AICR · OPEN PROBLEM

Classical Attacks on Classical-Alice Quantum-Bob Key Agreement Without Perfect Completeness

The one case the Simulation Conjecture barrier does not cover

Status. Statement: AI-written from Per Austrin, Hao Chung, Kai-Min Chung, Shiuan Fu, Yao-Ting Lin, Mohammad Mahmoody, *On the Impossibility of Key Agreements from Quantum Random Oracles*, Cryptology ePrint Archive 2022, not yet checked by a human. Perfect completeness is settled by the source (Theorem 3.1: an expected $d_A d_B/\lambda$ classical queries, success probability $1-\lambda$), and imperfect completeness — the case asserted here — is explicitly left open. For two quantum parties with imperfect completeness the paper proves a barrier (Theorem 7.6): a classical polynomial-query attack there would imply the Simulation Conjecture. That barrier's protocol (Protocol 7.7) has both parties run the hard-to-simulate quantum algorithm, so it yields no classical-Alice counterexample; the paper does not address whether some other barrier applies to the classical-Alice case.

Category. *Idealized Models & Non-Uniformity.*

Conjecture (informal). *The paper proves that if Alice is an ordinary classical party, Bob has a quantum computer with superposition access to a hash function modelled as a random oracle, they talk over a classical channel, and they always end up with the same key, then a purely classical eavesdropper who sees the transcript can find that key with a number of oracle queries roughly the product of the two parties' query counts. The proof leans hard on* always*: it works by showing that some oracle is consistent both with a fake Alice view the eavesdropper samples and with Bob's real state, and then invoking the fact that any such oracle forces the two keys to agree. If the honest parties are allowed to fail with even negligible probability, this last step gives nothing, and the authors say so and leave the question open. It is not an idle gap. The same paper shows that for two quantum parties, a classical attack in the imperfect-completeness setting would resolve a long-standing folklore conjecture about simulating quantum query algorithms classically, so an unconditional proof there is out of reach for now. But that barrier is built from a protocol in which both parties run a quantum algorithm whose acceptance probability no classical algorithm can estimate, so it says nothing when Alice is classical. This conjecture asks for the unconditional statement in exactly that case.*

1 The setting

Merkle's protocol [Mer74] gives a quadratic query gap between honest parties and an eavesdropper from a random oracle alone, and in the fully classical random oracle model that gap is known to be optimal [IR89, BM17]. The source paper carries the upper bound into the quantum world for the unbalanced case: Theorem 3.1 shows that if Alice is classical, Bob is quantum, they communicate classically, they use a random oracle h, and they *always* agree on a key, then a classical eavesdropper who is given the transcript finds the key with probability $1 - \lambda$ using $d_A \cdot d_B/\lambda$ classical queries. This models the realistic asymmetric setting of a quantum-capable server talking to a classical client, and it is unconditional — unlike the paper's other attack, which rests on the Polynomial Compatibility Conjecture.

The proof has three ingredients: the classical heavy-query learner

for Alice of [BM17, BKSY11], which needs no quantum generalisation here because Alice is classical; an independence lemma (Lemma 3.2) saying that, conditioned on the transcript and the oracle, the joint state of the two parties is a product state; and a consistency lemma (Lemma 3.3) saying that if the eavesdropper's learned partial oracle misses the maximal Fourier monomial of the joint state, some full oracle is consistent with both her sampled fake Alice view and the real execution. Perfect completeness is what turns that consistency into a win: the mere existence of one oracle consistent with the fake Alice view and Bob's registers forces the fake key to equal Bob's key. With a completeness error, however small, there can be consistent oracles on which the keys differ, and the final step of the argument evaporates. The authors flag this explicitly as future work.

