Trustworthiness of detectors in quantum key distribution with untrusted detectors

Bing Qi

Phys. Rev. A **91**, 020303 — Published 25 February 2015

DOI: 10.1103/PhysRevA.91.020303

Trustworthiness of detectors in quantum key distribution with untrusted detectors

Bing Qi[1, 2, *]

[1] *Quantum Information Science Group, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6418, USA*

[2] *Department of Physics and Astronomy, The University of Tennessee, Knoxville, TN 37996 - 1200, USA*

(Dated: January 23, 2015)

Measurement-device-independent quantum key distribution (MDI-QKD) protocol has been demonstrated as a viable solution to detector side-channel attacks. Recently, to bridge the strong security of MDI-QKD with the high efficiency of conventional QKD, the detector-device-independent (DDI) QKD has been proposed. One crucial assumption made in DDI-QKD is that the untrusted BSM located inside the receiver's laboratory cannot send any unwanted information to the outside. Here, we show that if the BSM is completely untrusted, a simple scheme would allow the BSM to send information to the outside. Combined with Trojan horse attacks, this scheme could allow an eavesdropper to gain information of the quantum key without being detected. To prevent the above attack, either countermeasures to Trojan horse attacks or some trustworthiness to the "untrusted" BSM device is required.

PACS numbers: 03.67.Dd

Introduction. Quantum key distribution (QKD) allows two authenticated users, normally referred to as Alice and Bob, to generate a private key through an insecure quantum channel controlled by an eavesdropper, Eve [1–5]. Idealized QKD protocols have been proved to be unconditionally secure against adversaries with unlimited computing power and technological capabilities [6]. However, practical implementations of QKD unavoidably contain imperfections which may be overlooked in the security proofs. The disconnection between QKD theory and its implementations has led to various "side-channel" attacks [7–10].

One important approach to enhance the security of practical QKD is to develop QKD protocols based on "untrusted" device [11–15]. Among them, the measurement-device-independent (MDI) QKD protocol [14], has received much attention [16]. The MDI-QKD protocol is automatically immune to all side-channel attacks associated with the measurement device which, arguably, is the weakest link in a QKD system. The feasibility of MDI-QKD has been demonstrated experimentally [17]. See [18] for a recent review.

Recently, to bridge the strong security of MDI-QKD with the high efficiency of conventional QKD, the detector-device-independent (DDI) QKD has been proposed by several groups [19–21]. In this paper, we scrutinize the underlying assumptions behind DDI-QKD. One crucial assumption is that the untrusted Bell state measurement (BSM) located inside the receiver's laboratory cannot send any "unwanted" information to the outside. Here, we show that if the BSM is completely untrusted, a simple scheme would allow the BSM to send information to the outside without being detected: Eve can place high-efficiency detectors inside the BSM and program it to selectively report a fraction of the total detection events. The time delay between adjacent reported events can be used by the BSM to send information. Combined with Trojan horse attacks, this scheme could allow Eve to gain information of the secure key without introducing any errors. Our results suggest that to establish the security of DDI-QKD, additional assumptions on the measurement device are required. It is thus very important to clearly spell out those assumptions and place them under scrutiny.

DDI-QKD. In a conventional QKD protocol (see Fig.1a), Alice prepares quantum states and sends them to Bob through an insecure quantum channel, while Bob performs measurement. In this configuration, it is reasonable to assume that the errors in quantum state preparation can be well controlled and quantified, since this can be done within Alice's well protected laboratory without Eve's interference. On the contrary, the quantum states received by Bob are highly unpredictable. Eve can interfere the measurement process by either manipulating Alice's signal [8] or sending her own signals to Bob [9]. The above observation could explain why most identified security loopholes in conventional QKD are associated with the measurement device [7–9].

