

# Quantum Diffie-Hellman Key Exchange

Dirk Fischer\*

In 2014, the author conceived of a quantal version of the classical cryptographic Diffie-Hellman key exchange protocol. However, a resulting paper was declined to be published (by a here not disclosed journal) due to the lack of a security proof. No further publication attempts then were made by the author. In the time being, the same idea was conceived by others as well, resulting in a number of publications by others regarding this (previously declined) topic. As the author is unaware of the exact order in time of his own ideas compared to the ones of others, of course no prior art or similar claim is intended to be made here. However, the significance of the author's original idea is underlined, despite of being rejected by peer review mechanisms. And despite the fact that, even today, quantal security proofs are - at best - in their infancy. The paper at hand therefore serves two purposes: First, it might serve others (especially young researchers) as an example to not feel discouraged by publication refusals, if they truly deem the own research as an important novelty. Second, it provides an easy to understand introduction to grasp the concept of a quantum Diffie-Hellman key exchange. All of the following paragraphs, including the following remainder of this abstract, are taken from the original 2014 publication attempt and are unchanged in comparison to the 2014 original:

In this work, a quantal version of the classical cryptographic Diffie-Hellman key exchange protocol is introduced. It is called Quantum Diffie-Hellman key exchange. Unlike for the existing quantum key distribution protocols, actual quantum states, and not their measurement outcomes, are regarded as finally exchanged keys/information. By implementation of that quantal Diffie-Hellman version, both communication parties in the end are in possession of identically prepared, and secret quantum states. Thus the cryptographically important principle of forward secrecy is now available in a quantum physical framework. As a merit of the quantum setting, an improvement of the classical Diffie-Hellman protocol is also achieved, as neither of the two parties exactly know the final, exchanged states.

PACS numbers: 03.65.Ta, 03.67.-a, 03.67.Dd

## *Introduction*

Since the introduction of the first quantum key exchange protocol [1], the two former disjoint disciplines of quantum physics and cryptography began to get closer to each other. The main novelty of quantum physical concepts in cryptographic settings is owing to the differences of Shannon information theory and quantum physical principles. For instance, usage of principles like complementarity, and non-cloneability [1], and entanglement [2] on the one hand imposes restrictions on information processing. But on the other hand it grants new ways of tackling cryptographic challenges. In the next sections, the author introduces quantal versions of a key exchange protocol, known as Diffie-Hellman key exchange [3]. By this, a well-established protocol, which is utilized in almost every secured communication environment, now experiences a transition to the framework of quantum physics.

This paper is organized as follows: In the first three sections, the Quantum Diffie-Hellman protocol will be introduced, an example given and a short discussion delivered. After that, a further modified protocol version is devised. It makes use of an also presented qubit authentication method, in order to harden the Quantum Diffie-Hellman scheme against a common type of attack. A final conclusion section then ends this paper.

## *Quantum Diffie-Hellman scheme*

There are two corner stones of the classical Diffie-Hellman key exchange protocol (short: DH) in general: First, that the finally exchanged keys themselves are not transmitted, nevertheless both communication parties afterwards are in possession of identical keys or key information. Second, that despite both parties in the end being in possession of identical keys, neither of the two actually knows or can easily determine, what secret information was contributed by the other party in order to complete the key exchange.

In the following, a novel quantum physics based DH protocol (short: qDH) will be discussed in detail. The counterpart to the final key in the regular DH protocol will be a quantum state in the qDH scheme. The prerequisites of the qDH protocol are discussed in the next lines. Given two communication parties A, and B, which are able to exchange qubits. Both parties A, and B, each hold a set of secret values, denoted by  $\mathcal{S}_A$ , and  $\mathcal{S}_B$ , respectively. These fixed secrets, during the course of the key exchange scheme, will serve to generate shared, secret states for A, and B. These generated secret states then are regarded as ephemeral ones, whereas the initial secrets contained in  $\mathcal{S}_A$ , and  $\mathcal{S}_B$ , are kept stored by A, and B, respectively. Prior to the qDH communication, both parties need to have agreed on a set  $\mathcal{S}_P$  of publicly known, shared information. As quantum states and their

