Quantum non-malleability and authentication
(extended abstract)

Gorjan Alagic and Christian Majenz

QMATH, Department of Mathematical Sciences
University of Copenhagen

galagic@gmail.com majenz@math.ku.dk

1. Introduction.

In its most basic form, encryption ensures the secrecy of transmissions against eavesdroppers. Besides secrecy, another desirable property is *non-malleability*, which guarantees that an active adversary cannot modify the plaintext by manipulating the ciphertext. In the classical setting, secrecy and non-malleability are independent: there are schemes which satisfy secrecy but are malleable, and schemes which are non-malleable but transmit the plaintext in the clear. If both secrecy and non-malleability are desired, then pairwise-independent permutations provide information-theoretically perfect (one-time) security [12].

In the setting of quantum information, encryption is the task of transmitting quantum states over a completely insecure quantum channel. Information-theoretic secrecy for quantum encryption is well-understood. Moreover, significant progress has been made on more advanced constructions, such as authenticated encryption [5], secure two-party computation [10], blind computation [8], quantum fully-homomorphic encryption [9], and many more. Despite this high-level progress, a basic aspect of quantum encryption remains largely unstudied. Indeed, *quantum non-malleability* was considered in only one previous work, by Ambainis, Bouda and Winter [4]. Their definition (which we call ABW-NM) requires secrecy, and that the "effective channel" $\mathsf{Dec} \circ \Lambda \circ \mathsf{Enc}$ of any adversary Λ is trivial[1].

Unlike non-malleability, the closely-related subject of quantum authentication (where decryption is allowed to reject) has received significant attention (see, e.g., [1, 5, 7, 10, 11].) In this setting, there are two definitions. The widely-adopted definition of Dupuis, Nielsen and Salvail (DNS-authentication) asks that, regardless of whether decryption rejects, the average effective channel of any adversary does not touch the plaintext [10]. A more recent definition of Garg, Yuen and Zhandry (GYZ-authentication) asks that, in the accept case, the adversary does not touch the plaintext with high probability over the key (rather than on average) [11].

2. Summary of results.

In this work, we devise a new definition of quantum non-malleability (denoted NM). We prove several new results about quantum non-malleability, quantum authentication, and the connections between these two concepts. A summary of our results is as follows; we will focus on the exact case, but all the definitions and results have appropriate relaxations to the approximate setting; see the full paper for details [2].

[1] More precisely, it is either the identity or replacement by a fixed state.

2.1. New definition. We give a new definition of quantum non-malleability (NM), which improves on ABW-NM in a number of ways:

1. it is expressed in terms of entropic quantities, generalizing classical definitions [12];
2. it prevents more powerful attacks, which make use of side information about the plaintext;
3. it is immune to a devastating "plaintext injection" attack, whereby an adversary against an ABW-NM scheme can send a plaintext of their choice to the receiver;
4. it does not require secrecy; instead, we show *quantum non-malleability implies quantum secrecy.*

The last point is analogous to the fact that quantum authentication implies encryption [5].

Informally, our definition states that any attack on the ciphertext will result in no information gain – except via a "trivial attack" which is always possible against any scheme. In this trivial attack, the adversary simply decides whether or not to destroy the ciphertext, and remembers that choice in their side information. A formal definition is as follows. The relevant quantum registers are: plaintext A, ciphertext C, user's reference R, and adversary's side information B.

Definition 1 *A scheme is non-malleable (NM) if for any ϱ_{ABR} and any attack $\Lambda_{CB\to C\tilde{B}}$, the effective channel $\tilde{\Lambda}_{AB\to A\tilde{B}} = \mathsf{Dec}\circ\Lambda\circ\mathsf{Enc}$ satisfies*

$$I(AR:\tilde{B})_{\tilde{\Lambda}(\varrho)} \leq I(AR:B)_{\varrho} + h(p_{=}(\Lambda,\varrho)).$$

The binary entropy term $h(p_{=}(\Lambda,\varrho))$ captures the information gain of the aforementioned trivial attack. Formally, $p_{=}(\Lambda,\varrho) = F\left(\mathrm{Tr}_{\tilde{B}}\Lambda((\cdot)_C\otimes\varrho_B)\right)^2$ is the squared entanglement fidelity of the attack map as it acts on the ciphertext register if ϱ_B is input in the side information register.

2.2. New results on non-malleability. As mentioned above, our first result is that NM implies secrecy. Here, secrecy stands for one of a number of equivalent notions of security; one may for instance use analogues of IND [6] or SEM [3] for computationally unbounded adversaries.

Theorem 2 *If a quantum encryption scheme is non-malleable (NM), then it is also secret.*

We remark that this is a significant departure from the classical case, where secrecy and non-malleability are independent properties.

Next, we show that NM implies ABW-NM, and give a separation scheme which is secure under ABW-NM but insecure under NM. As described in [2], this separation scheme is in fact susceptible to a powerful attack, whereby a simple adversary can replace the output of decryption with a plaintext of the adversary's choice.

Theorem 3 *If a quantum encryption scheme is NM, then it is also ABW-NM.*

On the other hand, if we restrict our attention to schemes where the encryption maps are unitary, then we are able to show the following.

Theorem 4 *Let Π be a quantum encryption scheme such that encryption E_k is unitary for all keys k. Then Π is NM if and only if $\{E_k\}_k$ is a two-design.*

Together with the results of [4], this implies that NM and ABW-NM are in fact equivalent for unitary schemes. Finally, we can also characterize NM schemes in the general (i.e., not necessarily unitary) case, as follows.

