PHYSICAL REVIEW A **94**, 030302(R) (2016)

Creation of backdoors in quantum communications via laser damage

Vadim Makarov,[1,2,3,*] Jean-Philippe Bourgoin,[1,2] Poompong Chaiwongkhot,[1,2] Mathieu Gagné,[4] Thomas Jennewein,[1,2,5] Sarah Kaiser,[1,2] Raman Kashyap,[4] Matthieu Legré,[6] Carter Minshull,[1] and Shihan Sajeed[1,3]

[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

[2]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

[3]*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1*

[4]*Department of Engineering Physics and Department of Electrical Engineering, École Polytechnique de Montréal, Montréal, QC, Canada H3C 3A7*

[5]*Quantum Information Science Program, Canadian Institute for Advanced Research, Toronto, ON, Canada M5G 1Z8*

[6]*ID Quantique SA, Chemin de la Marbrerie 3, 1227 Carouge, Geneva, Switzerland*

(Received 11 November 2015; published 15 September 2016)

Practical quantum communication (QC) protocols are assumed to be secure provided implemented devices are properly characterized and all known side channels are closed. We show that this is not always true. We demonstrate a laser-damage attack capable of modifying device behavior on demand. We test it on two practical QC systems for key distribution and coin tossing, and show that newly created deviations lead to side channels. This reveals that laser damage is a potential security risk to existing QC systems, and necessitates their testing to guarantee security.

DOI: 10.1103/PhysRevA.94.030302

Cryptography, an art of secure communication, has traditionally relied on either algorithmic or computational complexity [1]. Even the most state-of-the-art classical cryptographic schemes do not have a strict mathematical proof to ascertain their security. With the advance of quantum computing, it may be a matter of time before the security of the most widely used public-key cryptography protocols is broken [2]. Quantum communication protocols, on the other hand, have theoretical proofs of being unconditionally secure [3–9]. In theory, their security is based on the assumption of modeled behavior of implemented equipment. In practice, the actual behavior often deviates from the modeled one, leading to a compromise of security as has been seen so far in the case of quantum key distribution (QKD) [10–16]. However, it is widely assumed that as long as these deviations are properly characterized and security proofs are updated accordingly [5,17], implementations are unconditionally secure. In this work we show that satisfying this during the initial installation only is not enough to guarantee security. Even if a system is perfectly characterized and deviations are included into the security proofs, an adversary can still create a new deviation on demand and make the system insecure.

Before going into detail on how the adversary may do it, let us consider a few examples of deviations and their consequences. For example, a calibrated optical attenuator is required to set a precise value of the outgoing mean photon number μ in the implementations of ordinary QKD [18,19], decoy-state QKD [20], coherent-one-way QKD [21], measurement-device-independent QKD [22], continuous-variable QKD [23], digital signature [7], relativistic bit commitment [8], coin-tossing [24], and secret-sharing [9] protocols. An unexpected increase of this optical component's attenuation may cause a denial-of-service. However, a reduction in attenuation will increase μ, leading to a compromise of security via attacks that rely on measurement of multiphoton pulses [25,26]. For example, in QKD and secret-sharing this will allow eavesdropping of the key, and in bit commitment cheating the committed bit value. Some implementations use a detector for time synchronization [8,9,18,19,21–24]. Desensitizing it may result in the denial-of-service. However, several implementations require a calibrated monitoring detector for security purposes [8,9,18,19,21,23,24]. A reduction in its sensitivity may lead to security vulnerabilities such as a Trojan-horse attack that reads the state preparation [27]. This leaks the key in QKD, increases the cheating probability in coin-tossing [26], leaks the program and client's data in quantum cloud computing [6], and allows forging of digital signatures [7]. Many implementations use beamsplitters and rely on their precharacterized splitting ratio (e.g., [8,18–21,23,24]). A shift in the splitting ratio may lead to either the denial-of-service or security vulnerabilities (e.g., [28] or one of the above-mentioned attacks). A shift in characteristics of a phase modulator or a Faraday mirror may create imperfect qubits that will result in the denial-of-service or a breach in security [14,15,29]. If the dark count rate of single-photon detectors is increased, it may lead to the denial-of-service [30]. Even in device-independent QKD (DI-QKD) [31], the absence of information-leakage channels and memory is assumed [32]. Thus, there is a risk these assumptions may be compromised by deviations in device characteristics. To give a speculative illustration, let us suppose detectors in DI-QKD emit light on detection [33–35], and to prevent this leaking information about detection results, spectral filters and optical isolators are added to the devices. Then, unexpected deviations in characteristics of the latter components become important for security. In summary, quantum communication systems rely on multiple characteristics of many components for their correct operation, and a deviation might lead to severe security consequences.

