Side-Channel-Free Quantum Key Distribution

Samuel L. Braunstein and Stefano Pirandola
Computer Science, University of York, York YO10 5GH, United Kingdom
(Received 15 September 2011; published 30 March 2012)

Quantum key distribution (QKD) offers the promise of absolutely secure communications. However, proofs of absolute security often assume perfect implementation from theory to experiment. Thus, existing systems may be prone to insidious side-channel attacks that rely on flaws in experimental implementation. Here we replace all real channels with virtual channels in a QKD protocol, making the relevant detectors and settings inside private spaces inaccessible while simultaneously acting as a Hilbert space filter to eliminate side-channel attacks. By using a quantum memory we find that we are able to bound the secret-key rate below by the entanglement-distillation rate computed over the distributed states.

DOI: 10.1103/PhysRevLett.108.130502 PACS numbers: 03.67.Dd, 03.65.Ud, 03.67.Hk, 42.50.−p

In 1982 Richard Feynman conjectured the use of quantum systems as a technological platform for solving difficult calculations in physics. Eventually this insight lead to the field of quantum information processing. As part of the field's growth, it has partly diverged into the two main application domains: computation and communications, though much fundamental and technical overlap still exists. Interestingly, the key application that has started to mature and is now commercially available is quantum cryptography, or more precisely quantum key distribution (QKD) which has quickly moved from the purely theoretical [1–4] to a practical technology [5–8].

How can we explain the impressive industrial uptake of quantum cryptography and its ultimate aim to take over classical systems? The answer lies in the claim of ''absolute security'' [9]. Unfortunately, while the idea is very compelling, subtle details in implementation may introduce flaws that could, potentially, be open to attack. Specifically, attacks from so called ''side channels'' represent one of the most elusive threats in practical quantum cryptography, because a system could be vulnerable to side-channel attacks even if it is unbreakable in theory [10,11]. In fact, the recent approach of ''device-independent QKD'' [12] makes important advances in handling imperfect implementations, and can even be made by untrusted parties, but does not directly address all possible side-channel attacks, where, for example, detectors may directly receive external probing aimed at seeding or gleaning their readout.

In principle side-channel attacks affect both classical and quantum cryptography, but could be especially devastating for quantum cryptography, precisely because of the proclaimed absolute security ''guarantee''. The threat from such attacks has been demonstrated in both lab and installed field settings [11]. Thus, while practical QKD systems have been fighting a trade-off between distance and key generation rate, they are still facing the fundamental problem of guaranteed security, choosing to rely on theoretical promises of absolute security without having any way of authenticating them in practice.

Private spaces: general model.—Let us consider the scenario of Fig. 1. Two authenticated parties, Alice and Bob, control two private spaces, $\mathcal{A}$ and $\mathcal{B}$, respectively. Conventionally, these spaces are assumed completely inaccessible from the outside; i.e., no illegitimate system may enter $\mathcal{A}$ or $\mathcal{B}$. For this reason every kind of side-channel attack upon the private spaces is assumed excluded. In practice, however, any port can allow a side-channel to enter possibly probing any detector, state-generation or detector settings. To prevent or overcome such attacks, the QKD system must effectively isolate its private spaces: the private space must not be directly involved in either state preparation (for sending) or detection (of incoming states). To overcome such probing side-channel attacks, we propose performing state generation by collapse of a bipartite entangled state, so that any probe from outside is perfectly isolated from the state-generation ''machinery'' (see Ref. [13] for an extended discussion). Thus, in a manner akin to teleportation, we replace all real channels with virtual channels. This allows us to physically (and ''topologically'') separate all detectors and settings within the private space from external probing, while also

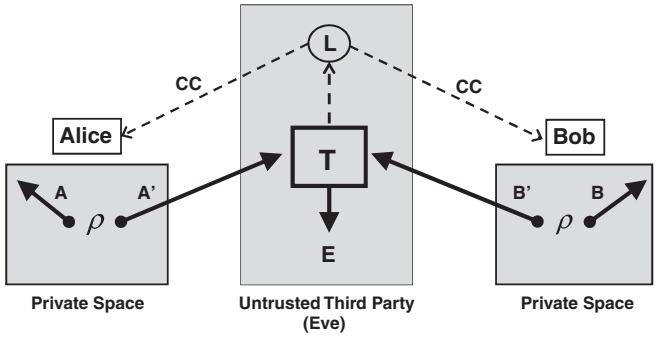

FIG. 1. Private space to private space. The UTP acts as a correlator.

acting as a Hilbert space filter [14] against any side channel.