Theorem 5 *A scheme is* NM *if and only if, for any* $\Lambda_{CB\to C\tilde{B}}$*, there exist maps* $\Lambda'_{B\to\tilde{B}}$*,* $\Lambda''_{B\to\tilde{B}}$ *such that the effective attack* $\tilde{\Lambda}_{AB\to A\tilde{B}}$ *has the form*

$$\tilde{\Lambda} = \mathrm{id}_A \otimes \Lambda' + \frac{1}{|C|^2-1}\left(|C|\left\langle D_K(\mathbb{1}_C)\right\rangle - \mathrm{id}\right)_A \otimes \Lambda''\,.$$

The maps Λ' and Λ'' can be computed directly from a description of Λ [2]. This characterization theorem shows that our notion provides *ciphertext non-malleability*: any ciphertext modification results in the plaintext being replaced by $D_K(\mathbb{1}_C)$.

2.3. New results on quantum authentication. The techniques we developed for quantum non-malleability also yield several new results on quantum authentication, as follows. We note that our definitions of authentication deviate slightly from the original versions [10, 11], in that decryption outputs a reject symbol in place of the plaintext (rather than setting a flag to "reject.")

First, we show how to build authentication from non-malleability. Given an encryption scheme $\Pi = \{E_k\}$, we define Π_t^{tag} to be a new scheme whose encryption is $\varrho \mapsto E_k\big(\varrho_A \otimes |0\rangle\langle 0|_B^{\otimes t}\big)E_k^\dagger$, and whose decryption rejects unless B measures to $|0^t\rangle$.

Theorem 6 *If a scheme* $\Pi = \{E_k\}$ *satisfies* NM*, then* Π_t^{tag} *is* 2^{2-t}*-*DNS*-authenticating.*

If the starting NM scheme is encryption via the Clifford group, then the result is the well-known Clifford scheme for authentication [1].

Next, we show that GYZ-authentication implies DNS-authentication.

Theorem 7 *If a scheme is* ε*-*GYZ*-authenticating, then it is also* $O(\sqrt{\varepsilon})$*-*DNS*-authenticating.*

This result is technically non-trivial: on one hand, GYZ requires high probability of success while DNS only needs success-on-average; on the other hand, GYZ requires nothing in the reject case while DNS still makes rather stringent demands.

Finally, we show that GYZ-authentication can be satisfied by a scheme which "tags" plaintexts as before, and encrypts with a unitary 2-design. This is a significant improvement over the analysis of [11], which required eight-designs for the same construction.

Theorem 8 *Let* $\Pi = \{E_k\}_k$ *be a* 2^{-t}*-approximate 2-design scheme. Then* Π_t^{tag} *is* $2^{-\Omega(t)}$*-*GYZ*-authenticating.*

Given the conclusions of Theorem 4, we may state this as follows: if a unitary scheme Π is non-malleable, then Π_t^{tag} is GYZ-authenticating. We remark that the simulation of adversaries in this proof is efficient, in the sense of [7].

3. Conclusion and open problems.

In this work, we introduced a new definition of quantum non-malleability, a core concept in encryption. Our notion addresses a major vulnerability in the previous definition, and can serve as a primitive for constructing authentication schemes. When using unitary 2-designs (e.g., the Clifford group) for non-malleable encryption, the resulting authentication schemes are secure under the strongest known definitions [11]. We remark that our work is also a natural starting point for future research on quantum non-malleability in the setting of many messages and computational security assumptions.

Bibliography

[1] Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. In *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 5-7, 2010. Proceedings*, pages 453–469, 2010.

[2] Gorjan Alagic and Christian Majenz. Quantum non-malleability and authentication. *CoRR*, abs/1610.04214, 2016. URL http://arxiv.org/abs/1610.04214.

[3] Gorjan Alagic, Anne Broadbent, Bill Fefferman, Tommaso Gagliardoni, Christian Schaffner, and Michael StJules. Computational security for quantum encryption. In *9th International Conference on Information Theoretic Security (ITICS), to appear.*, 2016.

[4] Andris Ambainis, Jan Bouda, and Andreas Winter. Nonmalleable encryption of quantum information. *Journal of Mathematical Physics*, 50(4):042106, 2009.

[5] Howard Barnum, Claude Crépeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 449–458. IEEE, 2002.

[6] Anne Broadbent and Stacey Jeffery. *Quantum Homomorphic Encryption for Circuits of Low T-gate Complexity*, pages 609–629. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-662-48000-7. doi: 10.1007/978-3-662-48000-7_30. URL http://dx.doi.org/10.1007/978-3-662-48000-7_30.

[7] Anne Broadbent and Evelyn Wainewright. Efficient simulation for quantum message authentication. *arXiv preprint arXiv:1607.03075*, 2016.

[8] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 517–526. IEEE, 2009.

[9] Y. Dulek, C. Schaffner, and F. Speelman. Quantum homomorphic encryption for polynomial-sized circuits. *ArXiv e-prints*, March 2016.

[10] Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Actively secure two-party evaluation of any quantum operation. In *Advances in Cryptology–CRYPTO 2012*, pages 794–811. Springer, 2012.

[11] Sumegha Garg, Henry Yuen, and Mark Zhandry. New security notions and feasibility results for authentication of quantum data. *arXiv preprint arXiv:1607.07759*, 2016.

[12] Akinori Kawachi, Christopher Portmann, and Keisuke Tanaka. Characterization of the relations between information-theoretic non-malleability, secrecy, and authenticity. In *International Conference on Information Theoretic Security*, pages 6–24. Springer, 2011.

[13] C. Portmann. Quantum authentication with key recycling. *ArXiv e-prints*, October 2016.