In MDI-QKD (see Fig.1b), both Alice and Bob prepare quantum states and send them to an untrusted third party, Charlie, who could be a collaborator of Eve. Charlie is supposed to measure the correlation between Alice's and Bob's quantum states and publicly announce the measurement results. Given Charlie's measurement results, Alice and Bob can further establish a secure key. The protocol has been designed in such a way that "only" the correlation (but not the quantum states themselves) can be determined by Charlie faithfully. On one hand, if Charlie executes the protocol honestly, he or Eve cannot gain any information of the secure key. On the other hand, any attempts by Charlie to gain information of the secure key will unavoidably introduce additional noise and can be detected. By allowing Eve to fully control

[*] qib1@ornl.gov

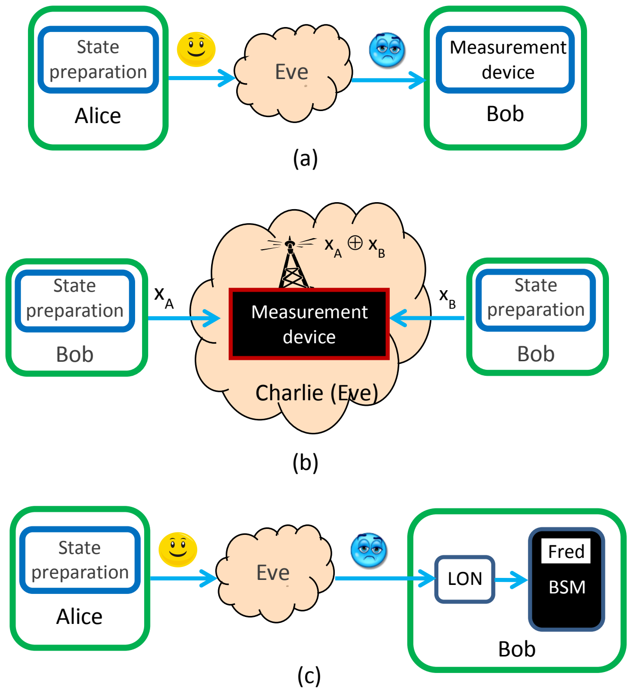

FIG. 1: (a) Conventional QKD; (b) Measurement-device-independent(MDI) QKD; (c) Detector-device-independent (DDI)QKD

the measurement device, the security of MDI-QKD can be established without making any assumptions about the measurement device, thus removing any potential detector side-channels. The security of MDI-QKD is based on the idea of time-reversed EPR QKD [22, 23].

In DDI-QKD (Fig.1c), Alice prepares a single-photon pulse in one of the four BB84 polarization states and sends it to Bob. On receiving the signals, Bob employs a "trusted" linear optics network (LON) to encode his information on a different (for example, spatial) degree of freedom of the incoming signals. Afterwards, the above signals are fed into an "untrusted" BSM device which is supposed to perform a single-photon BSM [24] and report the measurement results to Bob. Through an authenticated classical channel, Bob announces which transmitted signals have yielded successful BSM detection events, the Bell states obtained, and the basis information associated with those successful events. Alice and Bob estimate the quantum bit error rate (QBER) for the events when they happen to use the same basis. If the QBER is below certain threshold, they can further apply error correction and privacy amplification to generate a secure key. Note in Fig.1c, the untrusted BSM inside Bob's laboratory is represented by a "black box", which is controlled by Eve's partner (Fred) during the QKD process.

Proposed attack. In DDI-QKD, Bob's encoding device (LON in Fig.1c) sits between Eve and Fred. In practice, this design could be prone to Trojan horse attacks. We remark that the threats of Trojan horse attacks in conventional QKD and the corresponding countermeasures have been studied previously [25]. In conventional QKD, Eve can only access one end of each user's device. To launch a Trojan horse attack, Eve can send bright light pulses into the user's system and tries to gain information by measuring the back-reflected light. In practice, the QKD users can effectively reduce the risk of Trojan horse attack by using filters, optical circulators, isolators, and intensity monitors, etc. However, these countermeasures may not be effective in the case of DDI-QKD, where Eve and Fred together can access both ends of Bob's encoding device. For example, in each quantum transmission, Eve could send her own signal (which may contain a few photons) together with Alice's photon into Bob's laboratory. Both Alice's and Eve's signals go through Bob's encoding device and reach the BSM. Inside the BSM, Fred could determine Bob's bit information precisely by measuring the signal sent by Eve. In the meantime Fred can perform an honest Bell state measurement on Alice's photon. In principle, Fred can have a perfect copy of Bob's random bits without introducing any errors.