manipulation are considered in qDH,  $\mathcal{S}_P$  comprises state information of a common, initial state. It will be denoted by  $|0\rangle$ . Hence  $\mathcal{S}_P = \{|0\rangle, \dots\}$ , with “...” indicating optional presence of further, shared, and publicly known information. The initial state  $|0\rangle$  being shared, or publicly known, means that both A, and B must be able to prepare such an exact state. Starting point for any operation for A, and B then has to be such a mutually known state. Furthermore, the sets  $\mathcal{S}_A$ ,  $\mathcal{S}_B$ , and  $\mathcal{S}_P$  give rise to individual unitary operators  $U(\mathcal{S}_A)$ ,  $U(\mathcal{S}_B)$ , and  $U(\mathcal{S}_P)$ , respectively. Actual examples for  $\mathcal{S}_A$ ,  $\mathcal{S}_B$ ,  $\mathcal{S}_P$ , and their respective unitary operations will be given later. First, it is shown, how the general qDH scheme constitutes:

1. Party A prepares a qubit state  $|0\rangle$ , as given in the set of shared, publicly available information  $\mathcal{S}_P$ . Afterwards, A sends a qubit  $|\psi_A\rangle := U(\mathcal{S}_A)|0\rangle$  to B, where  $U(\mathcal{S}_A)$  is a unitary operator, depending on the secrets of set  $\mathcal{S}_A$ .
2. Party B receives  $|\psi_A\rangle$  and modifies it by its own unitary operator  $U(\mathcal{S}_B)$ , depending on the secrets of set  $\mathcal{S}_B$ . This results in  $|\psi_{BA}\rangle := U(\mathcal{S}_B)|\psi_A\rangle$ .
3. Now B performs similar steps than A did during the preceding two steps: B prepares a qubit state  $|0\rangle$ , as given in the set of shared, publicly available information  $\mathcal{S}_P$ . B sends a qubit  $|\psi_B\rangle := U(\mathcal{S}_B)|0\rangle$  to A.
4. Party A receives  $|\psi_B\rangle$  and modifies that state by with  $U(\mathcal{S}_A)$ , and unitary  $U(\mathcal{S}_P)$ . This results in  $|\psi_{AB}\rangle := U(\mathcal{S}_P)U(\mathcal{S}_A)|\psi_B\rangle$ .

Both communication parties are now in possession of the following states:  $|\psi_{AB}\rangle$  (for A at the end of step 4) and  $|\psi_{BA}\rangle$  (for B at the end of step 2). We now require, that

$$U(\mathcal{S}_P)U(\mathcal{S}_A)U(\mathcal{S}_B) = U(\mathcal{S}_B)U(\mathcal{S}_A). \quad (1)$$

Given that constraint for  $U(\mathcal{S}_A)$ ,  $U(\mathcal{S}_B)$ , and  $U(\mathcal{S}_P)$ , one has

$$|\psi_{AB}\rangle = |\psi_{BA}\rangle, \quad (2)$$

and thus both A, and B are in possession of identical quantum states, without ever having transmitted these final states and without both parties knowing the other party’s set of secrets.

The next table visualizes that concept. Each enframed row represents one of the above steps 1 to 4. That column headed by A represents actions on the side of A. Likewise for the column headed by B. The column “quantum channel” depicts the quantum states transmitted and their transmission direction.

In the following, a simple example for  $\mathcal{S}_A$ ,  $\mathcal{S}_B$ ,  $\mathcal{S}_P$ , and their respective unitary operators is presented.