In classical communications, there is no real concern about the possibility of a shift in device characteristics. Classical devices' security-critical parts can be physically separated from the communication channel and isolated from physical access by the adversary [36]. However, the front-end of

*makarov@vad1.com

a quantum communication system is essentially an analog optical system connected to the channel (at least, at our present level of the technology), and is easily accessible by the adversary. The latter can shoot a high-power laser from the communication channel to alter system component characteristics via laser damage [30]. The question is, what will this achieve? Will the adversary break some component needed for operation and cause the denial-of-service (which is not a useful outcome for her), or will she change some component in such a way as to facilitate a compromise of security? Further, will the security compromise be only possible in theory or be practical with today's technology? This cannot be predicted in advance, because system implementations contain many components and their laser damage thresholds and failure behavior are generally not precisely known. To assess the risk for quantum communications, we have performed tests on two extensively characterized, completely different and widely used implementations: a commercial fiber-optic system for QKD and coin-tossing with phase-encoded qubits [18,19], and a free-space system for QKD with polarization-encoded qubits [20]. In both systems, we have unfortunately observed the best possible outcome for the adversary. After the laser damage, the systems' security has become compromisable with today's technology.

Laser damage in fiber-optic system. As a representative of a fiber-optic quantum communication implementation, we chose a plug-and-play QKD [18] and loss-tolerant quantum coin tossing (QCT) [24]. Both were implemented using a commercial system Clavis2 from ID Quantique [19]. In both cases, Bob sends bright light pulses to Alice. Alice randomly encodes her secret bits by applying one out of four phases $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$, attenuates the pulses, and reflects them back to Bob [Fig. 1(a)]. The security of both protocols requires an upper bound on the mean photon number μ coming out of Alice. Otherwise, an eavesdropper Eve can perform a Trojan-horse attack [27] by superimposing extra light to the bright pulses on their way to Alice from Bob. If Alice is unaware of this and applies the same attenuation, then light coming out of her has a higher μ than allowed by the security proofs [5], making the implementations insecure. It is thus