At this point, security is not compromised since Fred is confined within Bob's laboratory. To prevent Fred from sending Bob's random bits to Eve, a crucial assumption is made in DDI-QKD [19, 20]: Fred is only allowed to report the BSM results to Bob; he cannot send any "unwanted" information to the outside. However, it could be difficult to justify the above assumption in practice. Below we will show that if the BSM is completely untrusted (as in the case of MDI-QKD), a simple scheme would allow the untrusted BSM to send information to Eve. Combined with Trojan horse attacks discussed above, this scheme could allow Eve to gain information of the final key without being detected.

A practical single photon detector (SPD) at telecom wavelength has a relatively low detection efficiency (typically 10% $\sim$ 30%). On the top of that, all the optical components inside a practical BSM introduce additional losses. Under normal operation, Bob would expect a low detection rate: most of the time, the BSM will report "no detection"; occasionally, the BSM will report a successful BSM result. However, Eve could place high-efficiency SPDs inside the BSM. In this case, the actual detection rate seen by Fred can be much higher than the one expected by Bob. Fred can easily simulate a low-efficiency BSM by reporting to Bob a small fraction of the total detection events. He can further take advantage of this "post-selection" process to send information to Eve.

Suppose Fred successfully detects the i^{th} signal sent by Alice; in the meantime, through the Trojan horse attack, he also learns Bob's bit information corresponding to the same transmission. Fred reports the BSM result to Bob honestly. He also determines the index number $i + k$ of the next BSM result to be reported based on the following rules: if Bob's i^{th} bit value is 1 (or 0), Fred will report a BSM result to make sure that k is an even (or odd) number. In other words, Fred encodes Bob's random bits on the time delays between adjacent BSM results reported to Bob. When Bob publicly announces which signals from Alice have been detected, Eve can decode Fred's information and have a perfect copy of Bob's random bits. Since Fred performs BSM on Alice's signals and reports the measurement results to Bob honestly, this attack will not introduce additional errors. To further conceal their attack, Eve and Fred can use pre-shared random numbers to determine whether an even or an odd number is used to encode bit 1 for each reported BSM event. Fred can also carefully control the reporting rate to match it with the one expected by Bob.

This attack, in its spirit, is similar to the memory attack on device-independent (DI) QKD [26], which might be applicable whenever an untrusted device is placed inside Alice's or Bob's secure laboratory. On the other hand, our attack is more feasible since Fred does not need to access the classical communication channel between Alice and Bob. All his activities are confined within the untrusted BSM, as assumed in DDI-QKD.

Discussion. To prevent the above attack, Bob could introduce various countermeasures to detect Trojan horse attacks, as suggested in [21]. In practice, additional filtering and random sampling systems could be introduced into Bob's system to characterize the input signals and mitigate the risk of Trojan horse attacks. However, it is very challenging in practice to implement single-mode filtering. Moreover, since Eve's signals may only contain a few photons, additional SPDs could be required to implement the above countermeasures. These SPDs may suffer from the same side-channel attacks as the SPDs for secure key generation, and may introduce new security loopholes.

Another approach to prevent the above attack is to make additional assumptions about the BSM device [27]. Instead of treating the BSM device as a completely "black box", Bob may approximately know what is inside the BSM. In this case, it may be possible to prove security without perfectly modeling the exact behavior of the BSM. Nevertheless, all the assumptions made about the BSM should be clearly specified and be placed under scrutiny. To highlight this point, we will show that if there are certain overlooked imperfections in the BSM, Eve could launch the detector blinding attack [9] to break the system.

Although the actual implementations of the BSM in [19–21] are slightly different, all of them employ four SPDs to identify the four Bell states. In normal QKD operation, if Alice and Bob use the same basis, only two out of the four SPDs have non-zero probability (50% each in the ideal case) to detect a photon. If they use different bases, all the four SPDs have non-zero probability (25% each in the ideal case) to click. In detector blinding attack [9], Eve first sends bright light into Bob's system to force the SPDs into the linear operation mode. Then she performs an intercept and resend attack: she intercepts Alice's signal, measures it in a randomly chosen basis, and resends a bright pulse to Bob according to her measurement result. The optical power of Eve's bright pulse will be distributed either between two SPDs (if Eve and Bob use the same basis) or among four SPDs (if they use different bases). By carefully controlling the power of the bright pulse, Eve can make sure that Bob registers a detection event only when they use the same basis. Equivalently, Eve has control of Bob's measurement basis and it is easy to show that in principle Eve can learn the whole key without introducing errors.