	A	quantum channel	B
1.	$ 0\rangle \mapsto U(\mathcal{S}_A) 0\rangle$ $:=  \psi_A\rangle$	$ \psi_A\rangle$ $\longrightarrow$	
2.			$ \psi_A\rangle \mapsto$ $U(\mathcal{S}_B) \psi_A\rangle$ $:=  \psi_{BA}\rangle$
3.		$ \psi_B\rangle$ $\longleftarrow$	$ 0\rangle \mapsto$ $U(\mathcal{S}_B) 0\rangle$ $:=  \psi_B\rangle$
4.	$ \psi_B\rangle \mapsto$ $U(\mathcal{S}_P)U(\mathcal{S}_A) \psi_B\rangle$ $:=  \psi_{AB}\rangle$		

TABLE I. Quantum Diffie-Hellman scheme. Starting from top, steps 1 to 4 show the sequential protocol process and actions between party A (column “A”) with secrets  $\mathcal{S}_A$ , and party B (column “B”) with secrets  $\mathcal{S}_B$ . The column “quantum channel” depicts actual the communication process and transmission direction via a quantum channel. Owing to Eq. (2), in the end both A, and B, possess identical states  $|\psi_{AB}\rangle$ , and  $|\psi_{BA}\rangle$ , respectively.

#### Example

Given  $\mathcal{S}_P = \{|0\rangle, \hat{\mathbf{n}}\}$  is shared between A, and B, or publicly available.  $\hat{\mathbf{n}}$  is a vector  $\in \mathbb{R}^3$  of unit modulus. Choosing  $\mathcal{S}_A = \{\alpha\}$ , with  $\alpha \in \mathbb{R}$  a secret real number, that  $\alpha$  serves as parameter for the operator

$$U(\mathcal{S}_A) = U(\alpha) := \exp\left(-i \frac{\alpha}{2} \hat{\mathbf{n}} \cdot \hat{\boldsymbol{\sigma}}\right). \quad (3)$$

Here,  $\hat{\boldsymbol{\sigma}}$  is a 3-vector with the Pauli matrices  $\boldsymbol{\sigma}_i$  ( $i = 1..3$ ) as components. This well-known operator is the qubit rotation around the axis  $\hat{\mathbf{n}}$  and angle  $\alpha$ . Similarly as in Eq. (3), if for B one defines  $\mathcal{S}_B = \{\beta\}$  with  $\beta \in \mathbb{R}$ , one has  $U(\mathcal{S}_B) = U(\beta) := \exp(-i(\beta/2)\hat{\mathbf{n}} \cdot \hat{\boldsymbol{\sigma}})$ . With that choice of secret-dependent operators, one has  $[U(\mathcal{S}_A), U(\mathcal{S}_B)] = 0$ , i.e., the commutator of  $U(\mathcal{S}_A)$  and  $U(\mathcal{S}_B)$  vanishes. This is just the requirement of Eq. (1), if  $U(\mathcal{S}_P)$  is additionally chosen to be the identity operator  $\mathbf{Id}$  in qubit space. In summary, this leads to the desired equality stated in Eq. (2).

Evidently, the presented example is only a very simple one, where commuting operators trivially led to the fulfillment of Eq. (1). However, it is also possible to utilize non-commuting operators. In a variety of cases, the so far unmotivated operator  $U(\mathcal{S}_P)$ , and set of shared, public information  $\mathcal{S}_P$ , then can be chosen accordingly to cancel out emerging cross-terms. Furthermore, for the sake of brevity, here only classical real numbers  $\alpha$  and  $\beta$  were used as individual secrets for  $\mathcal{S}_A$ , and  $\mathcal{S}_B$ , respectively. Of course, quantum states as well could serve as individual secrets.

#### Discussion

In comparison to [1] and [2], the final, pure quantum

states resemble the exchanged keys or information. No classical communication channel is required for the qDH scheme. Additionally, in the end both parties A, and B are in possession of identical quantum states, despite the fact, that neither A nor B exactly know their state or the other party's contributed secret value. As a consequence, this means even if, e.g., A revealed the secret set  $\mathcal{S}_A$ , the overall key state  $|\psi_{AB}\rangle$  still cannot not be unambiguously determined. Unless  $\mathcal{S}_B$  would be known as well. This is a general, important principle of the Diffie-Hellman key exchange scheme. It is known as forward secrecy and is, owing to the presented qDH scheme, now available in a purely quantum physical setting.

