Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems

Yi Zhao, Chi-Hang Fred Fung, Bing Qi, Christine Chen, and Hoi-Kwong Lo
Center for Quantum Information and Quantum Control, Department of Physics and Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, M5S 3G4, Canada
(Received 8 August 2007; published 28 October 2008)

Quantum-key-distribution (QKD) systems can send quantum signals over more than 100 km standard optical fiber and are widely believed to be secure. Here, we show experimentally a technologically feasible attack—namely, the time-shift attack—against a commercial QKD system. Our result shows that, contrary to popular belief, an eavesdropper, Eve, has a non-negligible probability (~4%) to break the security of the system. Eve's success is due to the well-known detection efficiency loophole in the experimental testing of Bell's inequalities. Therefore, the detection efficiency loophole plays a key role not only in fundamental physics, but also in technological applications such as QKD systems.

DOI: 10.1103/PhysRevA.78.042333 PACS number(s): 03.67.Dd, 03.67.Hk

Quantum key distribution (QKD) [1,2] provides a method to share a secret key between legitimate users called "Alice" (the sender) and "Bob" (the receiver). The unconditional security of QKD has been rigorously proved based on the laws of physics [3,4]. Even imperfect practical QKD systems have also been proved secure assuming some semirealistic models [5,6]. The decoy method [7] was proposed to dramatically improve the performance of a practical QKD system. Our group has implemented the decoy method experimentally over 15 and 60 km of telecom fibers [8]. Incidentally, QKD has found real-life applications in a recent Swiss election [9].

Recently, there has been a lot of theoretical interest in the connection between the security of QKD and fundamental physical principles such as violation of Bell's inequality and the no-signaling constraint [10] on spacelike observables. An ultimate goal, which has not yet been achieved [11], is to construct a device-independent security proof. As is well known, the experimental testing of Bell's inequalities often suffers from the detection efficiency loophole. Nonetheless, a fair sampling assumption may save the day [12]. However, as we will demonstrate below, the low detection efficiency of practical detectors not only violates the fair sampling assumption that would be needed in security proofs based on Bell-inequality violation, but also gives Eve (an eavesdropper) a powerful handle to break the security of a practical QKD system. Therefore, the detection efficiency loophole is of both conceptual and practical interest.

Our work is an illustration of general physical limitations, rather than a particular technological weakness. Indeed, a practical QKD system often includes two or more detectors. It is virtually impossible to manufacture identical detectors in practice. As a result, the two detectors of the same QKD system will exhibit different detection efficiencies as functions of either one or a combination of variables in the time, frequency, polarization, and/or spatial domains. If Eve manipulates a signal in these variables, she could effectively exploit the detection efficiency loophole to break the security of a QKD system. In our experiment, we consider Eve's manipulation of the time variable. Our work demonstrates the general problem of the detection efficiency loophole in practical QKD systems. Note that large detection efficiency mismatch suggests low detection efficiency because the detection efficiency mismatch will be minimal if all the detectors have very high (i.e., close to 1) efficiencies.

Recently, quantum hacking has attracted much scientific and popular attention [13]. Makarov *et al.* proposed a faked-state attack and studied its feasibility with their homemade QKD system [14,15]. Unfortunately, this attack is an intercept-resend attack, which is hard to implement in practice. Therefore, this attack has never been successfully demonstrated in experiments. Kim *et al.* simulated an entanglement probe attack on the Bennett-Brassard 1984 (BB84) protocol [16]. However, it serves to demonstrate the security rather than the insecurity of QKD systems because this attack has already been considered in standard security proofs. A study of the information leakage due to public announcement of the timing information by Bob was reported [17]. However, Bob does not need to make such an announcement in practice. In summary, despite numerous efforts, up until now, no one has even come close to hacking successfully a practical QKD system, let alone a commercial one.

Here, we present an experimental demonstration of a successful hacking against a commercial QKD system. It is highly surprising to break a well-designed commercial QKD system with only *current* technology. Our work shows clearly the slippery nature of QKD systems and forces us to reexamine the security of practical QKD systems and their applications in real life. The attack we use is the time-shift attack proposed by us in [18]. The time-shift attack is simple to implement as it does not involve any measurement or state preparation by Eve.