The above attack can be detected if the BSM is perfect. This is because when Bob and Eve use the same basis, Eve's bright pulse will be evenly distributed between two SPDs and result in an unusually high double-click rate [28]. However, if we allow certain imperfections in the BSM, Eve could refine her attack to make it undetectable. For example, if the SPDs have different wavelength-dependent efficiencies, Eve could reduce the double-click rate by tailoring the wavelength of her bright pulse. To rule out the possibility of the detector blinding attack in DDI-QKD, we need to quantify the imperfections inside the BSM carefully, or introduce other countermeasures.

In summary, we investigate some underlying assumptions in DDI-QKD. Our results show that if the BSM in DDI-QKD is completely untrusted, a simple attack could allow Eve to gain information of the quantum key without being detected. To prevent the above attack, either countermeasures to Trojan horse attacks or some trustworthiness to the "untrusted" BSM device is required. All these details should be clearly specified and included in the security analysis.

Acknowledgments. I would like to thank Marcos Curty for very helpful discussions. This work was performed at Oak Ridge National Laboratory, operated by UT-Battelle for the US Department of Energy under Contract No. DE-AC05-00OR22725.

[1] C. H. Bennett and G. Brassard, in *Proceedings of International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, (IEEE Press, New York, 1984), pp. 175-179.

[2] A. K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991).

[3] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).

[4] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, *Rev. Mod. Phys.* **81**, 1301 (2009).

[5] H.-K. Lo, M. Curty, and K. Tamaki, *Nature Photonics* **8**, 595 (2014).

[6] 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).

[7] V. Makarov and D. R. Hjelme, *J. Mod. Opt.* **52**, 691

(2005); V. Makarov, A. Anisimov, and J. Skaar, *Phys. Rev. A* **74**, 022313 (2006); L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, *Opt. Express* **18**, 27938 (2010); C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, Ch. Marquardt, V. Makarov, and G. Leuchs, *New J. Phys.* **13**, 013043 (2011); I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, *Nature Comm.* **2**, 349 (2011); D. J. Rogers, J. C. Bienfang, A. Nakassis, H. Xu, and C. W. Clark, *New J. Phys.* **9**, 319 (2007); V. Burenkov, B. Qi, B. Fortescue, and H.-K. Lo, *Quantum. Inf. Comput.* **14**, 217 (2014); H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauerth, and H. Weinfurter, *New J. Phys.* **13**, 073024 (2011); H.-W. Li, S. Wang, J.- Z. Huang, W. Chen, Z.- Q. Yin, F-. Y. Li, Z. Zhou, D. Liu, Y. Zhang, G.-C. Guo, W.-S. Bao, and Z.-F. Han, *Phys. Rev. A* **84**, 062308 (2011).

[8] B. Qi, C.-H. F. Fung, H.-K, Lo, and X. Ma, *Quantum Inf. Comput.* **7**, 73 (2007); Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K, Lo, *Phys. Rev. A* **78**, 042333 (2008).

[9] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, *Nature Photonics* **4**, 686 (2010).

[10] C.-H. F. Fung, B. Qi, K. Tamaki, and H.-K. Lo, *Phys. Rev. A* **75**, 032314 (2007); F. Xu, B. Qi, and H.-K. Lo, *New J. Phys.* **12**, 113026 (2010); S.-H. Sun, M.-S. Jiang, and L.-M. Liang, *Phys. Rev. A* **83**, 062331 (2011); Y.-L. Tang, H.-L. Yin, X. Ma, Chi-Hang Fred Fung, Y. Liu, H.-L. Yong, T.-Y. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, *Phys. Rev. A* **88**, 022308 (2013).

[11] D. Mayers and A. C.-C. Yao, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS98)* (IEEE Computer Society, Washington, DC, 1998), p. 503.

