Pr[\mathbf{H}(x) = y] \cdot \prod_{(x,y)\in F_2} \Pr[\mathbf{H}(x) = y]}{\prod_{(x,y)\in F_1 \cap F_2} \Pr[\mathbf{H}(x) = y]} \\
&= \frac{\Pr_{\mathbf{H}}[F_1] \cdot \Pr_{\mathbf{H}}[F_2]}{\Pr_{\mathbf{H}}[F_1 \cap F_2]}.
\end{aligned}$$

□

**Lemma 2.6** (Lemma 6.4 in [IR89])**.** *Let $E$ be any event defined over a random variable $\mathbf{x}$, and let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be a sequence of random variables all determined by $\mathbf{x}$. Let $D$ be the event defined over $(\mathbf{x}_1, \ldots)$ that holds if and only if there exists some $i \geq 1$ such that $\Pr[E \mid x_1, \ldots, x_i] \geq \lambda$. Then $\Pr[E \mid D] \geq \lambda$.*

[12] Our results extend to the case where the probabilities are not necessarily rational numbers, however, since every reasonable candidate random oracle we are aware of satisfies this rationality condition, and it avoids some technical subtleties, we restrict attention to oracles that satisfy it. In Section 4.2 we show how to remove this restriction.

**Lemma 2.7.** *Let $E$ be any event defined over a random variable $\mathbf{x}$, and let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be a sequence of random variables all determined by $\mathbf{x}$. Suppose $\Pr[E] \leq \lambda$ and $\lambda = \lambda_1 \cdot \lambda_2$. Let $D$ be the event defined over $(\mathbf{x}_1, \ldots)$ that holds if and only if there exists some $i \geq 1$ such that $\Pr[E \mid x_1, \ldots, x_i] \geq \lambda_1$. Then it holds that $\Pr[D] \leq \lambda_2$.*

*Proof.* Lemma 2.6 shows that $\Pr[E \mid D] \geq \lambda_1$. Now we prove the contrapositive of Lemma 2.7. If $\Pr[D] > \lambda_2$, then we would get $\Pr[E] \geq \Pr[E \wedge D] \geq \Pr[D] \cdot \Pr[E \mid D] > \lambda_1 \cdot \lambda_2 = \lambda$. □

## 2.1 Statistical Distance

**Definition 2.8** (Statistical Distance)**.** By $\Delta(\mathbf{x}, \mathbf{y})$ we denote the *statistical distance* between random variables $\mathbf{x}, \mathbf{y}$ defined as $\Delta(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \cdot \sum_z |\Pr[\mathbf{x} = z] - \Pr[\mathbf{y} = z]|$. We call random variables $\mathbf{x}$ and $\mathbf{y}$ *$\varepsilon$-close*, denoted by $\mathbf{x} \approx_\varepsilon \mathbf{y}$, if $\Delta(\mathbf{x}, \mathbf{y}) \leq \varepsilon$.

We use the following useful well-known lemmas about statistical distance.

**Lemma 2.9.** *$\Delta(\mathbf{x}, \mathbf{y}) = \varepsilon$ if and only if either of the following holds:*

1. *For every (even randomized) function $D$ it holds that $\Pr[D(\mathbf{x}) = 1] - \Pr[D(\mathbf{y}) = 1] \leq \varepsilon$.*
2. *For every event $E$ it holds that $\Pr_{\mathbf{x}}[E] - \Pr_{\mathbf{y}}[E] \leq \varepsilon$.*

*Moreover, if $\Delta(\mathbf{x}, \mathbf{y}) = \varepsilon$, then there is a deterministic (detecting) Boolean function $D$ that achieves $\Pr[D(\mathbf{x}) = 1] - \Pr[D(\mathbf{y}) = 1] = \varepsilon$.*

**Lemma 2.10.** *It holds that $\Delta((\mathbf{x}, \mathbf{z}), (\mathbf{y}, \mathbf{z})) = \mathbb{E}_{z \leftarrow \mathbf{z}} \Delta((\mathbf{x} \mid z), (\mathbf{y} \mid z))$.*

**Lemma 2.11.** *If $\Delta(\mathbf{x}, \mathbf{y}) \leq \varepsilon_1$ and $\Delta(\mathbf{y}, \mathbf{z}) \leq \varepsilon_2$, then $\Delta(\mathbf{x}, \mathbf{z}) \leq \varepsilon_1 + \varepsilon_2$.*

**Lemma 2.12.** *$\Delta((\mathbf{x}_1, \mathbf{x}_2), (\mathbf{y}_1, \mathbf{y}_2)) \geq \Delta(\mathbf{x}_1, \mathbf{y}_1)$.*

We use the convention for the notation $\Delta(\cdot, \cdot)$ that whenever $\Pr[\mathbf{x} \in E] = 0$ for some event $E$, we let $\Delta((\mathbf{x} \mid E), \mathbf{y}) = 1$ for every random variable $\mathbf{y}$.

**Lemma 2.13.** *Suppose $\mathbf{x}, \mathbf{y}$ are finite random variables, and suppose $G$ is some event defined over $\mathrm{Supp}(\mathbf{x})$. Then $\Delta(\mathbf{x}, \mathbf{y}) \leq \Pr_{\mathbf{x}}[G] + \Delta((\mathbf{x} \mid \neg G), \mathbf{y})$.*

*Proof.* Let $\delta = \Delta(\mathbf{x}, \mathbf{y})$. Let $\mathbf{g}$ be a Boolean random variable jointly distributed with $\mathbf{x}$ as follows: $\mathbf{g} = 1$ if and only if $\mathbf{x} \in G$. Suppose $\mathbf{y}$ is sampled independently of $(\mathbf{x}, \mathbf{g})$ (and so $(\mathbf{y}, \mathbf{g}) \equiv (\mathbf{y} \times \mathbf{g})$). By Lemmas 2.12 and 2.10 we have:

$$\begin{aligned}
\Delta(\mathbf{x}, \mathbf{y}) &\leq \Delta((\mathbf{x}, \mathbf{g}), (\mathbf{y}, \mathbf{g})) \\
&= \mathop{\mathbb{E}}_{g \leftarrow \mathbf{g}} \Delta((\mathbf{x} \mid g), (\mathbf{y} \mid g)) \\
&= \mathop{\mathbb{E}}_{g \leftarrow \mathbf{g}} \Delta((\mathbf{x} \mid g), \mathbf{y}) \\
&= \Pr[\mathbf{g} = 1] \cdot \Delta((\mathbf{x} \mid \mathbf{g} = 1), \mathbf{y}) + \Pr[\mathbf{g} = 0] \cdot \Delta((\mathbf{x} \mid \mathbf{g} = 0), \mathbf{y}) \\
&\leq \Pr[\mathbf{g} = 1] + \Delta((\mathbf{x} \mid \mathbf{g} = 0), \mathbf{y}) \\
&= \Pr_{\mathbf{x}}[G] + \Delta((\mathbf{x} \mid \neg G), \mathbf{y}).
\end{aligned}$$

□

**Definition 2.14** (Key Agreement)**.** A key agreement protocol consists of two interactive polynomial-time probabilistic Turing machines $(A, B)$ that both get $1^n$ as security parameter, each get secret randomness $\mathbf{r}_A, \mathbf{r}_B$, and after interacting for $\text{poly}(n)$ rounds $A$ outputs $s_A$ and $B$ outputs $s_B$. We say a key agreement scheme $(A, B)$ has completeness $\rho$ if $\Pr[s_A = s_B] \geq \rho(n)$. For an arbitrary oracle $O$, we define key agreement protocols (and their completeness) relative to $O$ by simply allowing $A$ and $B$ to be efficient algorithms relative to $O$.

**Security of Key Agreement Protocols.** It can be easily seen that no key agreement protocol with completeness $\rho > 0.9$ could be *statistically* secure, and that there is always a computationally unbounded eavesdropper Eve who can guess the shared secret key $s_A = s_B$ with probability at least $1/2 + \text{neg}(n)$. In this work we are interested in statistical security of key agreement protocols in the *random oracle model*. Namely, we would like to know how many oracle queries are required to break a key agreement protocol relative to a random oracle.

# 3 Proving the Main Theorem

In this section we prove the next theorem which implies our Theorem 1.2 as special case.

**Theorem 3.1.** *Let $\Pi$ be a two-party interactive protocol between Alice and Bob using a random oracle $\mathbf{H}$ (accessible by everyone) such that:*

- *Alice uses local randomness $r_A$, makes at most $n_A$ queries to $H$ and at the end outputs $s_A$.*
- *Bob uses local randomness $r_B$, makes at most $n_B$ queries to $H$ and at the end outputs $s_B$.*
- $\Pr[s_A = s_B] \geq \rho$ *where the probability is over the choice of* $(r_A, r_B, H) \leftarrow (\mathbf{r}_A, \mathbf{r}_B, \mathbf{H})$.

*Then, for every $0 < \delta < \rho$, there is a* deterministic *eavesdropping adversary Eve who only gets access to the public sequence of messages $M$ sent between Alice and Bob, makes at most $400 \cdot n_A \cdot n_B/\delta^2$ queries to the oracle $H$ and outputs $s_E$ such that* $\Pr[s_E = s_B] \geq \rho - \delta$.

## 3.1 Notation and Definitions

In this subsection we give some definitions and notations to be used in the proof of Theorem 3.1. W.l.o.g we assume that Alice, Bob, and Eve will never ask an oracle query twice. Recall that Alice (resp. Bob) asks at most $n_A$ (resp. $n_B$) oracle queries.

**Rounds.** Alice sends her messages in odd rounds and Bob sends his messages in even rounds. Suppose $i = 2j - 1$ and it is Alice's turn to send the message $m_i$. This round starts by Alice asking her oracle queries and computing $m_i$, then Alice sends $m_i$ to Bob, and this round ends by Eve asking her (new) oracle queries based on the messages sent so far $M^i = [m_1, \ldots, m_i]$. Same holds for $i = 2j$ by changing the role of Alice and Bob.

**Queries and Views.** By $Q_A^i$ we denote the set of oracle queries asked by Alice by the end of round $i$. By $P_A^i$ we denote the set of oracle query-answer pairs known to Alice by the end of round $i$ (i.e., $P_A^i = \{(q, H(q)) \mid q \in Q_A^i\}$). By $V_A^i$ we denote the view of Alice by the end of round $i$. This view consists of: Alice's randomness $r_A$, exchanged messages $M^i$ as well as oracle query-answer pairs $P_A^i$ known to Alice so far. By $Q_B^i, P_B^i, V_B^i$ (resp. $Q_E^i, P_E^i, V_E^i$) we denote the same variables defined for Bob (resp. Eve). Note that $V_E^i$ only consists of $(M^i, P_E^i)$ since Eve does not use any randomness. We also use $\mathcal{Q}(\cdot)$ as an operator that extracts the set of queries from set of query-answer pairs or views; namely, $\mathcal{Q}(P) = \{q \mid \exists\ a, (q, a) \in P\}$ and $\mathcal{Q}(V) = \{q \mid \text{the query } q \text{ is asked in the view } V\}$.

**Definition 3.2** (Heavy Queries)**.** For a random variable $\mathbf{V}$ whose samples $V \leftarrow \mathbf{V}$ are sets of queries, sets of query-answer pairs, or views, we say a query $q$ is $\varepsilon$-*heavy* for $\mathbf{V}$ if and only if $\Pr[q \in \mathcal{Q}(\mathbf{V})] \geq \varepsilon$.

**Executions and Distributions** A (full) *execution* of Alice, Bob, and Eve can be described by a tuple $(r_A, r_B, H)$ where $r_A$ denotes Alice's random tape, $r_B$ denotes Bob's random tape, and $H$ is the random oracle (note that Eve is deterministic). We denote by $\mathcal{E}$ the distribution over (full) executions that is obtained by running the algorithms for Alice, Bob and Eve with uniformly chosen random tapes $r_A, r_B$ and a uniformly sampled random oracle $H$. By $\Pr_{\mathcal{E}}[P_A^i]$ we denote the probability that a full execution of the system leads to $\mathbf{P}_A^i = P_A^i$ for a given $P_A^i$. We use the same notation also for other components of the system (by treating their occupance as events) as well.

For a sequence of $i$ messages $M^i = [m_1, \ldots, m_i]$ exchanged between the two parties and a set of query-answer pairs (i.e., a partial function) $P$, by $\mathcal{V}(M^i, P)$ we denote the joint distribution over the views $(V_A^i, V_B^i)$ of Alice and Bob in their own (partial) executions up to the point in the system in which the $i$'th message is sent (by Alice or Bob) conditioned on: the transcript of messages in the first $i$ rounds being equal to $M^i$ and $H(q) = a$ for all $(q, a) \in P$. Looking ahead in the proof, the distribution $\mathcal{V}(M^i, P)$ would be the conditional distribution of Alice's and Bob's views in eyes of the attacker Eve who knows the public messages and has learned oracle query-answer pairs described in $P$. For $(M^i, P)$ such that $\Pr_{\mathcal{E}}[M^i, P] > 0$, the distribution $\mathcal{V}(M^i, P)$ can be sampled by first sampling $(r_A, r_B, H)$ uniformly at random conditioned on being consistent with $(M^i, P)$ and then deriving Alice's and Bob's views $V_A^i, V_B^i$ from the sampled $(r_A, r_B, H)$.