The time-shift attack exploits the detection efficiency mismatch between the two detectors in a QKD system in the time domain. In QKD security proofs (e.g., Ref. [5]), a standard assumption is that the detection efficiencies for the bits "0" and "1" are equal. However, its validity is questionable [14,15,18]. For example, a typical time dependence of the detection efficiency of a practical fiber-based QKD system [with InGaAs avalanche photodiodes (APDs) of telecom wavelength operating at gated Geiger mode] is illustrated in Fig. 1(a). Note that, at time A, the detection efficiency for the bit "0" is much higher than that for the bit "1," while the

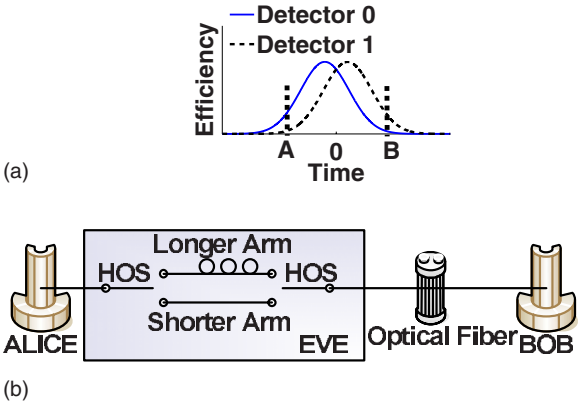

FIG. 1. (Color online) Conceptual drawings. (a) Conceptual efficiency mismatch. (b) A conceptual schematic of Eve's attack. HOS, high-speed optical switch.

opposite case can be found at B. The detection efficiency mismatch can only be confidently removed if the efficiencies are constant in the time domain. We remark that even nongated single-photon detectors such as Si APDs exhibit detection efficiency mismatch due to intrinsic dead time [19].

The idea of the time-shift attack is simple. Eve can shift the arrival time of each signal to either A or B randomly with probabilities p_A and $p_B=1-p_A$, respectively. Eve can carefully choose p_A to keep the number of Bob's detection events of 0's and 1's equal. Since Bob's measurement result will be biased toward 0 or 1 depending on the time shift (A or B), Eve can "steal" information without alerting Alice or Bob. A conceptual setup to launch the time-shift attack is shown in Fig. 1(b). Eve can choose to connect Alice and Bob through either a longer arm or a shorter arm so as to shift the signal around time A (a negative shift), or around time B (a positive shift).

The success of our demonstration is a big surprise because in our experiment, Eve *cannot* perform a quantum nondemolition (QND) measurement on the photon number or compensate any loss introduced by the attack, while Eve *can* have arbitrarily advanced technology in security proofs. In other words, our practical Eve is much weaker than the eavesdropper in security proofs. It is surprising to see an attack which can be implemented with current technology (e.g., the time-shift attack) do better than even the QND attack, which is significantly beyond current technology.

The experiment is performed on top of a modified commercial ID-500 QKD setup [20] manufactured by id Quantique. The schematic of our experimental setup is shown in Fig. 2.

The crucial issues in the experiment are the activation times of the two detectors (APDs in Fig. 2). The commercial QKD system has a built-in calibration program which sets the activation time of each detector independently. The activation times of the two detectors differ slightly due to the discrepancies in the lengths of the fibers connecting them. Ideally, to minimize the detection efficiency mismatch, the difference of the activation times should take a constant value. However, at times the difference in the activation times as set by the built-in calibration program deviates from this value, suggesting a larger efficiency mismatch. We observed the maximum value of the deviations as $\Delta\sim 100$ ps. To get statistics of this deviation, we ran the built-in calibration program for 2844 times, during which the deviation reaches Δ for 106 times. That is, the detection efficiency mismatch reaches its maximum value with a probability of $\sim 4\%$.