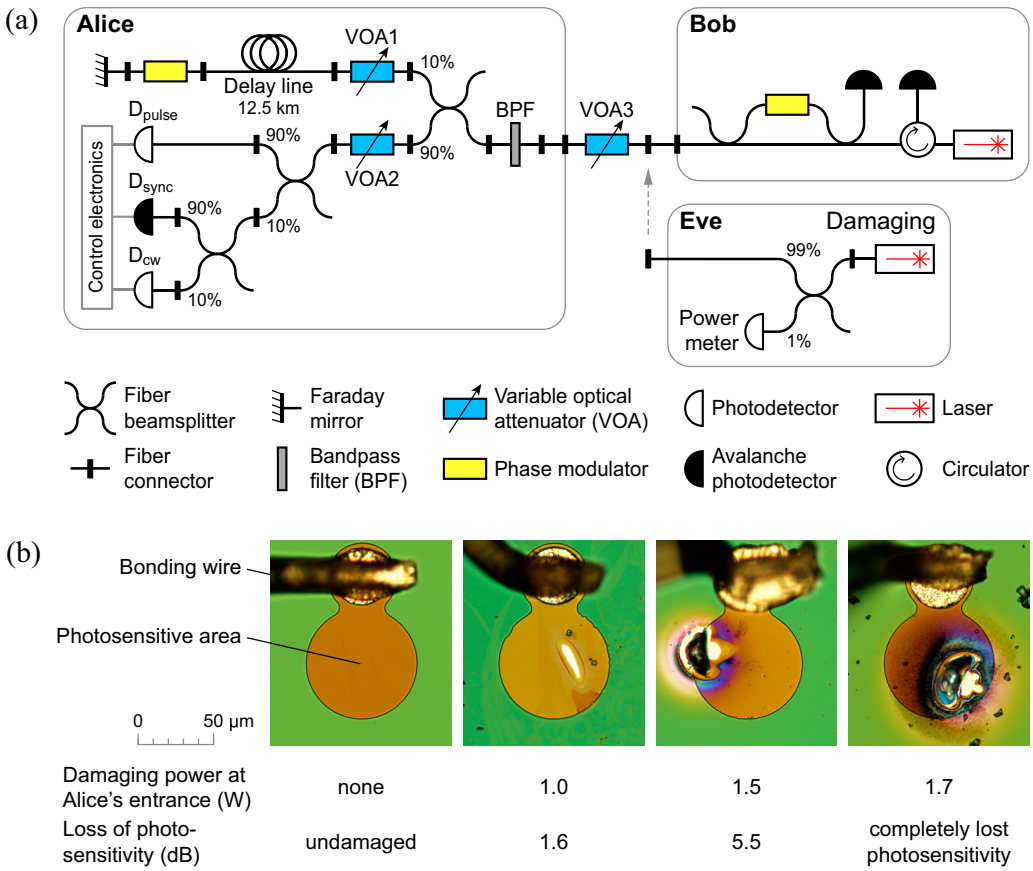

FIG. 1. Attack on fiber-optic system Clavis2. (a) Experimental setup. The system consists of Alice and Bob connected by a lossy fiber communication channel (simulated by variable optical attenuator VOA3). Bob sends to Alice pairs of bright coherent optical pulses, produced by his laser and two fiber arms of unequal length [18,19]. Alice uses fiber beamsplitters to divert parts of incoming pulse energy to monitoring detector D_{pulse}, synchronization detector D_{sync}, and line-loss measurement detector D_{cw}. She prepares quantum states by phase modulating the pulses, reflecting them at a Faraday mirror and attenuating to single-photon level with VOA1. Bob measures the quantum states by applying his basis choice via phase modulator and detecting outcome of quantum interference with single-photon avalanche photodetectors. Eve's damaging laser is connected to the channel manually. BPF: bandpass filter. (b) Pulse-energy-monitoring photodiode before and after damage. Brightfield microphotographs show the top view of decapsulated photodiode chips. The last two samples have holes melted through their photosensitive area. Scattered dark specks are debris from decapsulation.