For $(M^i, P)$ such that $\Pr_{\mathcal{E}}[M^i, P] > 0$, the event $\mathsf{Good}(M^i, P)$ is defined over the distribution $\mathcal{V}(M^i, P)$ and holds if and only if $Q_A^i \cap Q_B^i \subseteq \mathcal{Q}(P)$ for $Q_A^i, Q_B^i, \mathcal{Q}(P)$ determined by the sampled views $(V_A^i, V_B^i) \leftarrow \mathcal{V}(M^i, P)$ and $P$. For $\Pr_{\mathcal{E}}[M^i, P] > 0$ we define the distribution $\mathcal{GV}(M^i, P)$ to be the distribution $\mathcal{V}(M^i, P)$ conditioned on $\mathsf{Good}(M^i, P)$. Looking ahead to the proof the event $\mathsf{Good}(M^i, P)$ indicates that the attacker Eve has not "missed" any query that is asked by both of Alice and Bob (i.e. an intersection query) so far, and thus $\mathcal{GV}(M^i, P)$ refer to the same distribution of $\mathcal{V}(M^i, P)$ with the extra condition that so far no intersection query is missed by Eve.

### 3.2 Attacker's Algorithm

In this subsection we describe an attacker Eve who might ask $\omega(n_A n_B/\delta^2)$ queries, but she finds the key in the two-party key agreement protocol between Alice and Bob with probability $1 - O(\delta)$. Then we show how to make Eve "efficient" without decreasing the success probability too much.

**Protocols in Seminormal Form.** We say a protocol is in *seminormal form*[13] if **(1)** the number of oracle queries asked by Alice or Bob in each round is at most one, and **(2)** when the last message is sent (by Alice or Bob) the other party does not ask any oracle queries and computes its output without using the last message. The second property could be obtained by simply adding an extra message LAST at the end of the protocol. (Note that our results do not depend on the number of rounds.) One can also always achieve the first property *without* compromising the security as follows. If the protocol has $2 \cdot \ell$ rounds, we will increase the number of rounds to $2\ell \cdot (n_A + n_B - 1)$ as follows. Suppose it is Alice's turn to send $m_i$ and before doing so she needs to ask the queries $q_1, \ldots, q_k$ (perhaps adaptively) from the oracle. Instead of asking these queries from $H(\cdot)$ and sending $m_i$ in one round, Alice and Bob will run $2n_A - 1$ *sub-rounds* of interaction so that Alice will have enough number of (fake) rounds to ask her queries from $H(\cdot)$ one by one. More formally:

1. The messages of the first $2n_A - 1$ sub-rounds for an odd round $i$ will all be equal to $\perp$. Alice sends the first $\perp$ message, and the last message will be $m_i$ sent by Alice.

2. For $j \leq k$, before sending the message of the $2j - 1$'th sub-round Alice asks $q_j$ from the oracle. The number of these queries, namely $k$, might not be known to Alice at the beginning of round $i$, but since $k \leq n_A$, the number of sub-rounds are enough to let Alice ask all of her queries $q_1, \ldots, q_k$ without asking more than one query in each sub-round.

If a protocol is in semi-normal form, then in each round there is at most one query asked by the party who sends the message of that round, and we will use this condition in our analysis. Moreover, Eve can simply *pretend* that *any* protocol is in seminormal form by *imagining* in her head that the extra $\perp$ messages are being sent between every two real message. Therefore, w.l.o.g in the following we will assume that the two-party protocol $\Pi$ has $\ell$ rounds and is in seminormal form.[14] Finally note that we cannot simply "expand" a round $i$ in which Alice asks $k_i$ queries into $2k$ messages between Alice and Bob, because then Bob would know how many queries were asked by Alice, but if we do the transformation as described above, then the actual number of queries asked for that round could potentially remain secret.

**Construction 3.3.** Let $\varepsilon < 1/10$ be an input parameter. The adversary Eve attacks the $\ell$-round two-party protocol $\Pi$ between Alice and Bob (which is in seminormal form) as follows. During the attack Eve updates a set $P_E$ of oracle query-answer pairs as follows. Suppose in round $i$ Alice or Bob sends the message $m_i$. After $m_i$ is sent, if $\Pr_{\mathcal{E}}[\mathsf{Good}(M^i, P_E)] = 0$ holds at any moment, then Eve aborts. Otherwise, as long as there is any query $q \notin \mathcal{Q}(P_E)$ such that

$$\Pr_{(V_A^i, V_B^i) \leftarrow \mathcal{GV}(M^i, P_E)}[q \in \mathcal{Q}(V_A^i)] \geq \frac{\varepsilon}{n_B} \quad \text{or} \quad \Pr_{(V_A^i, V_B^i) \leftarrow \mathcal{GV}(M^i, P_E)}[q \in \mathcal{Q}(V_B^i)] \geq \frac{\varepsilon}{n_A}$$

(i.e., $q$ is $(\varepsilon/n_B)$-heavy for Alice or $(\varepsilon/n_A)$-heavy for Bob with respect to the distribution $\mathcal{GV}(M^i, P_E)$) Eve asks the lexicographically first such $q$ from $H(\cdot)$, and adds $(q, H(q))$ to $P_E$. At the end of round $\ell$ (when Eve is also done with asking her oracle queries), Eve samples $(V'_A, V'_B) \leftarrow \mathcal{GV}(M^\ell, P_E^\ell)$ and outputs Alice's output $s'_A$ determined by $V'_A$ as its own output $s_E$.

Theorem 3.1 directly follows from the next two lemmas.

[13] We use the term seminormal to distinguish it from the normal form protocols defined in [IR89].

[14] Impagliazzo and Rudich [IR89] use the term *normal form* for protocols in which each party asks *exactly one* query before sending their messages in every round.

**Lemma 3.4** (Eve Finds the Key)**.** *The output $s_E$ of Eve of Construction 3.3 agrees with $s_B$ with probability at least $\rho - 10\varepsilon$ over the choice of $(r_A, r_B, H)$.*

**Lemma 3.5** (Efficiency of Eve)**.** *The probability that Eve of Construction 3.3 asks more than $n_A \cdot n_B/\varepsilon^2$ oracle queries is at most $10\varepsilon$.*

Before proving Lemmas 3.4 and 3.5 we first derive Theorem 3.1 from them.

**Proof of Theorem 3.1.** Suppose we modify the adversary Eve and abort it as soon as it asks more than $n_A \cdot n_B/\varepsilon^2$ queries and call the new adversary EffEve. By Lemmas 3.4 and 3.5 the output $s_E$ of EffEve still agrees with Bob's output $s_B$ with probability at least $\rho - 10\varepsilon - 10\varepsilon = \rho - 20\varepsilon$. Theorem 3.1 follows by using $\varepsilon = \delta/20 < 1/10$ and noting that $n_A \cdot n_B/(\delta/20)^2 = 400 \cdot n_A \cdot n_B/\delta^2$. □

### 3.3 Analysis of Attack

In this subsection we will prove Lemmas 3.4 and 3.5, but before doing so we need some definitions.

**Events over $\mathcal{E}$.** Event $\mathsf{Good}$ holds if and only if $Q_A^\ell \cap Q_B^\ell \subseteq Q_E^\ell$ in which case we say that Eve has found all the *intersection queries*. Event $\mathsf{Fail}$ holds if and only if at *some* point during the execution of the system, Alice or Bob asks a query $q$, which was asked by the other party, but not already asked by Eve. If the first query $q$ that makes $\mathsf{Fail}$ happen is Bob's $j$'th query we say the event $\mathsf{BFail}_j$ has happened, and if it is Alice's $j$'th query we say that the event $\mathsf{AFail}_j$ has happened. Therefore, $\mathsf{BFail}_1, \ldots, \mathsf{BFail}_{n_B}$ and $\mathsf{AFail}_1, \ldots, \mathsf{AFail}_{n_B}$ are disjoint events whose union is equal to $\mathsf{Fail}$. Also note that $\neg\mathsf{Good} \Rightarrow \mathsf{Fail}$, because if Alice and Bob share a query that Eve never made, this must have happened *for the first time* at some point during the execution of the protocol (making $\mathsf{Fail}$ happen), but also note that $\mathsf{Good}$ and $\mathsf{Fail}$ are not necessarily complement events in general. Finally let the event $\mathsf{BGood}_j$ (resp. $\mathsf{AGood}_j$) be the event that when Bob (resp. Alice) asks his (resp. her) $j$'th oracle query, and this happens in round $i+1$, it holds that $Q_A^i \cap Q_B^i \subseteq Q_E^i$. Note that the event $\mathsf{BFail}_i$ implies $\mathsf{BGood}_i$ because if $\mathsf{BGood}_i$ does not hold, it means that Alice and Bob have *already* had an intersection query out of Eve's queries, and so $\mathsf{BFail}_i$ could not be the *first* time that Eve is missing an intersection query.

The following lemma plays a central role in proving both of Lemmas 3.5 and 3.4.

**Lemma 3.6** (Eve Finds the Intersection Queries)**.** *For all $i \in [n_B]$, $\Pr_{\mathcal{E}}[\mathsf{BFail}_i] \leq \frac{3\varepsilon}{2n_B}$. Similarly, for all $i \in [n_A]$, $\Pr_{\mathcal{E}}[\mathsf{AFail}_i] \leq \frac{3\varepsilon}{2n_A}$. Therefore, by a union bound, $\Pr_{\mathcal{E}}[\neg\mathsf{Good}] \leq \Pr_{\mathcal{E}}[\mathsf{Fail}] \leq 3\varepsilon$.*

We will first prove Lemma 3.6 and then will use this lemma to prove Lemmas 3.5 and 3.4. In order to prove Lemma 3.6 itself, we will reduce it to stronger statements in two steps i.e., Lemmas 3.7 and 3.8. Lemma 3.8 (called the graph characterization lemma) is indeed at the heart of our proof and characterizes the conditional distribution of the views of Alice and Bob conditioned on Eve's view.

#### 3.3.1 Eve Finds Intersection Queries: Proving Lemma 3.6

As we will show shortly, Lemma 3.6 follows from the following stronger lemma.

**Lemma 3.7.** *Let $B_i$, $M_i$, and $P_i$ denote, in order, Bob's view, the sequence of messages sent between Alice and Bob, and the oracle query-answer pairs known to Eve, all before the moment that Bob is going to ask his $i$'th oracle query that might happen be in a round $j$ that is* different *from $\geq i$.*[15] *Then, for* every *$(B_i, M_i, P_i) \leftarrow (\mathbf{B}_i, \mathbf{M}_i, \mathbf{P}_i)$ sampled by executing the system it holds that*

$$\Pr_{\mathcal{GV}(M_i,P_i)}[\mathsf{BFail}_i \mid B_i] \leq \frac{3\varepsilon}{2n_B}.$$

*A symmetric statement holds for Alice.*

We first see why Lemma 3.7 implies Lemma 3.6.

*Proof of Lemma 3.6 using Lemma 3.7.* It holds that

$$\Pr[\mathsf{BFail}_i] = \sum_{(B_i,M_i,P_i)\in \mathrm{Supp}(\mathbf{B}_i,\mathbf{M}_i,\mathbf{P}_i)} \Pr_{\mathcal{E}}[B_i, M_i, P_i] \cdot \Pr_{\mathcal{E}}[\mathsf{BFail}_i \mid B_i, M_i, P_i].$$

Recall that as we said the event $\mathsf{BFail}_i$ implies $\mathsf{BGood}_i$. Therefore, it holds that

$$\Pr_{\mathcal{E}}[\mathsf{BFail}_i \mid B_i, M_i, P_i] \leq \Pr_{\mathcal{E}}[\mathsf{BFail}_i \mid B_i, M_i, P_i, \mathsf{BGood}_i]$$

and by definition we have $\Pr_{\mathcal{E}}[\mathsf{BFail}_i \mid B_i, M_i, P_i, \mathsf{BGood}_i] = \Pr_{\mathcal{GV}(M_i,P_i)}[\mathsf{BFail}_i \mid B_i]$. By Lemma 3.7 it holds that $\Pr_{\mathcal{GV}(M_i,P_i)}[\mathsf{BFail}_i \mid B_i] \leq \frac{3\varepsilon}{2n_B}$, and so:

$$\Pr_{\mathcal{E}}[\mathsf{BFail}_i] \leq \sum_{(B_i,M_i,P_i)\in \mathrm{Supp}(\mathbf{B}_i,\mathbf{M}_i,\mathbf{P}_i)} \Pr_{\mathcal{E}}[B_i, M_i, P_i] \cdot \frac{3\varepsilon}{2n_B} = \Pr[\text{Bob asks} \geq i \text{ queries}] \cdot \frac{3\varepsilon}{2n_B} \leq \frac{3\varepsilon}{2n_B}.$$

□

In the following we will prove Lemma 3.7. In fact, we will not use the fact that Bob is about to ask his $i$'th query and will prove a more general statement. For simplicity we will use a simplified notation $M = M_i, P = P_i$. Suppose $M = M^j$ (namely the number of messages in $M$ is $j$). The following graph characterization of the distribution $\mathcal{V}(M, P)$ is at the heart of our analysis of the attacker Eve of Construction 3.3. We first describe the intuition and purpose behind the lemma.