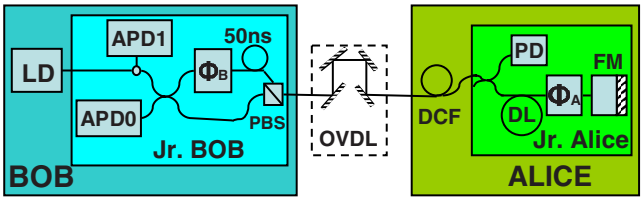

FIG. 2. (Color online) The schematic of experimental demonstration of the time-shift attack. Inside Jr. Bob (Jr. Alice): components in Bob's (Alice's) package of id Quantique QKD system. Our modifications: LD, narrow pulse laser diode; OVDL, optical variable delay line; DCF, dispersion compensating fiber. Original QKD system: APD, avalanche photon diode; $\Phi_{A/B}$, phase modulator; PBS, polarizing beam splitter; PD, classical photodetector; DL, delay line; FM, Faraday mirror.

After the calibration of the activation times, we use the optical variable delay line (OVDL in Fig. 2) to manually shift the arrival time of the signals, looking for instants that show large efficiency mismatch.

There are several challenges in this experiment. In our setup, the gating window for the detectors (APDs in Fig. 2) is $\sim$500 ps, which is also the laser pulse width. This will "blur" the efficiency mismatch. However, the commercial QKD system is not immune from the time-shift attack as Eve can simply apply the standard pulse compression technique to the bright pulses sent from Bob to Alice in the channel (e.g., [21]). In our experiment, we replaced the original laser source by a PicoQuant laser diode (LD in Fig. 2) with pulse width $\sim$100 ps, which is equivalent to the compression scheme mentioned above [22].

Another challenge is the chromatic dispersion in the fiber which broadens the laser pulses. We thus installed $\sim$2-km dispersion compensating fiber (DCF in Fig. 2). Ideally, Eve can prechirp the bright pulses that are sent from Bob to Alice. Note that both the prechirping and pulse compression can be done on the bright pulses from Bob to Alice without touching the quantum signal sent from Alice to Bob. Therefore, neither process would increase the channel loss when Alice sends quantum signals to Bob. We thus view the dispersion compensating fiber (DCF in Fig. 2) as part of Alice's local apparatus.

A third challenge is the optimization of the attack. Naively, Eve could simply select large shifts as they would definitely provide substantial intrinsic detection efficiency mismatches. However, they may be suboptimal for the attack because their low intrinsic detection efficiencies make the dark count significant, increasing the quantum bit error rate (QBER) and consequently the cost of the error correction. Therefore the task of choosing the shifts is nontrivial. The time-shift attack will introduce additional loss as the signals are shifted to the low-efficiency region. Nonetheless, since Alice and Bob's channel may not be a straight line and there may be additional loss due to components such as optical

switches, in practice Eve could lower the channel loss by, for example, replacing the existing channel with a better one without alerting Alice and Bob. Moreover, Alice and Bob are assumed to have no knowledge of the channel loss in standard security proofs [3–6]. If this assumption is violated, the secure key rate could be much higher. It is not rigorous to allow Alice and Bob to trust the loss of their channel since it is inconsistent with the security proofs [3–6]. Naively, one might think that Alice and Bob can catch Eve by observing an increase in channel loss during the quantum transmission phase (relative to the calibration phase). Such a naive thinking is incorrect because Eve may well be present during both the quantum transmission phase and the calibration phase. Therefore, Alice and Bob should not be able to see any difference in the channel loss in the two phases. So the power of the time-shift attack may be stronger than what one naively thinks.

We demonstrated the time-shift attack in the following way: first, the activation times of detectors were determined by the built-in program; second, the arrival times of the signals were shifted at a step of 50 ps (a narrower step was not necessary as the pulse width was ~100 ps); third, at each shifted time, Alice and Bob exchanged a key at an average photon number (at Alice's output) of 0.1; fourth, Bob calculated the counts of each detector and the error rates. The entire experiment after each calibration spanned ~15 min.