However, as it is for the original DH protocol, qDH as well is susceptible to so called man-in-the-middle attacks. The term man-in-the-middle-attack corresponds to malicious behaviour, where a third party C (besides parties A, and B) intercepts the quantum channel between A, and B and impersonates A, or B. In case neither party A, nor B, were aware of that interception, instead of a qDH exchange between A, and B [short: qDH(A, B)], one would have qDH(A, C) and qDH(C, B). In consequence, the parties A, and B would respectively share a respective, identical, secret quantum state with an attacker C. Albeit both A, and B would be in good faith of having succeeded in a proper qDH key exchange qDH(A, B).

This obstacle can be overcome with a modified qDH scheme, where a so called qubit authentication will be employed. For the DH scheme, an encryption based authentication needs to be introduced to counter a man-in-the-middle attack. For a quantum version, such mechanisms are not readily available. Hence, the classical authentication mechanism has to be substituted by another method, which will be derived in the next section. Afterwards, that new quantum authentication mechanism will be incorporated into qDH, leading to a so called authenticated qDH exchange protocol (short: aqDH).

#### *Qubit authentication*

In a authentication scheme for quantum states, comparison of some arbitrary quantum state against another, known one is necessary. Comparison is meant in the sense that, given a qubit  $|\psi_{\text{cmp}}\rangle$  whose complete state information is known and regarded as to-be-compared value, its equality to other (arbitrary, maybe unknown) states  $|\psi_{\text{arb}}\rangle$  is checked. For such a functionality, it is necessary that only those states, which are (up to a phase factor) identical to a desired state  $|\psi_{\text{cmp}}\rangle$ , yield an output interpreted as “yes”. And all other states result in an output, which can be unambiguously interpreted as “no”.

In the next lines, for qubits, such a comparison measurement scheme is devised:

1. Setup a projection operator  $\hat{P}_{\text{cmp}} := \langle \psi_{\text{cmp}}^\perp |$ , such that it projects onto the linear qubit subspace given by  $|\psi_{\text{cmp}}^\perp\rangle$  as base state, where  $|\psi_{\text{cmp}}^\perp\rangle$  denotes the orthogonal complement of  $|\psi_{\text{cmp}}\rangle$ . As only qubits are considered here and the dimensionality of the state space is equal to 2, that state  $|\psi_{\text{cmp}}^\perp\rangle$  is unique up to an irrelevant phase factor  $\kappa \in \mathbb{C}$  with  $\|\kappa\| = 1$ . As an example: In case of photons and their polarization as qubit state, this projection operator  $\hat{P}_{\text{cmp}}$  might be implemented by a polarization filter in the state space direction  $|\psi_{\text{cmp}}^\perp\rangle$ .
2. Setup a measurement apparatus, which resides after the projection operator  $\hat{P}_{\text{cmp}}$ . It only needs to check, whether a quantum state is present or not. In case of photons, this might simply be a fluorescent screen behind the projection operator of the above step 1.

Now an arbitrary qubit  $|\psi_{\text{arb}}\rangle$  consecutively undergoes the preceding steps 1 and 2. That indeed a distinctive, dichotomic answer is obtained, can be seen from the following facts: A general, arbitrary qubit  $|\psi_{\text{arb}}\rangle$  can always be uniquely decomposed according to

$$|\psi_{\text{arb}}\rangle = a|\psi_{\text{cmp}}\rangle + b|\psi_{\text{cmp}}^\perp\rangle, \quad (4)$$

with  $a, b \in \mathbb{C}$  and  $\|a\|^2 + \|b\|^2 = 1$ . In case of  $b = 0$ , i.e., in case of identical states (up to a phase factor), owing to Eq. (4) the above step 1 absorbs  $|\psi_{\text{arb}}\rangle$ , leading to no detection in step 2. This will be defined as the affirmative “yes” answer. All other cases, where  $a \neq 0$  and  $b \neq 0$  (i.e.,  $|\psi_{\text{arb}}\rangle \neq |\psi_{\text{cmp}}\rangle$ , and both not only differing by a phase factor), will lead to a detection in step 2, again owing to Eq. (4). This detection is regarded as a declining “no” answer. Evidently, such a comparison scheme is suitable to serve as an authentication mechanism. In the following, the presented scheme therefore will be denoted by  $\text{AUTH}(|\psi_{\text{cmp}}\rangle, |\psi_{\text{arb}}\rangle)$ , in case an arbitrary qubit  $|\psi_{\text{arb}}\rangle$  has to be compared to a known qubit  $|\psi_{\text{cmp}}\rangle$ .