**Intuition.** Lemma 3.8 below, intuitively, asserts that at any time during the execution of the protocol, while Eve is running her attack, the following holds. Let $(M, P)$ be the view of Eve at any moment. Then the distribution $\mathcal{V}(M, P)$ of Alice's and Bob's views conditioned on $(M, P)$ could be sampled using a "labeled" bipartite graph $G$ by sampling a uniform edge $e = (u, v)$ and taking the two labels of these two nodes (denoted by $A_u, B_v$). This graph $G$ has the extra property of being "dense" and close to being a complete bipartite graph.

**Lemma 3.8** (Graph Characterization of $\mathcal{V}(M, P)$)**.** *Let $M$ be the sequence of messages sent between Alice and Bob, let $P$ be the set of oracle query-answer pairs known to Eve by the end of the round in which the last message in $M$ is sent and Eve is also done with her learning queries. Let $\Pr_{\mathcal{V}(M,P)}[\mathsf{Good}(M, P)] > 0$. For* every *such $(M, P)$, there is a bipartite graph $G$ (depending on $M, P$) with vertices $(\mathcal{U}_A, \mathcal{U}_B)$ and edges $E$ such that:*

[15] Also note that $M_i$ is not necessarily the same as $M^i$. The latter refers to the transcript till the $i$'th message of the protocol is sent, while the former refers to the messages till Bob is going to ask his $i$'th messages (and might ask zero or more than one queries in some rounds).

1. *Every vertex $u$ in $\mathcal{U}_A$ has a corresponding view $A_u$ for Alice (which is consistent with $(M,P)$) and a set $Q_u = \mathcal{Q}(A_u) \setminus \mathcal{Q}(P)$, and the same holds for vertices in $\mathcal{U}_B$ by changing the role of Alice and Bob. (Note that every view can have multiple vertices assigned to it.)*

2. *There is an edge between $u \in \mathcal{U}_A$ and $v \in \mathcal{U}_B$ if and only if $Q_u \cap Q_v = \varnothing$.*

3. *Every vertex is connected to at least a $(1-2\varepsilon)$ fraction of the vertices in the other side.*

4. *The distribution $(V_A, V_B) \leftarrow \mathcal{GV}(M,P)$ is identical to: sampling a random edge $(u,v) \leftarrow E$ and taking $(A_u, B_v)$ (i.e., the views corresponding to $u$ and $v$).*

5. *The distributions $\mathcal{GV}(M,P)$ and $\mathcal{V}(M,P)$ have the same support set.*

Lemma 3.8 at the heart of the proof of our main theorem, and so we will first see how to use this lemma before proving it. In particular, we first use Lemma 3.8 to prove Lemma 3.7, and then we will prove Lemma 3.8.

**Proof of Lemma 3.7 using Lemma 3.8.** Let $B = B_i, M = M_i, P = P_i$ be as in Lemma 3.7 and $q$ be Bob's $i$'th query which is going to be asked after the last message $m_j$ in $M = M_i = M^j$ is sent to Bob. By Lemma 3.8, the distribution $\mathcal{GV}(M,P)$ conditioned on getting $B$ as Bob's view is the same as uniformly sampling a random edge $(u,v) \leftarrow E$ in the graph $G$ of Lemma 3.8 conditioned on $B_v = B$. We prove Lemma 3.7 even conditioned on choosing any vertex $v$ such that $B_v = B$. For such fixed $v$, the distribution of Alice's view $A_u$, when we choose a random edge $(u,v')$ conditioned on $v = v'$ is the same as choosing a random neighbor $u \leftarrow N(v)$ of the node $v$ and then selecting Alice's view $A_u$ corresponding to the node $u$. Let $S = \{u \in \mathcal{U}_A \text{ such that } q \in A_u\}$. Assuming $d(u)$ denotes the degree of $w$ for any node $w$ we have

$$\Pr_{u \leftarrow N(v)}[q \in A_u] \leq \frac{|S|}{d(v)} \leq \frac{|S|}{(1-2\varepsilon)\cdot|\mathcal{U}_A|} \leq \frac{|S|\cdot|\mathcal{U}_B|}{(1-2\varepsilon)\cdot|E|} \leq \frac{\sum_{u\in S} d(u)}{(1-2\varepsilon)^2\cdot|E|} \leq \frac{\varepsilon}{(1-2\varepsilon)^2\cdot n_B} < \frac{3\varepsilon}{2n_B}.$$

First note that proving the above inequality is sufficient for the proof of Lemma 3.7, because $\mathsf{BFail}_i$ is equivalent to $q \in A_u$. Now, we prove the above inequalities.

The second and fourth inequalities are due to the degree lower bounds of Item 3 in Lemma 3.8. The third inequality is because $|E| \leq |\mathcal{U}_A| \cdot |\mathcal{U}_B|$. The fifth inequality is because of the definition of the attacker Eve who asks $\varepsilon/n_B$ heavy queries for Alice's view when sampled from $\mathcal{GV}(M,P)$, as long as such queries exist. Namely, when we choose a random edge $(u,v) \leftarrow E$ (which by Item 4 of Lemma 3.8 is the same as sampling $(V_A, V_B) \leftarrow \mathcal{GV}(M,P)$), it holds that $u \in S$ with probability $\sum_{u\in S} d(u)/|E|$. But for all $u \in S$ it holds that $q \in Q_u$, and so if $\sum_{u\in S} d(u)/|E| > \varepsilon/n_B$ the query $q$ should have been learned by Eve already and so $q$ could not be in any set $Q_u$. The sixth inequality is because we are assuming $\varepsilon < 1/10$. □

### 3.3.2 The Graph Characterization: Proving Lemma 3.8

We prove Lemma 3.8 by first presenting a "product characterization" of the distribution $\mathcal{GV}(M,P)$.[16]

**Lemma 3.9** (Product Characterization)**.** *For any $(M,P)$ as described in Lemma 3.8 there exists a distribution $\mathbf{A}$ (resp. $\mathbf{B}$) over Alice's (resp. Bob's) views such that the distribution $\mathcal{GV}(M,P)$ is identical to the product distribution $(\mathbf{A} \times \mathbf{B})$ conditioned on the event $\mathsf{Good}(M,P)$. Namely,*

[16] A similar observation was made by [IR89], see Lemma 6.5 there.

$$\mathcal{GV}(M,P) \equiv ((\mathbf{A}\times\mathbf{B}) \mid \mathcal{Q}(\mathbf{A})\cap\mathcal{Q}(\mathbf{B})\subseteq\mathcal{Q}(P)).$$

*Proof.* Suppose $(V_A,V_B)\leftarrow\mathcal{V}(M,P)$ is such that $Q_A\cap Q_B\subseteq Q$ where $Q_A=\mathcal{Q}(V_A), Q_B=\mathcal{Q}(V_B)$, and $Q=\mathcal{Q}(P)$. For such $(V_A,V_B)$ we will show that $\Pr_{\mathcal{GV}(M,P)}[(V_A,V_B)]=\alpha(M,P)\cdot\alpha_A\cdot\alpha_B$ where: $\alpha(M,P)$ only depends on $(M,P)$, $\alpha_A$ only depends on $V_A$, and $\alpha_B$ only depends only on $V_B$. This means that if we let $\mathbf{A}$ be the distribution over $\mathrm{Supp}(V_A)$ such that $\Pr_{\mathbf{A}}[V_A]$ is proportional to $\alpha_A$ and let $\mathbf{B}$ be the distribution over $\mathrm{Supp}(V_B)$ such that $\Pr_{\mathbf{B}}[V_B]$ is proportional to $\alpha_B$, then $\mathcal{GV}(M,P)$ is proportional (and hence equal to) the distribution $((\mathbf{A}\times\mathbf{B}) \mid Q_A\cap Q_B\subseteq Q)$.

In the following we will show that $\Pr_{\mathcal{GV}(M,P)}[(V_A,V_B)]=\alpha(M,P)\cdot\alpha_A\cdot\alpha_B$. Since we are assuming $Q_A\cap Q_B\subseteq Q$ (i.e., that the event $\mathsf{Good}(M,P)$ holds over $(V_A,V_B)$) we have:

$$\Pr_{\mathcal{V}(M,P)}[(V_A,V_B)]=\Pr_{\mathcal{V}(M,P)}[(V_A,V_B)\wedge\mathsf{Good}(M,P)]=\Pr_{\mathcal{V}(M,P)}[\mathsf{Good}(M,P)]\Pr_{\mathcal{GV}(M,P)}[(V_A,V_B)]. \quad (1)$$

On the other hand, by definition of conditional probability we have[17]

$$\Pr_{\mathcal{V}(M,P)}[(V_A,V_B)]=\frac{\Pr_{\mathcal{E}}[(V_A,V_B,M,P)]}{\Pr_{\mathcal{E}}[(M,P)]}. \quad (2)$$

Therefore, by Equations (1) and (2) we have

$$\Pr_{\mathcal{GV}(M,P)}[(V_A,V_B)]=\frac{\Pr_{\mathcal{E}}[(V_A,V_B,M,P)]}{\Pr_{\mathcal{E}}[(M,P)]\cdot\Pr_{\mathcal{V}(M,P)}[\mathsf{Good}(M,P)]}. \quad (3)$$

The denominator of the righthand side of Equation (3) only depends on $(M,P)$ and so we can take $\beta(M,P)=\Pr_{\mathcal{E}}[(M,P)]\cdot\Pr_{\mathcal{V}(M,P)}[\mathsf{Good}(M,P)]$. In the following we analyze the numerator.

Recall that for a partial function $F$, by $\Pr_{\mathcal{E}}[F]$ we denote the probability that $H$ from the sampled execution $(r_A,r_B,H)\leftarrow\mathcal{E}$ is consistent with $F$; namely, $\Pr_{\mathcal{E}}[F]=\Pr_{\mathbf{H}}[F]$ (see Definition 2.2).

Let $P_A$ (resp. $P_B$) be the set of oracle query-answer pairs in $V_A$ (resp. $V_B$). We claim that:

$$\Pr_{\mathcal{E}}[(V_A,V_B,M,P)]=\Pr[\mathbf{r}_A=r_A]\cdot\Pr[\mathbf{r}_B=r_B]\cdot\Pr_{\mathcal{E}}[P_A\cup P_B\cup P].$$

The reason is that the necessary and sufficient condition that $(V_A,V_B,M,P)$ happens in the execution of the system is that when we sample a uniform $(r_A,r_B,H)$, $r_A$ equals Alice's randomness, $r_B$ equals Bob's randomness, and $H$ is consistent with $P_A\cup P_B\cup P$. These conditions implicitly imply that Alice and Bob will indeed produce the transcript $M$ as well.

Now by Lemma 2.5 and $(P_A\cap P_B)\setminus P=\varnothing$ we have $\Pr_{\mathcal{E}}[P_A\cup P_B\cup P]$ equals to:

$$\Pr_{\mathcal{E}}[P]\cdot\Pr_{\mathcal{E}}[(P_A\cup P_B)\setminus P]=\frac{\Pr_{\mathcal{E}}[P]\cdot\Pr_{\mathcal{E}}[P_A\setminus P]\cdot\Pr_{\mathcal{E}}[P_B\setminus P]}{\Pr_{\mathcal{E}}[(P_A\cap P_B)\setminus P]}=\Pr_{\mathcal{E}}[P]\cdot\Pr_{\mathcal{E}}[P_A\setminus P]\cdot\Pr_{\mathcal{E}}[P_B\setminus P].$$

Therefore, we get:

$$\Pr_{\mathcal{GV}(M,P)}[(V_A,V_B)]=\frac{\Pr[\mathbf{r}_A=r_A]\cdot\Pr[\mathbf{r}_B=r_B]\cdot\Pr_{\mathcal{E}}[P]\cdot\Pr_{\mathcal{E}}[P_A\setminus P]\cdot\Pr_{\mathcal{E}}[P_B\setminus P]}{\beta(M,P)}.$$

and so we can take $\alpha_A=\Pr[\mathbf{r}_A=r_A]\cdot\Pr_{\mathcal{E}}[P_A\setminus P]$, $\alpha_B=\Pr[\mathbf{r}_B=r_B]\cdot\Pr_{\mathcal{E}}[P_B\setminus P]$, and $\alpha(M,P)=\Pr_{\mathcal{E}}[P]/\beta(M,P)$. □

[17] Note that $V_A,V_B$ uniquely determine $M,P$ so $\Pr[V_A,V_B,M,P]=\Pr[V_A,V_B]$ holds for consistent $V_A,V_B,M,P$, but we choose to write the full event's description for clarity.

**Graph Characterization.** The product characterization of Lemma 3.9 implies that we can think of $\mathcal{GV}(M,P)$ as a distribution over random edges of some bipartite graph $G=(\mathcal{U}_A,\mathcal{U}_B,E)$ defined based on $(M,P)$ as follows.