In a real attack Eve should apply an alternative technique to obtain the efficiency mismatch as she has no access to Bob's apparatus [18]: she can gradually shift a small subset of the signals and set them to 0 or 1 and conclude the mismatch from Bob's detection announcement. Our experimental results show that the mismatch is stable throughout the 15-min span of our experiment. Therefore Eve has sufficient time to obtain the mismatch information and launch her attack.

The experimentally measured detector efficiencies are shown in Fig. 3 for the case where the deviation in activation times takes the maximal value Δt_{m}. It shows substantial detection efficiency mismatch. In particular, two shifts with large mismatches are found as in Table I.

The security of the QKD system is analyzed in the following way: one can estimate an upper bound K_{U} of the key length *given* the efficiency mismatch known by Eve and a lower bound K_{L} *ignoring* the time-shift attack (as Alice and Bob cannot detect the attack). If the upper bound is less than the lower bound (i.e., $K_{\mathrm{L}} > K_{\mathrm{U}}$), there must be some information leaked to Eve unknown to Alice or Bob.

We consider that Alice sends N bits to Bob, among which the same bases are used for $\tilde{N}$ bits and Bob detects $\tilde{N}Q$ signals (Q is the overall gain). Here we assume that infinite decoy-state protocol and one-way classical communications for post-processing are used.

Lower bound. The error correction will consume

$$r_{\mathrm{EC}} = \tilde{N} Q f(E) H_2(E) \tag{1}$$

bits, where E is the overall QBER, $H_2(x)$ is the standard binary Shannon entropy function, and $f(x)$ is the error correction inefficiency [23]. The net key length ignoring the time-shift attack is thus [5,24–26]

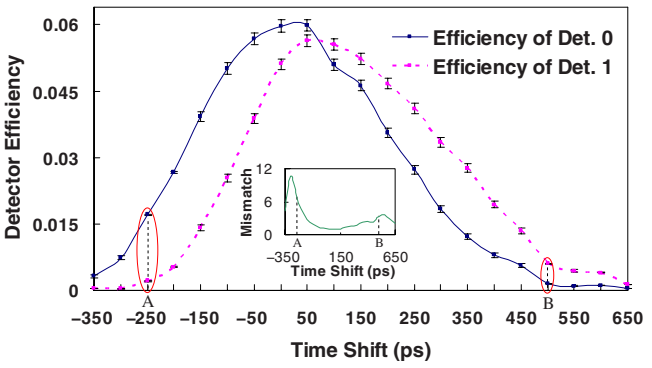

FIG. 3. (Color online) Efficiencies of the two detectors versus time shifts. Inset: the mismatch of detector efficiencies [defined as $\max(d_0/d_1, d_1/d_0)$]. The peak efficiencies of detectors are slightly different, suggesting the detection efficiency has slightly drifted since the factory setting. The data size for time shifts with large detection efficiency mismatch (−250 ps, −200 ps, 500 ps, 600 ps, and 650 ps) is chosen to be 20.97 Mbit to acquire accurate mismatch, while the data size for other shifts is chosen to be 1.05 Mbit to speed up the experiment.

$$K_{\mathrm{L}} = -r_{\mathrm{EC}} + \tilde{N}\{Q_1[1 - H_2(e_1)] + Q_0\}, \tag{2}$$

where Q_i and e_i are the gain and the QBER for the signals with i photons sent by Alice.

Upper bound. An upper bound is given by [27]

$$K_{\mathrm{U}} = -r_{\mathrm{EC}} + \tilde{N} Q \sum_{i=\{A,B\};j=\{0,1\}} [\Pr\{Z_2 = j | Z_1 = i\} \Pr\{Z_1 = i\} \times H_2(\Pr\{X = 0 | Z_1 = i, Z_2 = j\})], \tag{3}$$

where X, Z_1, and Z_2 are classical random variables representing Alice's initial bit, Eve's choice of the time shift for each bit, and the basis information, respectively.

The upper bound and the lower bound of the key rate can then be calculated from Eqs. (1)–(3) using data in Table I. The calculation results are shown in Table I(c). Y_0 is determined experimentally. Note that no double clicks were observed in our experiment. The fact that $K_{\mathrm{L}} > K_{\mathrm{U}}$ clearly indicates the success of the attack [28].