Section 7 of the paper shows the gap is not purely technical in general. If the folklore Simulation Conjecture [AA14] is false — in the polynomial-time form the paper calls Conjecture 7.2, that the acceptance probability of any polynomial-time quantum oracle algorithm can be estimated to within $1/\mathrm{poly}$ by a polynomial-query classical algorithm, for most oracles — then there is a single-bit QCCC protocol and an infinite set $\mathcal{U} \subseteq \mathbb{N}$ such that for every $\kappa \in \mathcal{U}$ the protocol has $(1-2^{-\kappa})$-completeness and $2^{-\kappa}$-IND-security against classical $\mathrm{poly}(\kappa)$-query eavesdroppers. So an unconditional classical attack covering imperfect completeness in general would settle a long-standing question about quantum speedups. Crucially, in that construction *both* parties run the hard-to-simulate quantum algorithm and independently round their own empirical estimates against a common random shift that Alice sends; a classical Alice cannot do her half. The classical-Alice case is therefore the one place where the paper's own barrier does not stand in the way, which is what makes it a target rather than a wish.

Progress to date. The paper proves the perfect-completeness case unconditionally and gives the two quantum lemmas (Lemma 3.2, independence of views as a product state conditioned on transcript and oracle; Lemma 3.3, consistency via the maximal Fourier monomial) that survive unchanged when completeness is imperfect — it is only the last step of Lemma 3.9 that uses $\Pr[k_A = k_B] = 1$. It also gives a sharper analysis showing the error is εd_B rather than $\varepsilon(d_A + d_B)$, by replacing the true joint state with the state of

a variant of Bob that fixes Alice's messages from the transcript. On the negative side, Theorem 7.6 shows what an attack for imperfect completeness costs in general, and Section 1.3 notes that the Aaronson–Ambainis conjecture can be applied to imperfect completeness but, as far as the authors can see, only when there is no communication.

2 Notation and parameters

$\kappa \in \mathbb{N}$ is the security parameter; all other quantities may depend on it. $h\colon \{0,1\}^n \to \{0,1\}^m$ denotes the random oracle, a uniformly random function with $n = n(\kappa)$ and $m = m(\kappa)$. A *quantum query* to h is an application of the unitary $|x\rangle|y\rangle \mapsto |x\rangle|y \oplus h(x)\rangle$; a *classical query* is the map $x \mapsto h(x)$ on a classical input. We write d_{A} and d_{B} for the number of oracle queries made by Alice and by Bob, t for the protocol transcript, $k_{\mathsf{A}}, k_{\mathsf{B}}, k_{\mathsf{E}}$ for the keys output by Alice, Bob and the eavesdropper, and γ for the completeness error. As usual $\mathrm{negl}(\kappa)$ denotes a function smaller than $1/\kappa^c$ for every constant $c > 0$ and all sufficiently large κ, and $\lambda \in (0,1)$ is the eavesdropper's failure parameter.

- κ: security parameter; all protocol quantities and the attack's query bound are measured against it.
- $d_{\mathsf{A}}, d_{\mathsf{B}}$: number of classical queries Alice makes and number of quantum queries Bob makes; their sum is assumed to be polynomial in κ.
- γ: completeness error, the probability that the two parties end up with different keys. The perfect case $\gamma \equiv 0$ is proved in the paper; the conjecture is the case $\gamma = \mathrm{negl}(\kappa)$.
- λ: the eavesdropper's allowed failure probability, an arbitrary constant in $(0,1)$; the query bound q may depend on it.
- m: output length of the random oracle. Deliberately unconstrained: the perfect-completeness attack's query cost does not depend on it.

3 The conjecture

Definition 1 (CAQB key agreement protocol in the QROM). For a security parameter κ, let $h\colon \{0,1\}^{n(\kappa)} \to \{0,1\}^{m(\kappa)}$ be a uniformly random function. A *classical-Alice quantum-Bob (CAQB) key agreement protocol* $\Pi_\kappa = (\mathsf{A}, \mathsf{B})$ is a two-party interactive protocol in which:

- A is a classical randomised algorithm that may make at most d_{A} classical queries to h;
- B is a quantum algorithm, starting from a fixed state, that may apply arbitrary unitaries and measurements to its own registers and make at most d_{B} quantum queries to h;
- every message exchanged is a classical string, and the concatenation of all messages, in order, is the *transcript* t;
- at the end A outputs a key k_{A} and B outputs a key k_{B}, both in $\{0,1\}^{\ell(\kappa)}$ for some $\ell(\kappa) \geq 1$.