**Construction 3.10** (Labeled graph $G=(\mathcal{U}_A,\mathcal{U}_B,E)$)**.** Every node $u\in\mathcal{U}_A$ will have a corresponding view $A_u$ of Alice that is in the support of the distribution $\mathbf{A}$ from Lemma 3.9. We also let the number of nodes corresponding to a view $V_A$ be proportional to $\Pr_{\mathbf{A}}[\mathbf{A}=V_A]$, meaning that $\mathbf{A}$ corresponds to the uniform distribution over the left-side vertices $\mathcal{U}_A$. Similarly, every node $v\in\mathcal{U}_B$ will have a corresponding view $B_v$ of Bob such that $\mathbf{B}$ corresponds to the uniform distribution over $\mathcal{U}_B$. Doing this is possible because the probabilities $\Pr_{\mathbf{A}}[\mathbf{A}=V_A]$ and $\Pr_{\mathbf{B}}[\mathbf{B}=V_B]$ are all rational numbers. More formally, since in Definition 2.2 of random oracles we assumed $\mathbf{H}(x)=y$ to be rational for all $(x,y)$, the probability space $\mathcal{GV}(M,P)$ only includes rational probabilities. Thus, if $W_1,\ldots,W_N$ is the list of all possible views for Alice when sampling $(V_A,V_B)\leftarrow\mathcal{GV}(M,P)$, and if $\Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[W_j=V_A]=c_j/d_j$ where $c_1,d_1,\ldots,c_N,d_N$ are all integers, we can put $(c_j/d_j)\cdot\prod_{i\in[N]}d_i$ many nodes in $\mathcal{U}_A$ representing the view $W_j$. Now if we sample a node $u\leftarrow\mathcal{U}_A$ uniformly and take $A_u$ as Alice's view, it would be the same as sampling $(V_A,V_B)\leftarrow\mathcal{GV}(M,P)$ and taking $V_A$. Finally, we define $Q_u=Q(A_u)\setminus Q(P)$ for $u\in\mathcal{U}_A$ to be the set of queries *outside of* $\mathcal{Q}(P)$ that were asked by Alice in the view $A_u$. We define $Q_v=Q(B_u)\setminus Q(P)$ similarly. We put an edge between the nodes $u$ and $v$ (denoted by $u\sim v$) in $G$ if and only if $Q_u\cap Q_v=\varnothing$.

It turns out that the graph $G$ is *dense* as formalized in the next lemma.

**Lemma 3.11.** *Let $G=(\mathcal{U}_A,\mathcal{U}_B,E)$ be the graph of Construction 3.10. Then for every $u\in\mathcal{U}_A, d(u)\geq|\mathcal{U}_B|\cdot(1-2\varepsilon)$ and for every $v\in\mathcal{U}_B$, $d(v)\geq|\mathcal{U}_A|\cdot(1-2\varepsilon)$ where $d(w)$ is the degree of the vertex $w$.*

*Proof.* First note that Lemma 3.9 and the description of Construction 3.10 imply that the distribution $\mathcal{GV}(M,P)$ is equal to the distribution obtained by letting $(u,v)$ be a random edge of the graph $G$ and choosing $(A_u,B_v)$. We will make use of this property.

We first show that for every $w\in\mathcal{U}_A$, $\sum_{v\in\mathcal{U}_B,w\not\sim v}d(v)\leq\varepsilon\cdot|E|$. The reason is that the probability of vertex $v$ being chosen when we choose a random edge is $\frac{d(v)}{|E|}$ and if $\sum_{v\in\mathcal{U}_B,w\not\sim v}\frac{d(v)}{|E|}>\varepsilon$, it means that $\Pr_{(u,v)\leftarrow E}[Q_w\cap Q_v\neq\varnothing]\geq\varepsilon$. Hence, because $|Q_w|\leq n_A$, by the pigeonhole principle there would exist $q\in Q_w$ such that $\Pr_{(u,v)\leftarrow E}[q\in Q_v]\geq\varepsilon/n_A$. But this is a contradiction, because if that holds, then $q$ should have been in $P$ by the definition of the attacker Eve of Construction 3.3, and hence it could not be in $Q_w$. The same argument shows that for every $w\in\mathcal{U}_B$, $\sum_{u\in\mathcal{U}_A,u\not\sim w}d(u)\leq\varepsilon|E|$. Thus, for every vertex $w\in\mathcal{U}_A\cup\mathcal{U}_B$, $\left|E^{\not\sim}(w)\right|\leq\varepsilon|E|$ where $E^{\not\sim}(w)$ denotes the set of edges that do not contain any neighbor of $w$ (i.e., $E^{\not\sim}(w)=\{(u,v)\in E\mid u\not\sim w\wedge w\not\sim v\}$). The following claim proves Lemma 3.11.

**Claim 3.12.** *For $\varepsilon\leq1/2$, let $G=(\mathcal{U}_A,\mathcal{U}_B,E)$ be a nonempty bipartite graph where $\left|E^{\not\sim}(w)\right|\leq\varepsilon|E|$ for all vertices $w\in\mathcal{U}_A\cup\mathcal{U}_B$. Then $d(u)\geq|\mathcal{U}_B|\cdot(1-2\varepsilon)$ for all $u\in\mathcal{U}_A$ and $d(v)\geq|\mathcal{U}_A|\cdot(1-2\varepsilon)$ for all $v\in\mathcal{U}_B$.*

*Proof.* Let $d_A=\min\{d(u)\mid u\in\mathcal{U}_A\}$ and $d_B=\min\{d(v)\mid v\in\mathcal{U}_B\}$. By switching the left and right sides if necessary, we may assume without loss of generality that

$$\frac{d_A}{|\mathcal{U}_B|}\leq\frac{d_B}{|\mathcal{U}_A|}. \tag{4}$$

So it suffices to prove that $1-2\varepsilon \leq \frac{d_A}{|\mathcal{U}_B|}$. Suppose $1-2\varepsilon > \frac{d_A}{|\mathcal{U}_B|}$, and let $u \in \mathcal{U}_A$ be the vertex that $d(u) = d_A < (1-2\varepsilon)\,|\mathcal{U}_B|$. Because for all $v \in \mathcal{U}_B$ we have $d(v) \leq |\mathcal{U}_A|$, thus, using Inequality (4) we get that $|E^{\sim}(u)| \leq d_A\,|\mathcal{U}_A| \leq d_B\,|\mathcal{U}_B|$ where $E^{\sim}(u) = E \setminus E^{\not\sim}(u)$. On the other hand since we assumed that $d(u) < (1-2\varepsilon)|\mathcal{U}_B|$, there are more than $2\varepsilon|\mathcal{U}_B|d_B$ edges in $E^{\not\sim}(u)$, meaning that $|E^{\sim}(u)| < \left|E^{\not\sim}(u)\right|/(2\varepsilon)$. But this implies

$$|E^{\not\sim}(u)| \leq \varepsilon|E| = \varepsilon\left(|E^{\not\sim}(u)| + |E^{\sim}(u)|\right) < \varepsilon|E^{\not\sim}(u)| + |E^{\not\sim}(u)|/2,$$

which is a contradiction for $\varepsilon < 1/2$. □

Finally we prove Item 5. Namely, for every $(A,B) \leftarrow \mathcal{V}(\mathbf{V}_A, \mathbf{V}_B)$, there is some $B'$ such that $(A,B')$ is in the support set of $\mathcal{GV}(\mathbf{V}_A, \mathbf{V}_B)$. The latter is equivalent to finding $B'$ that is consistent with $M, P$ and that $\mathcal{Q}(A) \cap \mathcal{Q}(B) \subseteq \mathcal{Q}(P)$. For sake of contradiction suppose this is not the case. Therefore, if we sample $B'$ from the distribution of $\mathbf{V}_B$ conditioned on $P, M$ then there is always an element in $\mathcal{Q}(A) \cap \mathcal{Q}(B')$ that is outside of $c\mathcal{Q}(P)$. By the pigeonhole principle, one of the queries in $\mathcal{Q}(A) \setminus \mathcal{Q}(P)$ would be at least $1/n_A$-heavy for the distribution $\mathcal{GV}(\mathbf{V}_A, \mathbf{V}_B)$ (in particular the $\mathbf{V}_B$ part). But this contradicts how the algorithm of Eve operates. □

**Remark 3.13** (Sufficient Condition for Graph Characterization)**.** It can be verified that the proof of the graph characterization of Lemma 3.8 only requires the following: At the end of the rounds, Eve has learned all the $(\varepsilon/n_B)$-heavy queries for Alice and all the $(\varepsilon/n_A)$-heavy queries for Bob with respect to the distribution $\mathcal{GV}(M,P)$. More formally, all we need is that when Eve stops asking more queries, if there is any query $q$ such that

$$\Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[q \in \mathcal{Q}(V_A)] \geq \frac{\varepsilon}{n_B} \quad \text{or} \quad \Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[q \in \mathcal{Q}(V_B)] \geq \frac{\varepsilon}{n_A}$$

then $q \in \mathcal{Q}(P)$. In particular, Lemma 3.8 holds even if Eve arbitrarily asks queries that are *not* necessarily heavy at the time being asked or chooses to ask the heavy queries in an arbitrary (different than lexicographic) order.

### 3.3.3 Eve Finds the Key: Proving Lemma 3.4

Now, we turn to the question of finding the secret. Theorem 6.2 in [IR89] shows that once one finds all the intersection queries, with $O(n^2)$ more queries they can also find the actual secret. Here we use the properties of our attack to show that we can do so even without asking more queries.

First we need to specify and prove the following corollary of of Lemma 3.8.

**Corollary 3.14** (Corollary of Lemma 3.8)**.** *Let Eve be the eavesdropping adversary of Construction 3.3 using parameter $\varepsilon$, and $\Pr_{\mathcal{V}(M^i,P_E^i)}[\mathsf{Good}(M^i,P_E^i)] > 0$ where $(M^i,P_E^i)$ is the view of Eve by the end of round $i$ (when she is also done with learning queries). For the fixed $i, M^i, P_E^i$, let $(\mathbf{V}_A,\mathbf{V}_B)$ be the joint view of Alice and Bob as sampled from $\mathcal{GV}(M^i,P_E^i)$. Then for some product distribution $(\mathbf{U}_A \times \mathbf{U}_B)$ (where $\mathbf{U}_A \times \mathbf{U}_B$ could also depend on $i, M^i, P_E^i$) we have:*

1. $\Delta((\mathbf{V}_A,\mathbf{V}_B),(\mathbf{U}_A \times \mathbf{U}_B)) \leq 2\varepsilon$.

2. *For* every *possible* $(A, B) \leftarrow \mathcal{V}(\mathbf{V}_A, \mathbf{V}_B)$ *(which by Item 5 is the same as the set of all* $(A, B) \leftarrow \mathcal{GV}(\mathbf{V}_A, \mathbf{V}_B)$*) we have:*

$$\Delta((\mathbf{V}_A \mid \mathbf{V}_B = B), \mathbf{U}_A) \leq 2\varepsilon,$$
$$\Delta((\mathbf{V}_B \mid \mathbf{V}_A = A), \mathbf{U}_B) \leq 2\varepsilon.$$

*Proof.* In the graph characterization $G = (\mathcal{U}_A, \mathcal{U}_B, E)$ of $\mathcal{GV}(M, P)$ as described in Lemma 3.8, every vertex is connected to $1 - 2\varepsilon$ fraction of the vertices of the other section, and consequently the graph $G$ has $1 - 2\varepsilon$ fraction of the edges of the complete bipartite graph with the same nodes $(\mathcal{U}_A, \mathcal{U}_B)$. Thus, if we take $\mathbf{U}_A$ the uniform distribution over $\mathcal{U}_A$ and $\mathbf{U}_B$ the uniform distribution over $\mathcal{U}_B$, they satisfy all the three inequalities. □

The process of sampling the components of the system can also be done in a "reversed" order where we first decide about whether some events are going to hold or not and then sample the other components conditioned on that.

**Notation.** In the following let $s(V)$ be the output determined by any view $V$ (of Alice or Bob)

**Construction 3.15.** Sample Alice, Bob, and Eve's views as follows.

1. Toss a coin $b$ such that $b = 1$ with probability $\Pr_{\mathcal{E}}[\mathsf{Good}]$.
2. If $b = 1$:
   (a) Sample Eve's final view $(M, P)$ conditioned on $\mathsf{Good}$.
   (b) i. Sample views of Alice and Bob $(V_A, V_B)$ from $\mathcal{GV}(M, P)$.
   ii. Eve samples $(V'_A, V'_B) \leftarrow \mathcal{GV}(M, P)$, and outputs $s_E = s(V'_A)$.
3. If $b = 0$:
   (a) Sample Eve's final view $(M, P)$ conditioned on $\neg\mathsf{Good}$.
   (b) i. Sample views $(V_A, V_B) \leftarrow (\mathcal{V}(M, P) \mid \neg\mathsf{Good})$.
   ii. Eve does the same as case $b = 1$ above.