#### *Authenticated Quantum Diffie-Hellman scheme*

With the single qubit authentication method at hand, it is now possible to improve the previously introduced qDH scheme and harden it against man-in-the-middle attacks. Unlike for authenticated Diffie-Hellman scheme in the classical cryptographic setting, for the quantum Diffie-Hellman scheme the main idea of implementing an authentication component is to insert a certain control qubit  $|\phi\rangle$  into the stream of exchanged qubits. Then, only in case of no malicious interception during the qDH scheme occurs, in the end a proper authentication of that additional qubit should be possible. The authenticated qDH scheme (aqDH) then reads as follows:

1. Additional to the first step in table I, a further, arbitrary but secret qubit  $|\phi\rangle \in \mathcal{S}_A$  is sent by A. Then party A sends the two qubits  $|\psi_A\rangle$  (defined

in step 1 of table I) and  $|\phi\rangle$  to B in a random order to avoid an attacker knowing which state the control state is. That random order is denoted by the notation of an unordered set  $\{|\psi_A\rangle, |\phi\rangle\}$ . The newly introduced  $|\phi\rangle$  will later serve for authentication and validation purposes.

2. In step 2 now, party B is not aware of the order of received qubits. Party B therefore randomly chooses one of the two received states. That chosen state is denoted by  $|\eta_1\rangle \in \{|\psi_A\rangle, |\phi\rangle\}$ . The remainder of the two received states is denoted by  $|\eta_2\rangle$ . Then B performs a transformation similar to step 2 of table I, this time on  $|\eta_1\rangle$ , leading to  $|\eta_1\rangle \mapsto U(\mathcal{S}_B)|\eta_1\rangle =: |\eta_{B1}\rangle$ . That state then is kept by B.
3. Different than for step 3 of table I, B here sends two qubits to A. This time also  $|\eta_2\rangle$  is sent to party A, besides the newly generated  $|\psi_B\rangle$  like in step 3 of table I. As it was in the preceding step 1, both qubits are sent in a random order, denoted by the unordered set  $\{|\psi_B\rangle, |\eta_2\rangle\}$ . Additionally, compared to table I, upon arrival of these states at A, via a classical communication channel, party B provides information to A, whether the received first or second state of step 2 was kept at B. Furthermore, the order of states sent to A within the first part of this step 3 is disclosed. Formally, this resembles an enhancement of  $\mathcal{S}_P$  by the information “order of qubit choice”, but only after the states were sent by B and received by A.
4. Similar to step 4 in table I,  $|\psi_B\rangle$  is transformed to  $|\psi_{AB}\rangle$ . This is possible, as A can deduce, in which order the qubits were sent by B and which qubit presumably is  $|\psi_B\rangle$  (in case of no interception). With the remaining  $|\eta_2\rangle$  then A performs an authentication procedure against the desired state  $|\phi\rangle$ , i.e.,  $\text{AUTH}(|\phi\rangle, |\eta_2\rangle)$ . The authentication procedure’s outcome then provides means to detect a qubit interception with probability  $p(\text{detect}) \geq \frac{3}{4}$ .

The overall scheme is also given in the following table II, which uses the same notation and convention as in table I. However, in table II the communication channel also needs to transport classical information, as seen in step 3 above.

Via that authenticated quantum Diffie-Hellman scheme, an interception of communication thus is detectable. This can be seen as follows: If an attacker intercepts all of the transmitted qubits, a detection with probability  $p(\text{detect}) = 1$  will occur. Hence, an adversary can intercept and replace at most only one qubit transmitted during step 1, and one qubit during step 3. The malicious interception succeeds, iff exactly  $|\psi_A\rangle$  (in step 1) and  $|\psi_B\rangle$  (in step 3) are chosen by the adversary.