The protocol has *completeness error* $\gamma(\kappa)$ if $\Pr[k_{\mathsf{A}} = k_{\mathsf{B}}] \geq 1 - \gamma(\kappa)$, the probability being over h, the coins of A, and the measurement outcomes of B. It has *perfect completeness* if $\gamma \equiv 0$.

Definition 2 (Classical eavesdropper). A *classical eavesdropper* E is a classical randomised algorithm, of unbounded computational power, which is given the transcript t of an execution of Π_κ, may make classical queries to the same oracle h used in that execution, and outputs a key $k_{\mathsf{E}} \in \{0,1\}^{\ell(\kappa)}$. Its query complexity is the expected number of classical queries it makes, over the choice of h, the execution, and its own coins.

Conjecture 1 (classical attacks without perfect completeness). *Let $\{\Pi_\kappa = (\mathsf{A}, \mathsf{B})\}_{\kappa \in \mathbb{N}}$ be any family of CAQB key agreement protocols in the QROM (Definition 1) such that*

1. *there is a polynomial p_0 with $d_{\mathsf{A}}(\kappa) + d_{\mathsf{B}}(\kappa) \le p_0(\kappa)$ for all κ, and*

2. *the completeness error satisfies $\gamma(\kappa) = \mathrm{negl}(\kappa)$.*

Then for every constant $\lambda \in (0,1)$ there exist a classical eavesdropper E (Definition 2) and a polynomial q such that for all sufficiently large κ, E makes at most $q(\kappa)$ classical queries to h in expectation and

$$\Pr\left[k_{\mathsf{E}} = k_{\mathsf{A}}\right] \;\ge\; 1 - \lambda,$$

where the probability is over the choice of h, the coins and measurement outcomes of A and B, and the coins of E. No bound is imposed on the domain size $2^{n(\kappa)}$, on the output length $m(\kappa)$, or on the computational (as opposed to query) complexity of E.

Remark 1 (expected versus worst-case queries). The query bound in Conjecture 1 is on the *expected* number of queries, matching Theorem 3.1 of the source and its Lemma 3.8, whose Construction 3.7 carries no worst-case bound. A worst-case bound is obtained from the expected one by truncation, at the cost of an additive constant in λ.

Remark 2 (why the statement stops at a classical Alice). The source poses this as "an intriguing question for future work" for all of its attacks, which include the attack in which Alice too is quantum. The conjecture above is deliberately narrowed to the classical-Alice case, which is what "the above result" in the paper's own sentence refers to. Widening it back to a quantum Alice would, by the paper's Theorem 7.6, make it imply the Simulation Conjecture, and so turn an unconditional target into a conditional one.

4 Bibliography

[AA14] S. Aaronson, A. Ambainis. *The Need for Structure in Quantum Speedups.* 2014. No venue printed in the source bibliography.

[BKSY11] Z. Brakerski, J. Katz, G. Segev, A. Yerukhimovich. *Limits on the Power of Zero-Knowledge Proofs in Cryptographic Constructions.* Theory of Cryptography Conference, pages 559–578, 2011.

[BM17] B. Barak, M. Mahmoody. *Merkle's Key Agreement Protocol is Optimal: An $O(n^2)$ Attack on Any Key Agreement from Random Oracles.* Journal of Cryptology, 30(3), 2017.

[BMG09] B. Barak, M. Mahmoody-Ghidary. *Merkle Puzzles are Optimal — An $O(n^2)$-Query Attack on Any Key Exchange from a Random Oracle.* Advances in Cryptology — CRYPTO 2009, LNCS 5677, pages 374–390, Springer, 2009.

[HY20] A. Hosoyamada, T. Yamakawa. *Finding Collisions in a Quantum World: Quantum Black-Box Separation of Collision-Resistance and One-Wayness.* International Conference on the Theory and Application of Cryptology and Information Security, pages 3–32, Springer, 2020.

[IR89] R. Impagliazzo, S. Rudich. *Limits on the Provable Consequences of One-Way Permutations.* Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC), pages 44–61, ACM Press, 1989.

[Mer74] R. Merkle. *C.s. 244 Project Proposal.* 1974. Facsimile available at http://www.merkle.com/1974.