In other words, $b = 1$ if and only if $\mathsf{Good}$ holds over the real views of Alice and Bob. We might use $b = 1$ and $\mathsf{Good}$ interchangeably (depending on which one is conceptually more convenient).

The attacker Eve of Construction 3.3 samples views $(V'_A, V'_B)$ from $\mathcal{GV}(M, P)$ in *both* cases of $b = 0$ and $b = 1$, and that is exactly what the Eve of Construction 3.15 does as well, and the pair $(s_E, s(V_B))$ in Constructions 3.3 vs. 3.15 are identically distributed. Therefore, our goal is to lower bound the probability of getting $s_E = s(V_B)$ where $s_E = s(V'_A)$ is the output of $V'_A$ and $s(V_B)$ is the output of $V_B$ (in Construction 3.15). We would show that this event happens in Step 2b with sufficiently large probability. (Note that it is also possible that $s_E = s(V_B)$ happens in Step 3b as well, but we ignore this case.)

In the following, let $\rho(M, P)$ and $\mathtt{win}(M, P)$ be defined as follows.

$$\rho(M, P) = \Pr_{(V_A, V_B) \leftarrow \mathcal{GV}(M,P)}[s(V_A) = s(V_B)]$$
$$\mathtt{win}(M, P) = \Pr_{(V_A, V_B) \leftarrow \mathcal{GV}(M,P), (V'_A, V'_B) \leftarrow \mathcal{GV}(M,P)}[s(V'_A) = s(V_B)]$$

where $(V_A, V_B)$ and $(V'_A, V'_B)$ are independent samples.

We will prove Lemma 3.4 using the following two claims.

**Claim 3.16.** *Suppose $P$ denotes Eve's set of oracle query-answer pairs after all of the messages in $M$ are sent. Assuming the probability of $\mathsf{Good}(M,P)$ is nonzero conditioned on $(M,P)$, for every $\varepsilon < 1/10$ used by Eve's algorithm of Construction 3.3 it holds that*

$$\mathtt{win}(M,P) \geq \rho(M,P) - 4\varepsilon.$$

Now we prove Claim 3.16.

*Proof of Claim 3.16.* Let $(\mathbf{U}_A \times \mathbf{U}_B)$ be the product distribution of Corollary 3.14 for the view of $(M,P)$. We would like to lower bound the probability of $s(V'_A) = s(V_B)$ where $(V_A, V_B)$ and $(V'_A, V'_B)$ are independent samples from the same distribution $(\mathbf{V}_A, \mathbf{V}_B) \equiv \mathcal{GV}(M,P)$. Since $M, P$ are fixed, for simplicity of notation, in the following we let $(\mathbf{V}_A, \mathbf{V}_B) \equiv \mathcal{GV}(M,P)$ without explicitly mentioning $M, P$. Also, in what follows, $\mathbf{V}_A$ (resp. $\mathbf{V}_B$) will denote the marginal distribution of the first (resp. second) component of $(\mathbf{V}_A, \mathbf{V}_B)$. We will also preserve $V_A, V_B$ to denote the real and Bob views sampled from $(\mathbf{V}_A, \mathbf{V}_B)$, and we will use $V'_A, V'_B$ to denote Eve's samples from the same distribution $(\mathbf{V}_A, \mathbf{V}_B)$.

For every possible view $A_0 \leftarrow \mathbf{V}_A$, let $\rho(A_0) = \Pr_{(A,B)\leftarrow(\mathbf{V}_A,\mathbf{V}_B))}[s(A) = s(B) \mid A = A_0]$. By averaging over Alice's view, it holds that $\rho(M,P) = \mathbb{E}_{(A,B)\leftarrow(\mathbf{V}_A,\mathbf{V}_B)}[\rho(A)]$. Similarly, for every possible view $A_0 \leftarrow \mathbf{V}_A$, let $\mathtt{win}(A_0) = \Pr_{(A,B)\leftarrow(\mathbf{V}_A,\mathbf{V}_B))}[s(A) = s(B)]$. By averaging over Alice's view, it holds that $\rho(M,P) = \mathbb{E}_{(A,B)\leftarrow(\mathbf{V}_A,\mathbf{V}_B)}[\rho(A)]$ and $\mathtt{win}(M,P) = \mathbb{E}_{(A,B)\leftarrow(\mathbf{V}_A,\mathbf{V}_B)}[\mathtt{win}(A)]$

In the following, we will prove something stronger than Claim 3.16 and will show that $\mathtt{win}(V'_A) \geq \rho(V'_A) - 4\varepsilon$ for *every* $V'_A \leftarrow \mathbf{V}_A$, and the claim follows by averaging over $V'_A \leftarrow \mathbf{V}_A$. Thus, in the following $V'_A$ will be the fixed sample $V'_A \leftarrow \mathbf{V}_A$. By Corollary 3.14, for every possible Alice's view $A \leftarrow \mathbf{V}_A$, the distribution of Bob's view sampled from $(\mathbf{V}_B \mid \mathbf{V}_A = A)$ is $2\varepsilon$-close to $\mathbf{U}_B$. Therefore, the distribution of $\mathbf{V}_B$ (without conditioning on $\mathbf{V}_A = A$) is also $2\varepsilon$-close to $\mathbf{U}_B$. By two applications of Lemma 2.9 we get

$$\begin{aligned}
\mathtt{win}(V'_A) &= \Pr_{V_B \leftarrow \mathbf{V}_B}[s(V'_A) = s(V_B)] \\
&\geq \Pr_{B \leftarrow \mathbf{U}_B}[s(V'_A) = s(B)] - 2\varepsilon \\
&\geq \Pr_{V'_B \leftarrow (\mathbf{V}_B \mid \mathbf{V}_A = V'_A)}[s(V'_A) = s(V'_B)] - 4\varepsilon \\
&= \rho(V'_A) - 4\varepsilon.
\end{aligned}$$

□

The following claim lower bounds the completeness of the key agreement protocol when conjuncted with reaching Step 2b in Construction 3.15.

**Claim 3.17.** *It holds that* $\Pr_{\mathcal{E}}[s(V_A) = s(V_B) \wedge \mathsf{Good}] \geq \rho - 3\varepsilon$.

*Proof.* By Lemma 3.6 it holds that $1 - 3\varepsilon \leq \Pr_{\mathcal{E}}[\mathsf{Good}]$. Therefore

$$\rho - 3\varepsilon \leq \Pr_{\mathcal{E}}[s(V_A) = s(V_B)] - \Pr_{\mathcal{E}}[\neg\mathsf{Good}] = \Pr_{\mathcal{E}}[s(V_A) = s(V_B) \wedge \mathsf{Good}].$$

□

**Proof of Lemma 3.4.** We will show a stronger claim that $\Pr[s(V'_A) = s(V_B) \wedge \mathsf{Good}] \geq \rho - 7\varepsilon$ which implies $\Pr[s(V'_A) = s(V_B)] \geq \rho - 7\varepsilon$ as well. By definition of Construction 3.15 and using Claims 3.16 and 3.17 we have:

$$\begin{aligned}
\Pr[s(V'_A) = s(V_B) \wedge \mathsf{Good}] &= \Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \mathop{\mathbb{E}}_{(M,P)\leftarrow((\mathbf{M},\mathbf{P})|\mathsf{Good})}[\mathtt{win}(M,P)] \\
&\geq \Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \mathop{\mathbb{E}}_{(M,P)\leftarrow((\mathbf{M},\mathbf{P})|\mathsf{Good})}[\rho(M,P) - 4\varepsilon] \\
&= \left(\Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \mathop{\mathbb{E}}_{(M,P)\leftarrow((\mathbf{M},\mathbf{P})|\mathsf{Good})}[\rho(M,P)]\right) - (4\Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \varepsilon) \\
&= \left(\Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \Pr[s(V_A) = s(V_B) \mid \mathsf{Good}]\right) - (4\Pr_{\mathcal{E}}[\mathsf{Good}] \cdot \varepsilon) \\
&\geq (\rho - 3\varepsilon) - (4\varepsilon) = \rho - 7\varepsilon.
\end{aligned}$$

□

### 3.3.4 Efficiency of Eve: Proving Lemma 3.5

Recall that Eve's criteria for "heaviness" is based on the distribution $\mathcal{GV}(M, P_E)$ where $M$ is the current sequence of messages sent so far and $P_E$ is the current set of oracle query-answer pairs known to Eve. This distribution is conditioned on Eve not missing any queries up to this point. However, because we have proven that the event $\mathsf{Fail}$ has small probability, queries that are heavy under $\mathcal{GV}(M, P_E)$ are also (typically) almost as heavy under the real distribution $\mathcal{V}(M, P_E)$. Intuitively this means that, on average, Eve will not make too many queries.

**Definition 3.18** (Coloring of Eve's Queries)**.** Suppose $(M^i, P_E)$ is the view of Eve at the moment Eve asks query $q$. We call $q$ a *red* query, denoted $q \in \mathsf{R}$, if $\Pr[\mathsf{Good}(M^i, P_E)] \leq 1/2$. We call $q$ a *green* query of Alice's type, denoted $q \in \mathsf{GA}$, if $q$ is not red and $\Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{V}(M^i,P_E)}[q \in \mathcal{Q}(V^i_A)] \geq \frac{\varepsilon}{2n_B}$. (Note that here we are sampling the views from $\mathcal{V}(M^i, P_E)$ and not from $\mathcal{GV}(M^i, P_E)$ and the threshold of "heaviness" is $\frac{\varepsilon}{2n_B}$ rather than $\frac{\varepsilon}{n_B}$.) Similarly, we call $q$ a green query of Bob's type, denoted $q \in \mathsf{GB}$, if $q$ is not red and $\Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{V}(M^i,P_E)}[q \in \mathcal{Q}(V^i_B)] \geq \frac{\varepsilon}{2n_A}$. We also let the set of all green queries to be $\mathsf{G} = \mathsf{GA} \cup \mathsf{GB}$.

The following claim shows that each of Eve's queries is either red or green.

**Claim 3.19.** *Every query $q$ asked by Eve is either in $\mathsf{R}$ or in $\mathsf{G}$.*

*Proof.* If $q$ is a query of Eve which is not red, then $\Pr_{\mathcal{V}(M^i,P_E)}[\mathsf{Good}(M^i, P_E)] \geq 1/2$ where $(M^i, P_E)$ is the view of Eve when asking $q$. Since Eve is asking $q$, either of the following holds:

1. $\Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{GV}(M^i,P_E)}[q \in \mathcal{Q}(V^i_A)] \geq \frac{\varepsilon}{n_B}$, or
2. $\Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{GV}(M^i,P_E)}[q \in \mathcal{Q}(V^i_B)] \geq \frac{\varepsilon}{n_A}$.

If case 1 holds, then

$$\begin{aligned}
\Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{V}(M^i,P_E)}[q \in \mathcal{Q}(V^i_A)] &\geq \Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{V}(M^i,P_E)}[\mathsf{Good}(M^i, P_E) \wedge q \in \mathcal{Q}(V^i_A)] \\
&= \Pr_{\mathcal{V}(M^i,P_E)}[\mathsf{Good}(M^i, P_E)] \cdot \Pr_{(V^i_A,V^i_B)\leftarrow\mathcal{GV}(M^i,P_E)}[q \in \mathcal{Q}(V^i_A)] \\
&\geq (\frac{1}{2}) \cdot \frac{\varepsilon}{n_B} = \frac{\varepsilon}{2n_B}
\end{aligned}$$

which implies that $q \in \mathsf{GA}$. Case 2 similarly shows that $q \in \mathsf{GB}$. □

We will bound the size of the queries of each color separately.

**Claim 3.20** (Bounding Red Queries)**.** $\Pr_{\mathcal{E}}[\mathsf{R} \neq \varnothing] \leq 6\varepsilon$.

**Claim 3.21** (Bounding Green Queries)**.** $\mathbb{E}_{\mathcal{E}}[|\mathsf{G}|] \leq 4n_A \cdot n_B/\varepsilon$. *Therefore, by Markov inequality,* $\Pr_{\mathcal{E}}[|\mathsf{G}| \geq n_A \cdot n_B/\varepsilon^2] \leq 4\varepsilon$.

**Proving Lemma 3.5.** Lemma 3.5 follows by a union bound and Claims 3.19, 3.20, and 3.21.

*Proof of Claim 3.20.* Claim 3.20 follows directly from Lemma 2.7 and Lemma 3.6 as follows. Let $\mathbf{x}$ (in Lemma 2.7) be $\mathcal{E}$, the event $E$ be $\mathsf{Fail}$, the sequence $\mathbf{x}_1, \ldots,$ be the sequence of pieces of information that Eve receives (i.e., the messages and oracle answers), $\lambda = 3\varepsilon$, $\lambda_1 = 1/2$ and $\lambda_2 = 6\varepsilon$. Lemma 3.6 shows that $\Pr[\mathsf{Fail}] \leq \lambda$. Therefore, if we let $D$ be the event that at some point conditioned on Eve's view the probability of $\mathsf{Fail}$ is more than $\lambda_1$, Lemma 2.7 shows that the probability of $D$ is at most $\lambda_2$. Also note that for every sampled $(M, P_E)$, $\Pr[\neg\mathsf{Good} \mid (M, P_E)] \leq \Pr[\mathsf{Fail} \mid (M, P_E)]$. Therefore, with probability at least $1 - \lambda_2 = 1 - 6\varepsilon$, during the execution of the system, the probability of $\mathsf{Good}(M, P_E)$ conditioned on Eve's view will never go below $1/2$. □

*Proof of Claim 3.21.* We will prove that $\mathbb{E}_{\mathcal{E}}[|\mathsf{GA}|] \leq 2n_A \cdot n_B/\varepsilon$, and $\mathbb{E}_{\mathcal{E}}[|\mathsf{GB}|] \leq 2n_A \cdot n_B/\varepsilon$ follows symmetrically. Using these two upper bounds we can derive Claim 3.21 easily.

