Provably secure key establishment against quantum adversaries

A. Belovs G. Brassard P. Høyer M. Kaplan S. Laplante L. Salvail

Not taking classified work within secret services into consideration, Merkle is the first person to have asked—and solved—the question of secure communications over insecure channels. In his seminal (rejected!) 1974 project for a Computer Security course at the University of California, Berkeley, he discovered that it is possible for two people who want to communicate securely to establish a secret key by communicating over an authenticated channel that provides no protection against eavesdropping. Merkle's solution to this conundrum offers *quadratic security* in the sense that if the legitimate parties—codenamed Alice and Bob—are willing to expend an effort in the order of N, for some security parameter N, they can establish a key that no eavesdropper—codenamed Eve—can discover with better than vanishing probability without expending an effort in the order of N^2.

Quadratic security may seem unattractive compared to the potential exponential security entailed by the subsequently discovered key establishment protocols of Diffie and Hellman and of Rivest, Shamir and Adleman, among others. However, the security of those currently ubiquitous cryptographic solutions will be compromised with the advent of full-scale quantum computers, as discovered by Shor more than two decades ago. And even if a quantum computer is never built, no one has been able to prove their security against classical attacks, nor that of quantum-resistant candidates based, for instance, on short vectors in lattices. Furthermore, Merkle had already understood in 1974 that quadratic security *could* be practical if the underlying one-way function (see below) can be computed very quickly: if it takes one nanosecond to compute the function and legitimate users are willing to spend one second each, a classical adversary who could only invert the function by exhaustive search would require fifteen expected *years* to break Merkle's original scheme.

The main interest of Merkle's solution is that it offers *provable* security, at least in the *query model* of computational complexity, a model closely related to the random oracle model. In this model, we assume the existence of a *black-box* function $f : D \to R$ from some domain D to some range R, so that the only way to learn something about this function is to query the value of $f(x)$ on inputs $x \in D$ that can be chosen arbitrarily. The *query complexity* of some problem given f is defined as the expected number of calls to f required to solve the problem, using the best possible algorithm. In our case of interest, we shall consider *random* black-box functions, meaning that for each $x \in D$, the value of $f(x)$ is chosen uniformly at random within R, independently of the value of $f(x')$ for any other $x' \in D$. Provided the size r of R is sufficiently large compared to the size d of D, such a random function is automatically one-to-one, except with vanishing probability. The main characteristic of these black-box random functions that is relevant to the proof of security of Merkle's scheme is that, given a randomly chosen point y in the image of f, the only (classical) approach to finding an x so that $f(x) = y$ is exhaustive search: we have to try x's one after another until a solution is found. Indeed, whenever we try some x' and find that $f(x') \neq y$, the *only* thing we have learned is that this particular x' is not a solution. Provided the function is indeed one-to-one, we expect to have to query the function $d/2$ times on average in order to find the unique solution.

One may argue that black-box random functions do not exist in real life, but we can replace them in practice with one-way functions—provided *they* exist!—which is what Merkle meant by "one-way encryption" in his 1974 class assignment. Thus, we can base the security of Merkle's scheme on the *generic* assumption that one-way functions exist, which is unlikely to be broken by a quantum computer, rather than the assumption that *specific* computational problems such as factorization or finding short vectors in lattices are difficult, at least the first one of which is known not to hold on a quantum computer.

It was apparently noticed for the first time by one of us in 2005, and published a few years later, that Merkle's original 1974 scheme, as well as his better known subsequently published *puzzles*, are broken by Grover's algorithm on a quantum computer. This attack assumes that the eavesdropper can query the function in quantum superposition, which is perhaps not reasonable if the function is provided as a *physical* classical black box, but is completely reasonable if it is given by the publicly-available *code* of a one-way function (as originally envisioned by Merkle). If the legitimate parties are also endowed with a quantum computer, the same paper gave an obvious fix, by which the legitimate parties can establish a key after $O(N)$ quantum queries to the black-box function, but no quantum eavesdropper can discover it with better than vanishing probability without querying the function $O(N^{3/2})$ times.

At the CRYPTO 2011 conference, several of us introduced a new quantum protocol that no quantum eavesdropper could break without querying the black-box functions $\Omega(N^{5/3})$ times. We also offered the first protocol provably capable of protecting *classical* codemakers against *quantum* codebreakers, although $O(N^{13/12})$ queries in superposition sufficed for the quantum eavesdropper to obtain the not-so-secret key. Unfortunately, our security proofs were worked out in the traditional computational complexity *worst-case* scenario. In other words, it was only proved that any quantum eavesdropper limited to $o(N^{5/3})$ or $o(N^{13/12})$ queries, depending on whether the legitimate parties are quantum or classical, would be likely to fail *on at least one possible instance* of the protocol. This did not preclude that most instances of the protocol could result in insecure keys against an eavesdropper who would work no harder than the legitimate parties. Said otherwise, our CRYPTO 2011 result was of limited cryptographic significance.