Within its own private space, each party (Alice or Bob) has a bipartite state ρ which entangles two systems: $\{A, A'\}$ for Alice, and $\{B, B'\}$ for Bob. Systems $\{A, B\}$ are kept within the private spaces, while systems $\{A', B'\}$ are sent to an untrusted third party (UTP), whose task is to perform a quantum measurement and communicate the corresponding result. This untrusted LOCC then allows the creation of correlations between the private systems $\{A, B\}$ that Alice and Bob can exploit to generate a secret key. In its simplest form an ideal side-channel free QKD scheme reduces to an entanglement swapping setup [15], with the dual teleportation channel acting as an ideal Hilbert space filter. What is unique about our protocol is the ability to completely protect private space settings and detectors from probing side-channel attacks.

In the worst case scenario, the UTP must be identified with Eve herself, whose aim is to eavesdrop the key, or even prevent Alice and Bob from generating the key (i.e., a denial of service). In the most general case, Eve applies a quantum instrument $\mathbf{T} = \{T_l\}_{l=1}^{l_{\max}}$ to the incoming systems $\{A', B'\}$. This is a quantum operation with both classical and quantum outputs. For each classical outcome l, there is a corresponding completely positive (CP) map T_l applied to the systems $\{A', B'\}$ [16]. This means that the global input state $\rho_{AA'} \otimes \rho_{BB'}$ is transformed into the conditional output state

$$\rho_{ABE}(l) \equiv \frac{1}{p(l)}(I_A \otimes I_B \otimes T_l)(\rho_{AA'} \otimes \rho_{BB'}), \tag{1}$$

where E represents an output quantum system in the hands of Eve, while $I_A \otimes I_B$ is the identity channel acting on the private systems $\{A, B\}$. Cleary each outcome l will be found with some probability $p(l)$, depending both on T_l and the input state. As a consequence the classical output of $\mathbf{T}$ can be simply represented by the stochastic variable $L \equiv \{l, p(l)\}$. The quantum output of $\mathbf{T}$ is represented by the system E which is correlated with the private systems $\{A, B\}$ via the conditional state $\rho_{ABE|L}$ specified by Eq. (1). E is the system that Eve will use for eavesdropping. For instance, most generally Eve can store all the output systems E (generated in many independent rounds of the protocol) into a big quantum memory. Then, she can detect the whole memory using an optimal quantum measurement (corresponding to a collective attack).

According to the agreed protocol, the UTP must send a classical communication (CC) to both Alice and Bob in order to "activate" the correlations. Here, Eve has another weapon in her hands, i.e., tampering with the classical outcomes. In order to decrease the correlations between the honest parties, Eve may process the output stochastic variable L via a classical channel

$$p(l'|l): L \to L', \tag{2}$$

and then communicate the fake variable $L' = \{l', p(l')\}$ to Alice and Bob, where

$$p(l') = \sum_l p(l', l), \qquad p(l', l) = p(l'|l)p(l). \tag{3}$$

This process projects the private systems $\{A, B\}$ onto the conditional state

$$\rho_{AB|L'} = \mathrm{Tr}_E(\rho_{ABE|L'}), \tag{4}$$

where

$$\rho_{ABE}(l') \equiv \frac{1}{p(l')}\sum_l p(l', l)\rho_{ABE}(l) = \sum_l p(l|l')\rho_{ABE}(l). \tag{5}$$

Notice that, if L' is completely unrelated to L, then Eve realizes a denial of service, being the communication of the fake variable equivalent to tracing over systems $\{A', B'\}$. In other words, for $p(l', l) = p(l')p(l)$, we have $\rho_{AB|L'} = \rho_A \otimes \rho_B$, where $\rho_A \equiv \mathrm{Tr}_{A'}(\rho_{AA'})$ and $\rho_B \equiv \mathrm{Tr}_{B'}(\rho_{BB'})$.

Secret-key rate: General analysis.—After M rounds of the protocol, Alice and Bob will share M copies $(\rho_{AB|L'})^{\otimes M}$. Note that, in general, Alice and Bob do not know anything about the physical process within the UTP; i.e., they do not know the couple $\{\mathbf{T}, L \to L'\}$. For this reason, what they actually get are M copies of an unknown state $\rho_{AB}^{?}$ plus classical information L'. However, by measuring a suitable number M' of these copies, they are able to deduce the explicit form of the conditional state $\rho_{AB|L'}$ for the remaining $N = M - M'$ copies (here M, M' and N are large numbers). Then, by applying local measurements, Alice on her private systems and Bob on his, they are able to extract two correlated classical variables, X and Y. Finally, from these variables, they can derive a shared secret key via the classical techniques of error correction (EC) and privacy amplification (PA). These procedures can be implemented using one-way classical communications between these two parties.