For a fixed query $q \in \{0,1\}^\ell$, let $I_q$ be the event, defined over $\mathcal{E}$, that Eve asks $q$ as a green query of Alice's type (i.e., $q \in \mathsf{GA}$). Let $F_q$ be the event that Alice actually asks $q$ (i.e., $q \in Q_A$). By linearity of expectation we have $\mathbb{E}_{\mathcal{E}}[|\mathsf{GA}|] = \sum_q \Pr[I_q]$ and $\sum_q \Pr[F_q] \leq |Q_A| \leq n_A$. Let $\gamma = \frac{\varepsilon}{2n_B}$. We claim that for all $q$ it holds that:

$$\Pr[I_q] \cdot \gamma \leq \Pr[F_q]. \tag{5}$$

First note that Inequality (5) implies Claim 3.21 as follows:

$$\mathop{\mathbb{E}}_{\mathcal{E}}[|\mathsf{GA}|] = \sum_q \Pr[I_q] \leq \frac{1}{\gamma} \sum_q \Pr[F_q] \leq \frac{n_A}{\gamma} = \frac{2n_A n_B}{\varepsilon}.$$

To prove Inequality (5), we use Lemma 2.7 as follows. The underlying random variable $\mathbf{x}$ (of Lemma 2.7) will be $\mathcal{E}$, the event $E$ will be $F_q$, the sequence of random variables $\mathbf{x}_1, \mathbf{x}, \ldots$ will be the sequence of pieces of information that Eve observes, $\lambda$ will be $\Pr[F_q]$, and $\lambda_1$ will be $\gamma$. If $I_q$ holds, it means that based on Eve's view the query $q$ has at least $\gamma$ probability of being asked by Alice (at some point before), which implies that the event $D$ (of Lemma 2.7) holds, and so $I_q \subseteq D$. Therefore, by Lemma 2.7 $\Pr[I_q] \leq \Pr[D] \leq \lambda/\lambda_1 = \Pr[F_q]/\gamma$ proving Inequality (5). □

**Remark 3.22** (Sufficient Condition for Efficiency of Eve)**.** The proof of Claims 3.19 and 3.21 only depend on the fact that all the queries asked by Eve are are either $(\varepsilon/n_B)$-heavy for Alice or $(\varepsilon/n_A)$-heavy for Bob with respect to the distribution $\mathcal{GV}(M, P)$. More formally, all we need is that whenever Eve asks a query $q$ it holds that

$$\Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[q \in \mathcal{Q}(V_A)] \geq \frac{\varepsilon}{n_B} \quad \text{or} \quad \Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[q \in \mathcal{Q}(V_B)] \geq \frac{\varepsilon}{n_A}.$$

In particular, the conclusions of Claims 3.19 and 3.21 hold regardless of which heavy queries Eve chooses to ask at any moment, and the only important thing is that all the queries asked by Eve were heavy at the time of being asked.

# 4 Extensions

In this section we prove several extensions to our main result that can all be directly obtained from the results proved in Section 3. The main goal of this section is to generalize our main result to a broader setting so that it could be applied in subsequent work more easily. We assume the reader is familiar with the definitions given in Sections 2 and 3.

## 4.1 Making the Views Almost Independent

In this section we will prove Theorem 1.3 along with several other extensions. These extensions were used in [DSLMM11] to prove black-box separations for certain optimally-fair coin-tossing protocols. We first mention these extensions informally and then will prove them formally.

**Average Number of Queries:** We will show how to decrease the number of queries asked by Eve by a factor of $\Omega(\varepsilon)$ if we settle for bounding the *average* number of queries asked by Eve. This can always be turned into a an attack of worst-case complexity by putting the $\Theta(\varepsilon)$ multiplicative factor back and applying the Markov inequality.

**Changing the Heaviness Threshold:** We will show that the attacker Eve of Construction 3.3 is "robust" with respect to choosing its "heaviness" parameter $\varepsilon$. Namely, if she changes the parameter $\varepsilon$ arbitrarily during her attack, as long as $\varepsilon \in [\varepsilon_1, \varepsilon_2]$ for some $\varepsilon_1 < \varepsilon_2$, we can still show that Eve is both "successful" and "efficient" with high probability.

**Learning the Dependencies:** We will show that our adversary Eve can, with high probability, learn the "dependency" between the views of Alice and Bob in any two-party computation. Dachman et al. [DSLMM11] were the first to point out that such results can be obtained from results proved in original publication of this work [BMG09]. Haitner et al. [HOZ13], relying some of the results proved in [BMG09], proved a variant of the first part of our Theorem 1.3 in which $n$ bounds *both* of $n_A$ and $n_B$.

**Lightness of Queries:** We observe that with high probability the following holds at the end of every round conditioned on Eve's view: For every query $q$ *not* learned by Eve, the probability of $q$ being asked by Alice or Bob remains "small". Note that here we are *not* conditioning on the event $\mathsf{Good}(M, P)$.

Now we formally prove the above extensions.

The following definition defines a *class* of attacks that share a specific set of properties.

**Definition 4.1.** For $\varepsilon_1 \leq \varepsilon_2$, we call Eve an $(\varepsilon_1, \varepsilon_2)$-attacker, if Eve performs her attack in the framework of Construction 3.3, but instead of using a single parameter $\varepsilon$ it uses $\varepsilon_1 \leq \varepsilon_2$ as follows.

1. **All queries asked are heavy according to parameter $\varepsilon_1$.** Every query $q$ asked by Eve, at the time of being asked, should be either $(\varepsilon_1/n_B)$-heavy for Alice or $(\varepsilon_1/n_A)$-heavy for Bob with respect to the distribution $\mathcal{GV}(M, P)$ where $(M, P)$ is the view of Eve when asking $q$.

2. **No heavy query, as parameterized by $\varepsilon_2$, remains unlearned.** At the end of every round $i$, if $(M, P)$ is the view of Eve at that moment, and if $q$ is any query that is either $(\varepsilon_2/n_B)$-heavy for Alice or $(\varepsilon_2/n_A)$-heavy for Bob with respect to the distribution $\mathcal{GV}(M, P)$, then Eave has to have learned that query already to make sure $q \in \mathcal{Q}(P)$.

**Comparison with Eve of Construction 3.3.** The Eve of Construction 3.3 is an $(\varepsilon,\varepsilon)$-attacker, but for $\varepsilon_1 < \varepsilon_2$ the class of $(\varepsilon_1,\varepsilon_2)$-attackers include algorithms that could not necessarily be described by Construction 3.3. For example, an $(\varepsilon_1,\varepsilon_2)$-attackers can chose any $\varepsilon \in [\varepsilon_1,\varepsilon_2]$ and run the attacker of Construction 3.3 using parameter $\varepsilon$, or it can even keep changing its parameter $\varepsilon \in [\varepsilon_1,\varepsilon_2]$ *along the execution* of the attack. In addition, the attacker of Construction 3.3 needs to choose the *lexicographically first* heavy query, while an $(\varepsilon_1,\varepsilon_2)$-attacker has the freedom of choosing *any* query so long as it is $(\varepsilon_1/n_B)$-heavy for Alice or $(\varepsilon_1/n_A)$-heavy for Bob. Finally, an $(\varepsilon_1,\varepsilon_2)$-attacker could use its own randomness $r_E$ that affects its choice of queries, as long as it respects the two conditions of Definition 4.1.

**Definition 4.2** (Self Dependency)**.** For every joint distribution $(\mathbf{x},\mathbf{y})$, we call $\mathsf{SelfDep}(\mathbf{x},\mathbf{y}) = \Delta((\mathbf{x},\mathbf{y}),(\mathbf{x}\times\mathbf{y}))$ the *self (statistical) dependency* of a $(\mathbf{x},\mathbf{y})$ where in $(\mathbf{x}\times\mathbf{y})$ we sample $\mathbf{x}$ and $\mathbf{y}$ independently from their marginal distributions.

The following theorem formalizes Theorem 1.3. The last part of the theorem is used by [DSLMM11] to prove lower-bounds on coin tossing protocols from one-way functions. We advise the reader to review the notations of Section 3.1 as we will use some of them here for our modified variant of $(\varepsilon_1,\varepsilon_2)$-attackers.

**Theorem 4.3** (Extensions to Main Theorem)**.** *Let,* $\Pi, r_A, n_A, r_B, n_B, H, s_A, s_B, \rho$ *be as in Theorem 3.1 and suppose* $\varepsilon_1 \le \varepsilon_2 < 1/10$*. Let Eve be* any $(\varepsilon_1,\varepsilon_2)$*-attacker who is modified to stop asking any queries as soon as she is about to ask a red query (as defined in Definition 3.18). Then the following claims hold.*

1. **Finding outputs:** *Eve's output agrees with Bob's output with probability* $\rho - 16\varepsilon_2$*.*
2. **Average number of queries:** *The expected number of queries asked by Eve is at most* $4n_An_B/\varepsilon_1$*. More generally, if we let* $Q_\varepsilon$ *to be the number of (green) queries that are asked because of being* $\varepsilon$*-heavy for a fixed* $\varepsilon \in [\varepsilon_1,\varepsilon_2]$*, it holds that* $\mathbb{E}[|Q_\varepsilon|] \le 4n_An_B/\varepsilon$*.*
3. **Self-dependency at every fixed round.** *For any fixed round* $i$*, it holds that*
$$\mathop{\mathbb{E}}_{(M,P)\leftarrow(\mathbf{M}^i,\mathbf{P}^i_E)}[\mathsf{SelfDep}(\mathcal{V}(M,P))] \le 21\cdot\varepsilon_2.$$
4. **Simultaneous self-dependencies at all rounds.** *For every* $\alpha,\beta$ *such that* $0<\alpha<1$*,* $0<\beta<1$*, and* $\alpha\cdot\beta \ge \varepsilon_2$*, with probability at least* $1-9\alpha$ *the following holds: at the end of* every *round* $i$*, we have* $\mathsf{SelfDep}(\mathcal{V}(M^i,P^i_E)) \le 9\beta$*.*
5. **Simultaneous lightness at all round.** *For every* $\alpha,\beta$ *such that* $0<\alpha<1$*,* $0<\beta<1$*, and* $\alpha\cdot\beta \ge \varepsilon_2$*, with probability at least* $1-9\alpha$ *the following holds: at the end of* every *round, if* $q \notin \mathcal{Q}(P)$ *is any query not learned by Eve so far we have*
$$\Pr_{(V_A,V_B)\leftarrow\mathcal{V}(M,P)}[q\in\mathcal{Q}(V_A)] < \frac{\varepsilon_2}{n_B}+\beta \quad \text{and} \quad \Pr_{(V_A,V_B)\leftarrow\mathcal{V}(M,P)}[q\in\mathcal{Q}(V_B)] < \frac{\varepsilon_2}{n_A}+\beta.$$
6. **Dependency and lightness at every fixed round.** *For every round* $i$ *and every* $(M,P)\leftarrow(\mathbf{M}^i,\mathbf{P}^i_E)$ *there is a product distribution* $(\mathbf{W}_A\times\mathbf{W}_B)$ *such that the following two hold:*
   (a) $\mathbb{E}_{(M,P)}[\Delta(\mathcal{V}(M,P),(\mathbf{W}_A\times\mathbf{W}_B))] \le 15\varepsilon_2$*.*

*(b) With probability* $1 - 6\varepsilon_2$ *over the choice of* $(M, P)$ *(which determines the distributions* $\mathbf{W}_A, \mathbf{W}_B$ *as well), we have* $\Pr[q \in \mathcal{Q}(\mathbf{W}_A)] < \frac{\varepsilon_2}{n_B}$ *and* $\Pr[q \in \mathcal{Q}(\mathbf{W}_B)] < \frac{\varepsilon_2}{n_A}$.

In the rest of this section we prove Theorem 4.3. To prove all the properties, we first assume that the adversary is an $(\varepsilon_1, \varepsilon_2)$-attacker, denoted by UnbEve (Unbounded Eve), and then will analyze how stopping UnbEve upon reaching a red query (i.e., converting it into Eve) will affect her execution.

Remarks 3.13 and 3.22 show that many of the results proved in the previous section extend to the more general setting of $(\varepsilon_1, \varepsilon_2)$-attackers.