We conclude with a few general lessons. First, countermeasures often exist for known attacks. For instance, the "four-state setting" proposal (which suggests that for the phase-encoding BB84 protocol, Bob's phase modulation is randomly selected from a set of four values instead of two values) can shield the time-shift attack [29]. Second, the implementation of a countermeasure may open up new security loopholes. For instance, we noted in [30] that the four-state measurement scheme will be vulnerable to a combined large pulse [31] and time-shift attack. Once an attack is known, the prevention is usually easy. However, we have a third lesson: unanticipated attacks are most dangerous.

The time-shift attack is demonstrated on a bidirectional system. However, it is a threat to a general class of QKD systems (including unidirectional setup) and protocols (e.g.,

TABLE I. Experimental results. (a) The number of detections. (b) The number of detections given that Alice and Bob use the same basis. $\tilde{N}=10\,481\,280$ bits. Y is Bob's bit value. (c) Parameters for computing the key length.

(a)

Label	Shift (ps)	d_0	d_1	N
A	−250	10992	1541	20 966 400
B	500	1231	4059	20 966 400

(b)

Time shift A (−250 ps)				Time shift B (500 ps)			
Z_2	X	$Y=1$	$Y=0$	Z_2	X	$Y=1$	$Y=0$
0	1	336	139	0	1	979	31
0	0	65	2557	0	0	41	260
1	1	333	120	1	1	1022	37
1	0	59	2634	1	0	35	279
QBER: 0.06135				QBER: 0.05365			

(c)

Theoretical		Experimental					
$f(x)$	p_A	μ	Y_0	$d_{0/1}$	E	K_U	K_L
1.22	23.0%	0.1	2.26×10^{-5}	3479	5.68%	1131 bit	1297 bit

[2]). Moreover, we are concerned with the general physical limitations of the detection efficiency loophole, rather than a specific technological problem. The time-shift attacks can be easily generalized to spatial-, spectral-, and polarization-shift attacks exploiting the efficiency mismatch in the corresponding domains [29]. On the practical side, our work highlights the significance of side-channel attacks [32,33] in QKD. Historically, the existence of a side-channel attack went back to the first QKD experiment, which was unconditionally secure to any eavesdropper who happens to be deaf [34].

The time-shift attack, like any other quantum hacking attack, was demonstrated on a particular implementation of QKD. Therefore, it may not directly apply to all QKD systems. However, any QKD is done on a particular implementation. If we cannot trust a particular implementation of QKD, one should never use QKD in the first place. The plug-and-play structure that we have successfully attacked was the most widely used commercial system, i.e., it is a standard system. It is thus unclear if any existing commercial system is secure from unanticipated attacks because of the detection efficiency mismatch problem. Indeed, we emphasize that the detection efficiency mismatch is a very general problem. It is hard to build two identical detectors. Nonetheless, the security proof for a QKD system with detection efficiency mismatch was recently devloped [12].

In summary, we report an experimental demonstration of a technologically feasible attack against a commercial QKD system. Our results clearly show that even QKD systems built by *trustworthy* manufacturers may contain subtle flaws that will allow Eve to break it with current technologies. The success of the attack highlights the importance to battle-test practical QKD systems and work on security proofs with *testable* assumptions. It is remarkable that the detection efficiency loophole plays a key role in both fundamental physics and technological applications (e.g., QKD systems) [33]. How to close the detection efficiency loophole and side-channel attacks will be an important subject for future investigations.

ACKNOWLEDGMENTS

We acknowledge generous help from A. Kurtsiefer, C. Lamas-Linares, X. Ma, Vadim Makarov, and id Quantique. Support of the funding agencies CFI, CIPI, the CRC program, CIFAR, MEDT, MITACS, NSERC, Perimeter Institute, QuantumWorks, OIT, CQIQC, and PREA is gratefully acknowledged.