[12] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, *Phys. Rev. Lett.* **98**, 230501 (2007).

[13] N. Gisin, S. Pironio, and N. Sangouard, *Phys. Rev. Lett.* **105**, 070501 (2010).

[14] H.-K. Lo, M. Curty, and B. Qi, *Phys. Rev. Lett.* **108**, 130503 (2012).

[15] S. L. Braunstein and S. Pirandola, *Phys. Rev. Lett.* **108**, 130502 (2012).

[16] K. Tamaki, H.-K. Lo, C.-H. F. Fung, and B. Qi, *Phys. Rev. A* **85**, 042307 (2012); X. Ma and M. Razavi, *Phys. Rev. A* **86**, 062319 (2012); T.-T. Song, Q.-Y. Wen, F.-Z. Guo, and X.-Q. Tan, *Phys. Rev. A* **86**, 022332 (2012); S.-H. Sun, M. Gao, C.-Y. Li, and L.-M. Liang, *Phys. Rev. A* **87**, 052329 (2013); F. Xu, B. Qi, Z. Liao, and H.-K. Lo, *Appl. Phys. Lett.* **103**, 061101 (2013); M. Curty, F. Xu, W. Cui, C. C. W. Lim, K. Tamaki, and H.-K. Lo, *Nature Comm.* **5**, 3732 (2014).

[17] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, *Phys. Rev. Lett.* **111**, 130501 (2013); T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. P. Temporão, and J. P. von der Weid, *Phys. Rev. A* **88**, 052303 (2013); Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li, X. Ma, J. S. Pelc, M. M. Fejer, Q. Zhang, and J.-W. Pan, *Phys. Rev. Lett.* **111**, 130502 (2013); Z. Tang, Z. Liao, F. Xu, B. Qi, L. Qian, and H.-K. Lo, *Phys. Rev. Lett.* **112**, 190503 (2014); Y.-L. Tang *et al.*, *Phys. Rev. Lett.* **113**, 190501 (2014); Y.-L. Tang *et al.*, *IEEE J. Sel. Topics Quantum Electron.* **21**, 6600407 (2014).

[18] F. Xu, M. Curty, B. Qi, and H.-K. Lo, *IEEE J. Sel. Topics Quantum Electron.* **21**, 6601111 (2015).

[19] P. González, L. Rebón, T. Ferreira da Silva, M. Figueroa, C. Saavedra, M. Curty, G. Lima, G. B. Xavier, and W. A. T. Nogueira, arXiv:1410.1422v2 [quant-ph] (2014).

[20] C. C. W. Lim, B. Korzh, A. Martin, F. Bussières, R. Thew, and H. Zbinden, *Appl. Phys. Lett.* **105**, 221112 (2014).

[21] W.-F. Cao, Y.-Z. Zhen, Y.-L. Zheng, Z.-B. Chen, N.-L. Liu, K. Chen, J.-W. Pan, arXiv:1410.2928v1 [quant-ph] (2014).

[22] E. Biham, B. Huttner, and T. Mor, *Phys. Rev. A* **54**, 2651 (1996).

[23] H. Inamori, *Algorithmica* **34**, 340 (2002).

[24] Y.-H. Kim, *Phys. Rev. A* **67**, 040301(R) (2003).

[25] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, *Phys. Rev. A* **73**, 022320 (2006); N. Jain, E. Anisimova, I. Khan, V. Makarov, Ch. Marquardt, and G. Leuchs, *New J. Phys.* **16**, 123030 (2014); N. Jain, B. Stiller, I. Khan, V. Makarov, Ch. Marquardt, and G. Leuchs, *IEEE J. Sel. Topics Quantum Electron.* **21**, 6600710 (2015).

[26] J. Barrett, R. Colbeck, and A. Kent, *Phys. Rev. Lett.* **110**, 010503 (2013).

[27] Marcos Curty, private communication.

[28] T. Ferreira da Silva, Gustavo C. do Amaral, G. B. Xavier, G. P. Temporão, and J. P. von der Weid, *IEEE J. Sel. Topics Quantum Electron.* **21**, 6600309 (2014).