**Claim 4.4.** *All the following lemmas, claims, and corollaries still hold when we use an arbitrary* $(\varepsilon_1, \varepsilon_2)$*-attacker and* $\varepsilon_1 < \varepsilon_2 < 1/10$*:*

1. *Lemma 3.8 using* $\varepsilon = \varepsilon_2$.
2. *Corollary 3.14 using* $\varepsilon = \varepsilon_2$.
3. *Lemma 3.6 using* $\varepsilon = \varepsilon_2$.
4. *Lemma 3.4 using* $\varepsilon = \varepsilon_2$.
5. *Claim 3.20 using* $\varepsilon = \varepsilon_2$.
6. *Claim 3.19 by using* $\varepsilon = \varepsilon_1$ *in the definition of green queries.*
7. *Claim 3.21 by using* $\varepsilon = \varepsilon_1$ *in the definition of green queries. More generally, the proof of Claim 3.21 works directly (without any change) if we run a* $(\varepsilon_1, \varepsilon_2)$ *attack, but define the green queries using a parameter* $\varepsilon \in [\varepsilon_1, \varepsilon_2]$ *(and only count such queries, as green ones).*

*Proof.* Item 1 follows from Remark 3.13 and the the second property of $(\varepsilon_1, \varepsilon_2)$-attackers. All Items 2–5 follow from Item 1 because the proofs of the corresponding statements in previous section *only* rely (directly or indirectly) on Lemma 3.8.

Items 6 and 7 follow from Remark 3.22 and the first property of $(\varepsilon_1, \varepsilon_2)$-attackers. □

**Finding Outputs.** By Item 4 of Claim 4.4, UnbEve hits Bob's output with probability at least $\rho - 10\varepsilon_2$. By Item 5 of Claim 4.4, the probability that UnbEve asks any red queries is at most $6\varepsilon_2$. Therefore, Eve's output will agree with Bob's output with probability at least $\rho - 10\varepsilon - 6\varepsilon = \rho - 16\varepsilon$.

**Number of Queries.** By Item 7, the expected number of green queries asked by UnbEve is at most $4n_A n_B/\varepsilon_1$. As also specified in Item 7, the more general upper bound, for an arbitrary parameter $\varepsilon \in [\varepsilon_1, \varepsilon_2]$, holds as well.

**Dependencies.** We will use the following definition which relaxes the notion of self dependency by computing the statistical distance of $(\mathbf{x}, \mathbf{y})$ to the closest product distribution (that might be different from $(\mathbf{x} \times \mathbf{y})$).

**Definition 4.5** (Statistical Dependency)**.** For two jointly distributed random variables $(\mathbf{x}, \mathbf{y})$, let the *statistical dependency* of $(\mathbf{x}, \mathbf{y})$, denoted by $\mathsf{StatDep}(\mathbf{x}, \mathbf{y})$, be the minimum statistical distance of $(\mathbf{x}, \mathbf{y})$ from all product distributions defined over $\mathrm{Supp}(\mathbf{x}) \times \mathrm{Supp}(\mathbf{y})$. More formally:

$$\mathsf{StatDep}(\mathbf{x}, \mathbf{y}) = \inf_{(\mathbf{a} \times \mathbf{b})} \Delta((\mathbf{x}, \mathbf{y}), (\mathbf{a} \times \mathbf{b}))$$

in which $\mathbf{a} \times \mathbf{b}$ are distributed over $\mathrm{Supp}(\mathbf{x}) \times \mathrm{Supp}(\mathbf{y})$.

By definition, we have $\mathsf{StatDep}(\mathbf{x}, \mathbf{y}) \leq \mathsf{SelfDep}(\mathbf{x}, \mathbf{y})$. The following lemma by [MMP14] shows that the two quantities can not be too far.

**Lemma 4.6** (Lemma A.6 in [MMP14])**.** $\mathsf{SelfDep}(\mathbf{x}, \mathbf{y}) \leq 3 \cdot \mathsf{StatDep}(\mathbf{x}, \mathbf{y})$.

**Remark 4.7.** We note that, $\mathsf{SelfDep}(\mathbf{x}, \mathbf{y})$ can, in general, be larger than $\mathsf{StatDep}(\mathbf{x}, \mathbf{y})$. For instance consider the following joint distribution over $(\mathbf{x}, \mathbf{y})$ where $\mathbf{x}$ and $\mathbf{y}$ are both Boolean variables: $\Pr[\mathbf{x} = 0, \mathbf{y} = 0] = 1/3, \Pr[\mathbf{x} = 1, \mathbf{y} = 0] = 1/3, \Pr[\mathbf{x} = 1, \mathbf{y} = 1] = 1/3, \Pr[x = 0, y = 1] = 0$. It is easy to see that $\mathsf{SelfDep}(\mathbf{x}, \mathbf{y}) = 2/9$, but $\Delta((\mathbf{x}, \mathbf{y}), (\mathbf{a} \times \mathbf{b})) = 1/6 < 2/9$ for a product distribution $(\mathbf{a} \times \mathbf{b})$ defined as follows: $\mathbf{a} \equiv \mathbf{x}$ and $\Pr[\mathbf{b} = 0] = \Pr[\mathbf{b} = 1] = 1/2$.

The following lemma follows from Lemma 2.13 and the definition of statistical dependency.

**Lemma 4.8.** *For jointly distributed* $(\mathbf{x}, \mathbf{y})$ *and event* $E$ *defined over the support of* $(\mathbf{x}, \mathbf{y})$*, it holds that* $\mathsf{StatDep}(\mathbf{x}, \mathbf{y}) \leq \Pr_{(\mathbf{x},\mathbf{y})}[E] + \mathsf{StatDep}((\mathbf{x}, \mathbf{y}) \mid \neg E)$*. We take the notational convention that whenever* $\Pr_{(\mathbf{x},\mathbf{y})}[E] = 0$ *we let* $\mathsf{StatDep}((\mathbf{x}, \mathbf{y}) \mid \neg E) = 1$.

*Proof.* Let $(\mathbf{a} \times \mathbf{b})$ be such that $\Delta(((\mathbf{x}, \mathbf{y}) \mid \neg E), (\mathbf{a} \times \mathbf{b})) \leq \delta$. For the same $(\mathbf{a} \times \mathbf{b})$, by Lemma 2.13 it holds that $\Delta((\mathbf{x}, \mathbf{y}), (\mathbf{a} \times \mathbf{b})) \leq \Pr_{(\mathbf{x},\mathbf{y})}[E] + \delta$. Therefore

$$\begin{aligned}\mathsf{StatDep}(\mathbf{x}, \mathbf{y}) = \inf_{(\mathbf{a} \times \mathbf{b})} \Delta((\mathbf{x}, \mathbf{y}), (\mathbf{a} \times \mathbf{b})) &\leq \Pr_{(\mathbf{x},\mathbf{y})}[E] + \inf_{(\mathbf{a} \times \mathbf{b})} \Delta(((\mathbf{x}, \mathbf{y}) \mid \neg E), (\mathbf{a} \times \mathbf{b})) \\ &\leq \Pr_{(\mathbf{x},\mathbf{y})}[E] + \mathsf{StatDep}((\mathbf{x}, \mathbf{y}) \mid \neg E).\end{aligned}$$

□

**Self-dependency at every fixed round.** By Item 2 of Claim 4.4, we get that by running UnbEve we obtain $\mathsf{StatDep}(\mathcal{GV}(M, P)) \leq 2\varepsilon_2$ where $(M, P)$ is the view of UnbEve at the end of the protocol. By also Lemma 4.8 we get:

$$\begin{aligned}\mathsf{StatDep}(\mathcal{V}(M, P)) &\leq \Pr_{\mathcal{E}}[\neg\mathsf{Good} \mid (M, P)] + \mathsf{StatDep}(\mathcal{GV}(M, P)) \\ &\leq \Pr_{\mathcal{E}}[\neg\mathsf{Good} \mid (M, P)] + 2\varepsilon_2.\end{aligned}$$

Therefore, by Item 3 of Claim 4.4 and Lemma 4.6 we get

$$\begin{aligned}\mathop{\mathbb{E}}_{(M,P) \leftarrow (\mathbf{M},\mathbf{P})}[\mathsf{StatDep}(\mathcal{V}(M, P))] &\leq 3 \cdot \left( \mathop{\mathbb{E}}_{(M,P) \leftarrow (\mathbf{M},\mathbf{P})} [\mathsf{StatDep}(\mathcal{V}(M, P))] \right) \\ &\leq 3 \cdot \left( \mathop{\mathbb{E}}_{(M,P) \leftarrow (\mathbf{M},\mathbf{P})} \left[ \Pr_{\mathcal{E}}[\neg\mathsf{Good} \mid (M, P)] \right] + 2\varepsilon_2 \right) \\ &\leq 3 \cdot \left( \Pr_{\mathcal{E}}[\neg\mathsf{Good}] + 2\varepsilon_2 \right) \leq 3 \cdot 5\varepsilon_2 = 15\varepsilon_2\end{aligned}$$

Since the probability of UnbEve asking any red queries is at most $6\varepsilon_2$ (Item 5 of Claim 4.4), therefore when we run Eve, it holds that $\mathbb{E}_{(M,P)\leftarrow(\mathbf{M},\mathbf{P})}[\mathsf{StatDep}(\mathcal{V}(M,P))]$ increases at most by $6\varepsilon_2$ compared to when running UnvEve. This is because whenever we halt the execution of Eve (which happens with probability at most $6\varepsilon_2$) this can lead to statistical dependency of $\mathcal{V}(M,P)$ at most 1. Therefore, if we use Eve instead of UnbEve, it holds that

$$\mathop{\mathbb{E}}_{(M,P)\leftarrow(\mathbf{M},\mathbf{P})}[\mathsf{StatDep}(\mathcal{V}(M,P))] \leq 15\varepsilon_2 + 6\varepsilon_2 = 21\varepsilon_2.$$

**Simultaneous self-dependencies at all rounds.** First note that $0 < \alpha < 1$, $0 < \beta < 1$, and $\alpha \cdot \beta \geq \varepsilon_2$ imply that $\alpha \geq \varepsilon_2$ and $\beta \geq \varepsilon_2$. By Item 3 of Claim 4.4, when we run UnbEve, it holds that $\Pr_{\mathcal{E}}[\mathsf{Fail}] \leq 3\varepsilon_2$, so by Lemma 2.7 we conclude that with probability at least $1-3\alpha$ it holds that during the execution of the protocol, the probability of $\mathsf{Fail}$ (and thus, the probability of $\neg\mathsf{Good}(M,P)$) conditioned on Eve's view always remains at most $\beta$. Therefore, by Item 2 of Claim 4.4 and Lemma 4.8, with probability at least $1-3\alpha$ the following holds at the end of *every* round (where $(M,P)$ is Eve's view at the end of that round)

$$\begin{aligned}\mathsf{StatDep}(\mathcal{V}(M,P)) &\leq \Pr_{\mathcal{E}}[\neg\mathsf{Good} \mid (M,P)] + \mathsf{StatDep}(\mathcal{GV}(M,P)) \\ &\leq \beta + 2\varepsilon_2 \leq 3\beta.\end{aligned}$$

Using Lemma 4.6 we obtain the bound $\mathsf{SelfDep}(\mathcal{V}(M,P)) \leq 9\beta$. Since the probability of UnbEve asking any red queries is at most $6\varepsilon_2$, by a union bound we conclude that with probability at least $1-3\alpha-6\varepsilon_2 > 1-9\alpha$, we still get $\mathsf{SelfDep}(\mathcal{V}(M,P)) \leq 9\beta$ at the end of every round.

**Simultaneous lightness at all rounds.** As shown in the previous item, for such $\alpha, \beta$, with probability at least $1-9\alpha$ it holds that during the execution of the protocol, the probability of $\mathsf{Fail}$ (and thus, the probability of $\neg\mathsf{Good}(M,P)$) conditioned on Eve's view always remains at most $\beta$. Now suppose $(M,P)$ be the view of Eve at the end of some round where $\Pr_{\mathcal{V}(M,P)}[\neg\mathsf{Good}(M,P)] \leq \beta$. By the second property of $(\varepsilon_1,\varepsilon_2)$-attackers, it holds that:

$$\Pr_{(V_A,V_B)\leftarrow\mathcal{V}(M,P)}[q \in \mathcal{Q}(V_A)] \leq \Pr_{\mathcal{V}(M,P)}[\neg\mathsf{Good}(M,P)] + \Pr_{(V_A,V_B)\leftarrow\mathcal{GV}(M,P)}[q \in \mathcal{Q}(V_A)] \leq \varepsilon_2/n_B + \beta.$$

The same proof shows that a similar statement holds for Bob.