Let us bound the secret-key rate of the protocol. For simplicity we omit here the conditioning on L', so that Eq. (4) simply becomes $\rho_{AB} = \mathrm{Tr}_E(\rho_{ABE})$. It is understood that the final result must be averaged over L'. Independently from its generation, the (generally) mixed state ρ_{AB} can be purified in a pure state $\Phi_{ABe} = |\Phi\rangle\langle\Phi|_{ABe}$ by introducing a suitable system "e" to be assigned to Eve (this is generally larger than the E system considered before). After this purification, the scenario is the one depicted in Fig. 2. Here, for every bipartition of the systems, $\{AB, e\}$, $\{Ae, B\}$, or $\{Be, A\}$, the corresponding reduced states have the same von Neumann entropy. In particular, we have $S(\rho_{AB}) = S(\rho_e)$.

Now suppose that Alice performs a POVM $\mathcal{M}_A = \{\hat{A}(x)\}$ on her system A with classical outcome x. This measurement projects Φ_{ABe} onto the conditional state

$$\Phi_{Be}(x) = \frac{1}{p(x)}\,\mathrm{Tr}_A[\hat{A}(x)\Phi_{ABe}\hat{A}(x)^\dagger], \tag{6}$$

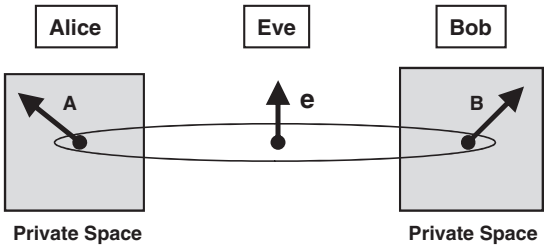

FIG. 2. Purified scenario.

where

$$p(x) = \mathrm{Tr}_{ABe}[\hat{A}(x)\Phi_{ABe}\hat{A}(x)^\dagger]. \quad (7)$$

Thus Alice encodes the stochastic variable $X = \{x, p(x)\}$ in the nonlocal ensemble $\mathcal{E}_{Be} \equiv \{\Phi_{Be}(x), p(x)\}$. Given the conditional state $\Phi_{Be|X}$ of Eq. (6), Bob and Eve can only access their local states, respectively, given by

$$\rho_B(x) = \mathrm{Tr}_e[\Phi_{Be}(x)], \qquad \rho_e(x) = \mathrm{Tr}_B[\Phi_{Be}(x)]. \quad (8)$$

Thus, on his side, Bob has the ensemble $\mathcal{E}_B \equiv \{\rho_B(x), p(x)\}$, whose measurement estimates Alice's variable X. Assuming that Bob has a quantum memory, he can collect all the private systems B associated to the N rounds of the protocol. Then, asymptotically for $N \to \infty$, Bob can reach the Holevo bound [17]

$$I(X\!:B) = S(\rho_B) - \sum_x p(x)S[\rho_B(x)]. \quad (9)$$

At the same time, Eve's information is bounded by

$$I(X\!:e) = S(\rho_e) - \sum_x p(x)S[\rho_e(x)]. \quad (10)$$

Assuming one-way CCs from Alice to Bob (for implementing EC and PA), we can write the secret-key rate as a difference of Holevo informations [18], i.e.,

$$R = I(X\!:B) - I(X\!:e). \quad (11)$$

If we now assume that Alice's POVM is rank one, then the conditional state $\Phi_{Be|X}$ is pure and, therefore, $\rho_{B|X}$ and $\rho_{e|X}$ have the same entropy, i.e., $S[\rho_B(x)] = S[\rho_e(x)]$. As a consequence, we can write

$$R = S(\rho_B) - S(\rho_e) = S(\rho_B) - S(\rho_{AB}) = I(A\rangle B), \quad (12)$$

where $I(A\rangle B)$ is the coherent information between Alice and Bob. Thus the secret-key rate is lower bounded by the entanglement-distillation rate.

Secret-key rate: Detailed analysis.—Here we make a more detailed analysis which is more closely connected to the scenario of Fig. 1. In fact, the rate R of Eq. (12) comes from the general configuration of Fig. 2, which is independent from the actual process generating the final state of Alice and Bob. If we explicitly consider the peculiarities of the scheme of Fig. 1, then we could achieve a larger rate $R^* \geq R$. This new rate can be achieved if Alice and Bob have some knowledge of the classical unreliability of the UTP, i.e., of the amount of information which is "absorbed" by the classical channel $L \to L'$. Thus, if Eve tries to tamper with the overall security by employing fake CCs, then Alice and Bob can potentially extract a secret-key with rate larger than the entanglement-distillation rate.