	A	comm. channel	B
1.	$ 0\rangle \mapsto  \psi_A\rangle,$ $ \phi\rangle$	$\{ \psi_A\rangle,  \phi\rangle\}$ $\longrightarrow$	
2.			Choose $ \eta_1\rangle \in \{ \psi_A\rangle,  \phi\rangle\}$ , remaining qubit denoted $ \eta_2\rangle$ . $ \eta_1\rangle \mapsto U(\mathcal{S}_B) \eta_1\rangle,$ $=:  \eta_{B1}\rangle$
3.		$\{ \eta_2\rangle,  \psi_B\rangle\}$ $\longleftarrow$  “order of qubit choice”  $\longleftarrow$	$ 0\rangle \mapsto  \psi_B\rangle$
4.	$ \psi_B\rangle \mapsto  \psi_{AB}\rangle,$ $\text{AUTH}( \phi\rangle,  \eta_2\rangle).$		

TABLE II. Authenticated Quantum Diffie-Hellman scheme. Starting from top, steps 1 to 4 show the sequential protocol process and actions of party A (column “A”) with secrets  $\mathcal{S}_A$ , and party B (column “B”) with secrets  $\mathcal{S}_B$ . The column “comm. channel” depicts the actual communication and transmission direction via a combined quantum- and classical channel.  $|\phi\rangle \in \mathcal{S}_A$  serves as an interception detection state. In steps 1 and 3, qubits are sent in a random order, only known to A, and B, respectively. In step 3, additionally classical information is sent by B upon arrival of the qubits at party A.  $\text{AUTH}(|\phi\rangle, |\eta_2\rangle)$  represents an execution of the qubit authentication scheme. Due to the qubits sent in random order, the authentication detects tampering with  $p(\text{detect}) \geq 3/4$ .

This happens with probability  $p(\text{attack}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ . As a result, with  $p(\text{detect}) = 1 - p(\text{attack})$ , one has a probability of  $p(\text{detect}) = \frac{3}{4}$  for an attacker to be detected after one protocol iteration if qubits are intercepted randomly. In that case, for  $n$  iterations of qDH, one has  $p(\text{detect}, n) = 1 - \left(\frac{1}{4}\right)^n$  for at least one attack detection.

### Conclusion

We close our considerations by stating, that a quantal version (qDH) of the classical Diffie-Hellman (DH) key exchange protocol was presented. In an actual example, the DH remainder class arithmetic on finite fields was substituted within qDH by arithmetics of (commutating) spin state rotations. The presented, general principle enables purely quantum physical frameworks to implement the principle of forward secrecy.

And unlike for existing quantum key exchange protocols, the finally obtained quantum states themselves represent the key agreed on, not their measurement outcomes. As an improvement, compared to the classical Diffie-Hellman key exchange, for the quantum version neither

party is able to uniquely determine the final secret states, which is a result of the quantum mechanical projection postulate and the no-cloning theorem. Furthermore, at least for qDH, no classical communication channel is necessary.

However, qDH needed a modification to provide resilience against man-in-the-middle attacks. Therefore, with the help of a newly devised qubit authentication method, for qubits the authenticated quantum Diffie-Hellman key exchange (aqDH) was introduced. By that, the risk of an adversary undetectedly intercepting the quantum communication channel, or impersonating one of the communication parties was mitigated. For  $n$  protocol executions, a malicious interception is detected with a probability of  $p(\text{detect}, n) \geq 1 - (\frac{1}{4})^n$ . But that resistance

comes at the cost of a classical communication channel, which needs to be used in order to complete an authenticated quantum Diffie-Hellman key exchange.

- 
- \* Dirk.Fischer@uni-dortmund.de; The author is currently not affiliated with the University of Dortmund
- [1] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, New York, 1984, volume 175, p. 8.
  - [2] Artur K. Ekert, *Phys. Rev. Lett.*, **67**, 661 (1991)
  - [3] E. Rescorla, *RFC 2631: Diffie-Hellman Key Agreement Method*, June 1999