[1] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), pp. 175–179.
[2] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[3] D. Mayers, J. ACM **48**, 351 (2001); H.-K. Lo and H. F. Chau, Science **283**, 2050 (1999); P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).
[4] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996).
[5] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **4**, 325 (2004).
[6] H. Inamori, N. Lütkenhaus, and D. Mayers, Eur. Phys. J. D **41**, 599 (2007).
[7] W. Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003); H.-K. Lo, X. Ma, and K. Chen, *ibid.* **94**, 230504 (2005); X.-B. Wang, *ibid.* **94**, 230503 (2005); X. Ma, B. Qi, Y. Zhao, and H. K. Lo, Phys. Rev. A **72**, 012326 (2005); X.-B. Wang, *ibid.* **72**, 012322 (2005).
[8] Y. Zhao, B. Qi, X. Ma, H. K. Lo, and L. Qian, Phys. Rev. Lett. **96**, 070502 (2006); Y. Zhao *et al.*, in *Proceedings of IEEE International Symposium on Information Theory* (IEEE, New York, 2006), pp. 2094–2098.
[9] Economist, Oct 18, 2007 issue.
[10] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[11] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[12] P. M. Pearle, Phys. Rev. D **2**, 1418 (1970).
[13] G. Brumfiel, Nature (London) **447**, 372 (2007).
[14] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006).
[15] V. Makarov and J. Skaar, Quantum Inf. Comput. **8**, 0622 (2008).
[16] T. Kim, I. Stork genannt Wersborg, F. N. C. Wong, and J. H. Shapiro, Phys. Rev. A **75**, 042327 (2007).
[17] A. Lamas-Linares and C. Kurtsiefer, Opt. Express **15**, 9388 (2007).
[18] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **7**, 73 (2007).
[19] D. J. Rogers, J. C. Bienfang, A. Nakassis, H. Xu, and C. W. Clark, New J. Phys. **9**, 319 (2007).
[20] D. Stucki, N. Gisin, O. Guinnard, G. Robordy, and H. Zbinden, New J. Phys. **4**, 41 (2002).
[21] B. Kibler *et al.*, Electron. Lett. **43**, 915 (2007).
[22] Pulse compression is a standard experimental technique that has been routinely performed in industries for decades. It would be of implementational interest to see the impact of the pulse compression on the time-shift attack, though not of physics interest.
[23] G. Brassard and L. Salvail, *Lecture Notes in Computer Science* (Springer, Berlin, 1994), Vol. 765, pp. 410–423.
[24] H.-K. Lo, Quantum Inf. Comput. **5**, 413 (2005).
[25] M. Koashi, J. Phys.: Conf. Ser. **36**, 98 (2006).
[26] M. Ben-Or, M. Horodecki, D. W. Leung, D. Mayers, and J. Oppenheim, in *Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005*, edited by J. Kilian, Lecture Notes in Computer Science Vol. 3378 (Springer-Verlag, Berlin, 2005), pp. 386–406.
[27] R. Renner and R. König, in *Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005*, edited by J. Kilian, Lecture Notes in Computer Science Vol. 3378 (Springer-Verlag, Berlin, 2005), pp. 407–425.
[28] The most general security analysis suggests $K_L=1297$ bits secret key. However, the maximum length of the secret key is $K_U=1131$ bits due to the time-shift attack. Therefore (loosely speaking), Eve can successfully decrypt *no less than* 166 bits from the secret key (which is wrongly presumed to be unconditionally secure).
[29] C.-H. F. Fung *et al.*, e-print arXiv:0802.3788.
[30] H.-K. Lo, talk delivered at "Theory and Realisation of Quantum Key Distribution" Workshop at IQC, Waterloo, June 2007, slides avaliable on-line at http://www.iqc.ca/quantumworld/index.php?id=3$pid=21.
[31] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Phys. Rev. A **73**, 022320 (2006).
[32] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, Jet Cryptogr. **5**, 3 (1992).
[33] R. Alleaume *et al.*, e-print arXiv:quant-ph/0701168.
[34] G. Brassard, *Proceedings of IEEE Information Theory Workshop on Theory and Practice in Information Theoretic Security*, Awajl Island, Japan, pp. 19–23, October 2005, e-print arXiv:quant-ph/0604072.