In this section, we take the different conditionings (by L and L') explicitly into account. After the CC of $L' = \{l', p(l')\}$, Alice and Bob possess the conditional state $\rho_{AB}(l')$ of Eq. (4). Let us assume that Alice performs a POVM $\mathcal{M}_A = \{\hat{A}(x)\}$ on her system A with classical outcome x. This generates the doubly conditional state

$$\rho_B(x, l') = \frac{1}{p(x|l')}\,\mathrm{Tr}_A[\hat{A}(x)\rho_{AB}(l')\hat{A}(x)^\dagger], \quad (13)$$

where

$$p(x|l') = \mathrm{Tr}_{AB}[\hat{A}(x)\rho_{AB}(l')\hat{A}(x)^\dagger]. \quad (14)$$

Averaging over the CCs, the output of Alice's measurement is the unconditional variable $X = \{x, p(x)\}$, where

$$p(x) = \sum_{l'} p(x|l')p(l') = \mathrm{Tr}_A[\hat{A}(x)\rho_A\hat{A}(x)^\dagger]. \quad (15)$$

This is the secret variable to be estimated by Bob. In his private system B, Bob has the ensemble

$$\mathcal{E}_B = \{p(x, l'), \rho_B(x, l')\}, \quad (16)$$

where $p(x, l') = p(x|l')p(l')$. Clearly, this ensemble depends on both X and L'. Exploiting his knowledge of L', Bob applies a conditional measurement $\mathcal{M}_{B|L'}$ to his system B which estimates the value x encoded by Alice. Asymptotically (i.e., for $N \to \infty$), using a quantum memory and averaging over the CCs (i.e., over L'), Bob can reach the conditional Holevo information [19]

$$I(X\!:B|L') = \sum_{l'} p(l')I(X\!:B|L' = l'). \quad (17)$$

For Eve we have to consider the different conditioning given by L. Thus, the conditional state that Eve shares with Alice is

$$\rho_{AE|L} = \mathrm{Tr}_B(\rho_{ABE|L}), \quad (18)$$

which becomes $\rho_{E|XL}$ after Alice's projection. Explicitly this state is given by

$$\rho_E(x, l) = \frac{1}{p(x|l)}\,\mathrm{Tr}_A[\hat{A}(x)\rho_{AE}(l)\hat{A}(x)^\dagger], \quad (19)$$

where

$$p(x|l) = \mathrm{Tr}_{AB}[\hat{A}(x)\rho_{AE}(l)\hat{A}(x)^\dagger]. \quad (20)$$

Thus, Eve has the ensemble

$$\mathcal{E}_E = \{p(x, l), \rho_E(x, l)\}, \quad (21)$$

where $p(x, l) = p(x|l)p(l)$. Asymptotically, Eve can eavesdrop $I(X\!:E|L)$ bits per copy [20].

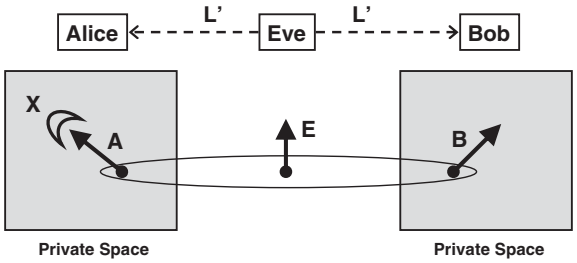

FIG. 3. Conditional state $\rho_{ABE|L'}$ projected onto $\rho_{BE|XL'}$.

As a result, we can write the secret-key rate

$$R^* = I(X:B|L') - I(X:E|L). \quad (22)$$

This quantity can be rewritten as $R^* = R' + \Delta$, where

$$R' \equiv I(X:B|L') - I(X:E|L'), \quad (23)$$

and $\Delta \equiv I(X:E|L') - I(X:E|L)$, quantifies the information which is absorbed by the classical channel $L \to L'$. We call Δ the "classical cheating" by Eve. Clearly, we have $\Delta = 0$ for $L' = L$. R' is the "apparent rate", which refers to the apparent scenario where Alice, Bob and Eve are all subject to the same conditioning L'. In other words, R' is computed assuming the total state $\rho_{ABE|L'}$, which is then projected onto $\rho_{BE|XL'}$ by Alice's measurement (see Fig. 3).

We can now easily prove that the secret-key rate is larger than the entanglement-distillation rate. We have the following result (see Ref. [13] for the proof).