**Dependency and lightness at every fixed round.** Let $(\mathbf{W}_A, \mathbf{W}_B) \equiv \mathcal{GV}(M,P)$. The product distribution we are looking for will be $\mathbf{W}_A \times \mathbf{W}_B$. When we run UnbEve, by Lemma 3.6 it holds that $\mathbb{E}_{(M,P)}[\Delta((\mathbf{W}_A,\mathbf{W}_B),\mathcal{V}(M,P))] \leq 3\varepsilon_2$, because otherwise the probability of $\mathsf{Fail}$ will be more than $3\varepsilon_2$. Also, by Corollary 3.14 it holds that $\mathsf{StatDep}(\mathcal{V}(M,P)) \leq 2\varepsilon_2$, and by Lemma 4.6, it holds that $\mathsf{SelfDep}(\mathcal{V}(M,P)) = \Delta(\mathcal{V}(M,P),(\mathbf{W}_A \times \mathbf{W}_B)) \leq 6\varepsilon_2$. Thus, when we run UnbEve, we get $\mathbb{E}_{(M,P)}[\Delta((\mathbf{W}_A \times \mathbf{W}_B),\mathcal{V}(M,P))] \leq 9\varepsilon_2$. By Claim 3.20, the upper bound of $9\varepsilon_2$ when we modify UnbEve to Eve (by not asking red queries), could increase only by $6\varepsilon_2$. This proves the first part.

To prove the second part, again we use Claim 3.20 which bounds the probability of asking a red query by $6\varepsilon_2$. Also, as long as we do not halt Eve (i.e., no red query is asked), Eve and UnbEve remain the same, and the lightness claims hold for UnbEve by definition of the attacker UnbEve.

### 4.2 Removing the Rationality Condition

In this subsection we show that *all* the results of this paper, except the graph characterization of Lemma 3.8, hold even with respect to random oracles that are not necessarily rational according to Definition 2.2. We will show that a variant of Lemma 3.8, which is sufficient for all of our applications, still holds. In the following, by an *irrational random oracle* we refer to a random oracle that satisfies Definition 2.2 except that its probabilities might not be rational.

**Lemma 4.9** (Characterization of $\mathcal{V}(M,P)$)**.** *Let $H$ be an irrational oracle, let $M$ be the sequence of messages sent between Alice and Bob so far, and let $P$ be the set of oracle query-answer pairs known to Eve (who uses parameter $\varepsilon$) by the end of the round in which the last message in $M$ is sent. Also suppose $\Pr_{\mathcal{V}(M,P)}[\mathsf{Good}(M,P)] > 0$. Let $(\mathbf{V}_A, \mathbf{V}_B)$ be the joint view of Alice and Bob as sampled from $\mathcal{GV}(M,P)$, and let $\mathcal{U}_A = \mathrm{Supp}(\mathbf{V}_A), \mathcal{U}_B = \mathrm{Supp}(\mathbf{V}_B)$. Let $G = (\mathcal{U}_A, \mathcal{U}_B, E)$ be a bipartite graph with vertex sets $\mathcal{U}_A, \mathcal{U}_B$ and connect $u_A \in \mathcal{U}_A$ to $u_B \in \mathcal{U}_B$ if and only if $\mathcal{Q}(u_A) \cap \mathcal{Q}(u_B) \subseteq \mathcal{Q}(P)$. Then there exists a distribution $\mathbf{U}_A$ over $\mathcal{U}_A$ and a distribution $\mathbf{U}_B$ over $\mathcal{U}_B$ such that:*

1. *For every vertex $u \in \mathcal{U}_A$, it holds that $\Pr_{v \leftarrow \mathbf{U}_B}[u \not\sim v] \le 2\varepsilon$, and similarly for every vertex $u \in \mathcal{U}_B$, it holds that $\Pr_{v \leftarrow \mathbf{U}_A}[u \not\sim v] \le 2\varepsilon$.*
2. *The distribution $(V_A, V_B) \leftarrow \mathcal{GV}(M,P)$ is identical to: sampling $u \leftarrow \mathbf{U}_A$ and $v \leftarrow \mathbf{U}_B$ conditioned on $u \sim v$, and outputting the views corresponding to $u$ and $v$.*

*Proof Sketch.* The distributions $\mathbf{U}_A$ and $\mathbf{U}_B$ are in fact the same as the distributions $\mathbf{A}$ and $\mathbf{B}$ of Lemma 3.9. The rest of the proof is identical to that of Lemma 3.8 *without any* vertex repetition. In fact, repetition of vertices (to make the distributions uniform) cannot be necessarily done anymore because of the irrationality of the probabilities. Here we explain the alternative parameter that takes the role of $|E^{\not\sim}(u)|/|E|$. For $u \in \mathcal{U}_A$ let $q^{\not\sim}(u)$ be the probability that if we sample an edge $e \leftarrow (\mathbf{V}_A, \mathbf{V}_B)$, it does not contain $u$ as Alice's view, and define $q^{\not\sim}(u)$ for $u \in \mathcal{U}_B$ similarly. It can be verified that by the very same argument as in Lemma 3.8, it holds that $q^{\not\sim}(u) \le \varepsilon$ for every vertex $u$ in $G$. The other steps of the proof remain the same. □

The characterization of $\mathcal{V}(M,P)$ by Lemma 4.9 can be used to derive Corollary 3.14 directly (using the same distributions $\mathbf{U}_A$ and $\mathbf{U}_B$). Remark 3.13 also holds with respect to Lemma 4.9. Here we show how to derive Lemma 3.7 and the rest of the results will follow immediately.

**Proving Lemma 3.7.** Again, we prove Lemma 3.7 even conditioned on choosing any vertex $v$ that describes Bob's view. For such vertex $v$, the distribution of Alice's view, when we choose a random edge $(u, v') \leftarrow (\mathbf{V}_A, \mathbf{V}_B)$ conditioned on $v = v'$ is the same as choosing $u \leftarrow \mathbf{U}_A$ conditioned on $u \sim v$. Let's call this distribution $\mathbf{U}_A^v$. Let $S = \{u \in \mathcal{U}_A \mid q \in A_u\}$ where $q$ is the next query of Bob as specified by $v$. Let $p(S) = \sum_{u \in S} \Pr[\mathbf{U}_A = u], q(S) = \Pr_{(u,v) \leftarrow (\mathbf{V}_A, \mathbf{V}_B)}[u \in S]$, and let $p(E) = \Pr_{u \leftarrow \mathbf{U}_A, v \leftarrow \mathbf{U}_B}[u \sim v]$. Also let $p^{\sim}(v) = \sum_{u \sim v} \Pr[\mathbf{U}_A = u]$. Then, we have:

$$\Pr_{u \leftarrow \mathbf{U}_A^v}[q \in A_u] \le \frac{p(S)}{p^{\sim}(v)} \le \frac{p(S)}{1-2\varepsilon} \le \frac{p(S)}{(1-2\varepsilon)\cdot p(E)} \le \frac{q(S)}{(1-2\varepsilon)^2 \cdot p(E)} \le \frac{\varepsilon}{(1-2\varepsilon)^2 \cdot n_B} < \frac{3\varepsilon}{2n_B}.$$

The second and fourth inequalities are due to the degree lower bounds of Item 1 in Lemma 4.9. The third inequality is because $p(E) < 1$. The fifth inequality is because of the definition of the attacker Eve who asks $\varepsilon/n_B$ heavy queries for Alice's view when sampled from $\mathcal{GV}(M,P)$, as long as such queries exist. The sixth inequality is because we are assuming $\varepsilon < 1/10$. □

**Acknowledgement.** We thank Russell Impagliazzo for very useful discussions and the anonymous reviewers for their valuable comments.

## References

[BBE92] Charles H. Bennett, Gilles Brassard, and Artur K. Ekert, *Quantum cryptography*, Scientific American **267** (1992), no. 4, 50–57.

[BGI08] Eli Biham, Yaron J. Goren, and Yuval Ishai, *Basing weak public-key cryptography on strong one-way functions*, TCC (Ran Canetti, ed.), Lecture Notes in Computer Science, vol. 4948, Springer, 2008, pp. 55–72.

[BHK+11] Gilles Brassard, Peter Høyer, Kassem Kalach, Marc Kaplan, Sophie Laplante, and Louis Salvail, *Merkle puzzles in a quantum world*, CRYPTO (Phillip Rogaway, ed.), Lecture Notes in Computer Science, vol. 6841, Springer, 2011, pp. 391–410.

[BKSY11] Zvika Brakerski, Jonathan Katz, Gil Segev, and Arkady Yerukhimovich, *Limits on the power of zero-knowledge proofs in cryptographic constructions*, TCC (Yuval Ishai, ed.), Lecture Notes in Computer Science, vol. 6597, Springer, 2011, pp. 559–578.

[BMG09] Boaz Barak and Mohammad Mahmoody-Ghidary, *Merkle puzzles are optimal - an O ( n 2 )-query attack on any key exchange from a random oracle*, CRYPTO (Shai Halevi, ed.), Lecture Notes in Computer Science, vol. 5677, Springer, 2009, pp. 374–390.

[BR93] Mihir Bellare and Phillip Rogaway, *Random oracles are practical: A paradigm for designing efficient protocols*, ACM Conference on Computer and Communications Security, 1993, pp. 62–73.

[BS08] Gilles Brassard and Louis Salvail, *Quantum merkle puzzles*, International Conference on Quantum, Nano and Micro Technologies (ICQNM), IEEE Computer Society, 2008, pp. 76–79.

[CGH04] Canetti, Goldreich, and Halevi, *The random oracle methodology, revisited*, JACM: Journal of the ACM **51** (2004), no. 4, 557–594.

[Cle86] Richard Cleve, *Limits on the security of coin flips when half the processors are faulty (extended abstract)*, Annual ACM Symposium on Theory of Computing (Berkeley, California), 28–30 May 1986, pp. 364–369.

[DH76] Whitfield Diffie and Martin Hellman, *New directions in cryptography*, IEEE Transactions on Information Theory **IT-22** (1976), no. 6, 644–654.

[DSLMM11] Dana Dachman-Soled, Yehuda Lindell, Mohammad Mahmoody, and Tal Malkin, *On the black-box complexity of optimally-fair coin tossing*, TCC (Yuval Ishai, ed.), Lecture Notes in Computer Science, vol. 6597, Springer, 2011, pp. 450–467.

[GGKT05] Rosario Gennaro, Yael Gertner, Jonathan Katz, and Luca Trevisan, *Bounds on the efficiency of generic cryptographic constructions*, SIAM journal on Computing **35** (2005), no. 1, 217–246.

[Gro96] Lov K. Grover, *A fast quantum mechanical algorithm for database search*, Annual ACM Symposium on Theory of Computing (STOC), 22–24 May 1996, pp. 212–219.

[HHRS07] Iftach Haitner, Jonathan J. Hoch, Omer Reingold, and Gil Segev, *Finding collisions in interactive protocols – A tight lower bound on the round complexity of statistically-hiding commitments*, Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE, 2007, pp. 669–679.

[Hol15] Thomas Holenstein, *Complexity theory*, 2015, http://www.complexity.ethz.ch/education/Lectures/ComplexityFS15/skript_printable.pdf.

[HOZ13] Iftach Haitner, Eran Omri, and Hila Zarosim, *Limits on the usefulness of random oracles*, Theory of Cryptography, TCC (Amit Sahai, ed.), Lecture Notes in Computer Science, vol. 7785, Springer, 2013, pp. 437–456.

[IR89] Russell Impagliazzo and Steven Rudich, *Limits on the provable consequences of one-way permutations*, Annual ACM Symposium on Theory of Computing (STOC), 1989, Full version available from Russell Impagliazzo's home page https://cseweb.ucsd.edu/~russell/secret.ps, pp. 44–61.

[KSY11] Jonathan Katz, Dominique Schröder, and Arkady Yerukhimovich, *Impossibility of blind signatures from one-way permutations*, TCC (Yuval Ishai, ed.), Lecture Notes in Computer Science, vol. 6597, Springer, 2011, pp. 615–629.

[Mer74] Ralph C. Merkle, *C.S. 244 project proposal*, http://merkle.com/1974/, 1974.

[Mer78] Ralph C. Merkle, *Secure communications over insecure channels*, Communications of the ACM **21** (1978), no. 4, 294–299.

[MMP14] Mohammad Mahmoody, Hemanta K Maji, and Manoj Prabhakaran, *Limits of random oracles in secure computation*, Proceedings of the 5th conference on Innovations in theoretical computer science, ACM, 2014, pp. 23–34.

[MMV11] Mohammad Mahmoody, Tal Moran, and Salil P. Vadhan, *Time-lock puzzles in the random oracle model*, CRYPTO (Phillip Rogaway, ed.), Lecture Notes in Computer Science, vol. 6841, Springer, 2011, pp. 39–50.

[MP12] Mohammad Mahmoody and Rafael Pass, *The curious case of non-interactive commitments - on the power of black-box vs. non-black-box use of primitives*, CRYPTO (Reihaneh Safavi-Naini and Ran Canetti, eds.), Lecture Notes in Computer Science, vol. 7417, Springer, 2012, pp. 701–718.

[RSA78] Ronald L. Rivest, Adi Shamir, and Leonard M. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Communications of the ACM **21** (1978), no. 2, 120–126.

[RTV04] Omer Reingold, Luca Trevisan, and Salil P. Vadhan, *Notions of reducibility between cryptographic primitives*, TCC (Moni Naor, ed.), Lecture Notes in Computer Science, vol. 2951, Springer, 2004, pp. 1–20.