In subsequent work (arXiv/1108.2316v2), we claimed to have provided a proper average-case analysis of our protocols, rendering them cryptographically meaningful, so that any quantum eavesdropper has a vanishing probability of learning the key after only $o(N^{5/3})$ or $o(N^{7/6})$ queries [1], where the probabilities are taken not only over the execution of the eavesdropping algorithm but also over the instance of the protocol run by the legitimate parties. In the same paper, we also extended our results to two sequences of protocols based on the k-SUM problem, where $k \geq 2$ is an integer parameter, in which the legitimate parties query the black-box random functions $O(kN)$ times. It was claimed that any quantum eavesdropper has a vanishing probability of learning the key after $o(N^{\frac{1}{2}+\frac{k}{k+1}})$ or $o(N^{1+\frac{k}{k+1}})$ queries, against the classical or the quantum protocol parametrized by k, respectively. Again, this was claimed to hold not only in the cryptographically-challenged worst-case scenario, but also when the probabilities are taken over the protocols being run by the legitimate parties.

Unfortunately, all our average-case analyses were incorrect! The case $k = 2$ can be fixed rather easily, hence the insufficiency of $o(N^{5/3})$ queries for a quantum-against-quantum protocol and of $o(N^{7/6})$ queries for a classical-against-quantum protocol in a cryptographically significant setting can be derived from the incorrect arguments provided in arXiv/1108.2316v2 (this will indeed be fixed in a forthcoming v3!). However, we also claimed that the case $k > 2$ could be proved in ways "similar to" when $k = 2$. This was a mistake due to a fundamental difference

[1] For classical legitimate parties, our original $o(N^{13/12})$ had been improved to $o(N^{7/6})$.

in the k-SUM problem whether $k = 2$ or $k > 2$. Whereas the 2-SUM problem is easily seen to be random self-reducible, so that its hardness in worst case implies its hardness on average, this does not seem to be the case for the k-SUM problem when $k > 2$. In particular, the worst-case lower bound proved by Belovs and Špalek in 2013 on the difficulty of solving the k-SUM problem on a quantum computer does not extend in any obvious way to a lower bound on average. And without such an average lower bound, our previously claimed results went up in smoke for $k > 2$. Furthermore, for a technical reason, even such an average lower bound would not suffice as we also need an average-case composition theorem.

In this work, we overcome all these difficulties and give the first correct and cryptographically significant [2] proof for our earlier claims. Consequently, we prove that for any $\varepsilon > 0$ there is a *classical* protocol that allows the legitimate parties to establish a common key after $O(N)$ expected queries to black-box random functions, yet any *quantum* eavesdropper will have a vanishing probability of learning their key after $O(N^{1.5-\varepsilon})$ queries to the same oracle. The vanishing probability is over the randomness in the actual run of the protocol followed by that of the eavesdropper's algorithm. If we allow the legitimate parties to use quantum computers as well, their advantage over the quantum eavesdropper becomes arbitrarily close to the quadratic advantage that classical legitimate parties enjoyed over classical eavesdroppers in the seminal 1974 work of Merkle.

Our results require new tools in quantum query complexity, which are of independent interest. In particular, we introduce techniques to lower-bound the quantum query complexity of distinguishing between two probability distributions, which we use to extend the adversary lower bound method in order to handle average-case complexity, but they could have other uses in cryptography. This approach is necessary for the distributions of inputs considered here because the associated decision problems become trivial on average, which prevents us from applying the (very few) methods previously developed to determine average-case quantum query complexity lower bounds. Furthermore, we prove the required composition theorem, which applies to our lower bound method. Using these two tools, we prove that any quantum eavesdropper who does not make a prohibitive number of calls to the black-box functions will fail to break a typical instance of the protocol, except with vanishing probability.

This work fits in the general framework of "Cryptography in a quantum world", which addresses the question: "Is the fact that we live in a quantum world a blessing or a curse for codemakers?". It is a blessing if we allow quantum communication, thanks to Quantum Key Establishment (aka Quantum Key Distribution—QKD), at least if the protocols can be implemented faithfully enough to close the door on quantum hackers. On the other hand, it is a curse if we continue to use the current cryptographic infrastructure, which intends to secure the Internet, but at the risk of falling prey to upcoming quantum computers. However, it is mostly a draw in the realm of provable query complexity in the black-box model considered in this work since codemakers enjoy a quadratic (or arbitrarily close to being quadratic) advantage over codebreakers in both an all-classical or an all-quantum world, at least in terms of query complexity (but see footnote 2 again).

The full paper is available at `http://arxiv.org/abs/1704.08182`. It has been accepted for oral presentation at the 12th Conference on Theory of Quantum Computation, Communication and Cryptography (TQC), to take place in Paris in June 2017, and will appear in its Proceedings.

[2] It is not entirely cryptographically meaningful to restrict the analysis to the number of calls to the black-box functions, taking no account of the computing *time* that may be required outside those calls. However, if we also restrict the legitimate expected time to be in $O(N)$, then our quantum protocol with $k = 3$ remains valid and provably resists any $o(N^{7/4})$-time quantum eavesdropping attack, which was previously claimed, but with a fundamentally incorrect proof. This is reasonably close to quadratic security to be of potential practical use.