Theorem.—Suppose that Eve measures the incoming systems but cheats on the results using a classical channel $L \to L'$. Then, Alice and Bob's secret-key rate satisfies

$$R^* \geq I(A\rangle B|L') + \Delta, \quad (24)$$

where $I(A\rangle B|L')$ is the coherent information conditioned to Eve's fake variable L', and Δ is the classical cheating.

Our analysis leaves an intriguing open question. It would be wonderful to provide an explicit example where simultaneously $\Delta > 0$ and $I(A\rangle B|L') = 0$, so that $R^* > 0$. This would imply secret-key distillation without entanglement distillation. More generally, we cannot exclude the possibility that $R^* > I(A\rangle B|L')$ by using POVMs which are not rank one.

Conclusion.—We have shown that virtual channels may replace real channels in the QKD setting so as to remove any possibility of side-channel attacks. In its simplest setting, our QKD protocol corresponds to an entanglement swapping experiment, where the dual teleportation channels act as ideal Hilbert space filters to wipe out side-channel attacks. The authenticated users' private spaces are designed so that any incoming quantum signal is topologically excluded from access to detectors, detector settings or state-generation settings, thus side-channel probing attacks of the private spaces are eliminated. Finally, an external untrusted party performs a suitable LOCC (such as a Bell-state measurement) to create correlations necessary for shared key generation.

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
[2] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[3] M. Hillery, Phys. Rev. A **61**, 022309 (2000).
[4] N. J. Cerf, M. Lévy, and G. Van Assche, Phys. Rev. A **63**, 052311 (2001).
[5] F. Grosshans *et al.*, Nature (London) **421**, 238 (2003).
[6] A. M. Lance *et al.*, Phys. Rev. Lett. **95**, 180503 (2005).
[7] V. Scarani *et al.*, Rev. Mod. Phys. **81**, 1301 (2009).
[8] SECOQC, 2007, http://www.secoqc.net.
[9] N. Gisin *et al.*, Rev. Mod. Phys. **74**, 145 (2002).
[10] N. Lütkenhaus and A. J. Shields, New J. Phys. **11**, 045005 (2009).
[11] B. Qi *et al.*, Quantum Inf. Comput. **7**, 73 (2007); C.-H. F. Fung *et al.*, Phys. Rev. A **75**, 032314 (2007); Y. Zhao *et al.*, Phys. Rev. A **78**, 042333 (2008); L. Lydersen *et al.*, Nature Photon. **4**, 686 (2010); L. Lydersen *et al.*, Nature Photon. **4**, 801 (2010); I. Gerhardt *et al.*, Nature Commun. **2**, 349 (2011); L. Lydersen *et al.*, New J. Phys. **13**, 113042 (2011).
[12] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004); J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005); A. Acin, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006); A. Acin *et al.*, Phys. Rev. Lett. **98**, 230501 (2007); N. Gisin, S. Pironio, and N. Sangouard, Phys. Rev. Lett. **105**, 070501 (2010).
[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.108.130502 for a detailed discussion about the private space model and proofs of the theorems presented.
[14] H.-K. Lo and H. F. Chau, Science **283**, 2050 (1999).
[15] E. Biham, B. Huttner, and T. Mor, Phys. Rev. A **54**, 2651 (1996); H. Inamori, Algorithmica **34**, 340 (2002).
[16] Summing over l, we have a completely positive trace preserving (CPTP) map.
[17] A. S. Holevo, Probl. Inf. Transm. **9**, 177 (1973).
[18] I. Devetak and A. Winter, Proc. R. Soc. A **461**, 207 (2005).
[19] Equivalently, we can adopt the EHS representation [13], where the ensemble $\mathcal{E}_B$ and the stochastic variables X and L' are described by a unique classical-quantum state $\rho_{\mathbf{XL'}B} = \sum_{x,l'} p(x,l')|x\rangle\langle x|_{\mathbf{X}} \otimes |l'\rangle\langle l'|_{\mathbf{L'}} \otimes \rho_B(x,l)$. The Holevo quantity of Eq. (17) corresponds to the conditional quantum mutual entropy $I(\mathbf{X}:B|\mathbf{L'})$ computed over this state.
[20] Equivalently, we can consider the classical-quantum state $\rho_{\mathbf{XL}E} = \sum_{x,l} p(x,l)|x\rangle\langle x|_{\mathbf{X}} \otimes |l\rangle\langle l|_{\mathbf{L}} \otimes \rho_E(x,l)$, and compute $I(\mathbf{X}:E|\mathbf{L}) = I(X:E|L)$.