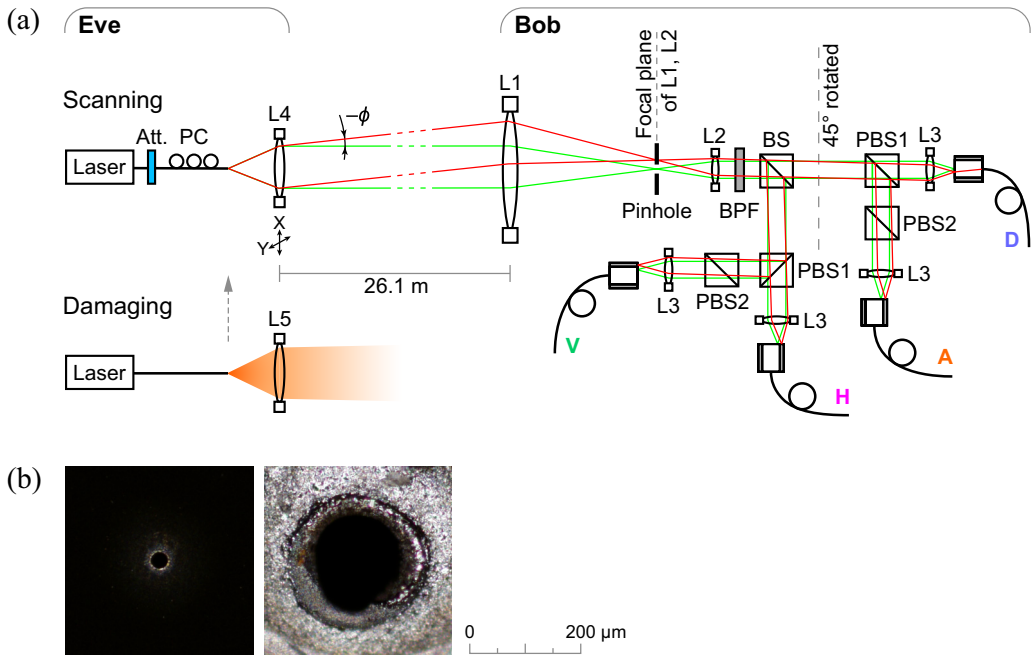

FIG. 2. Attack on free-space QKD system. (a) Experimental setup. QKD receiver Bob consists of two lenses L1, L2 reducing input beam diameter, 50:50 beamsplitter (BS), and two arms measuring photons in HV and DA polarizations using polarizing beamsplitters (PBS) [16,20]. Photons are focused by lenses L3 into multimode fibers leading to single-photon detectors. Setup drawing is not to scale. Eve's apparatus contains a scanning laser source that tilts the beam angle (ϕ,θ) by laterally shifting lens L4. Green marginal rays denote initial Eve's alignment, replicating the alignment Alice-Bob at $\phi = \theta = 0$. Red marginal rays show a tilted scanning beam missing fiber cores V, H, A, but coupling into D. Eve's damaging laser source can be manually inserted in place of the scanning source. Att.: attenuator; PC: polarization controller. (b) Spatial filter before and after damage. Darkfield microphotographs show front view of the pinhole. See Supplemental Material Sec. IV [37] for real-time video recording of laser damage to the pinhole inside Bob.

crucial for the security of both protocols that Alice monitors the incoming pulse energy. This is achieved by employing a pulse-energy-monitoring detector [D_{pulse} in Fig. 1(a)]. A portion of the incoming light is fed to D_{pulse} such that whenever extra energy is injected, an alarm is produced [26]. The sensitivity of D_{pulse} is factory calibrated, thus closing the side channel associated with the Trojan-horse attack.

Our testing showed that this countermeasure is vulnerable to laser damage. During normal QKD operation, we disconnected the fiber channel Alice-Bob temporarily and connected Eve [Fig. 1(a)]. She then injected 1550 nm laser light from an erbium-doped fiber amplifier for 20–30 s, delivering continuous-wave (cw) high power into Alice's entrance. 44% of this power reached the fiber-pigtailed InGaAs *p-i-n* photodiode D_{pulse} (JDSU EPM 605LL), and damaged it partially or fully. It became either less sensitive to incoming light (by 1–6 dB after 0.5–1.5 W illumination at Alice's entrance) or completely insensitive (after $\geqslant$1.7 W). The physical damage is shown in Fig. 1(b). No other optical component was damaged at this power level. We repeated the experiment with six photodiode samples. In half of these trials, QKD continued uninterrupted and kept producing more key after we reconnected the channel back to Bob, as if nothing has happened. In the other half, a manual software restart was needed. However, in all the trials the damage was sufficient to permanently open the system up to the Trojan-horse attack. As modeled in Ref. [26], in the QKD protocol, Eve can eavesdrop partial or full key using today's best technology if the sensitivity of D_{pulse} drops by more than 5.6 dB. In the QCT implementation, a sensitivity reduction by 2.6 dB can increase Bob's cheating probability above a classical level, removing any quantum advantage of coin-tossing. Laser damage thus compromises both the QKD and QCT implementations. See Supplemental Material Sec. I [37] for details. ID Quantique is developing countermeasures for their affected QKD system.

Laser damage in free-space system. As a representative of free-space quantum communication, we chose a long-distance satellite QKD prototype operating at 532 nm wavelength [20] employing Bennett-Brassard 1984 (BB84) protocol [3]. At each time slot, Alice randomly sends one out of four polarizations: horizontal (H), vertical (V), +45 ° (D), or −45 ° (A) using a phase-randomized attenuated laser. Bob randomly measures in either horizontal-vertical (HV) or diagonal-antidiagonal (DA) basis, using a polarization-beamsplitter receiver [Fig. 2(a)]. It has been shown in Ref. [16] that an eavesdropper can, in practice, tilt the beam going towards Bob by an angle (ϕ,θ) such that the beam misses, partially or fully, the cores of fibers leading to three detectors while being relatively well coupled into the core leading to the fourth detector, as illustrated in Fig. 2(a). This happens because real-world optical alignments are inherently imperfect and manufacturing precision is finite. By sending light at different spatial angles, the eavesdropper can have control over Bob's basis and measurement outcome and steal the key unnoticed [13,16,38]. This attack can be prevented by placing a spatial filter or "pinhole" at the focal plane of lenses L1 and L2, as shown in Fig. 2(a) [16]. Since the pinhole limits the field of view, any light entering at a higher spatial angle is blocked and

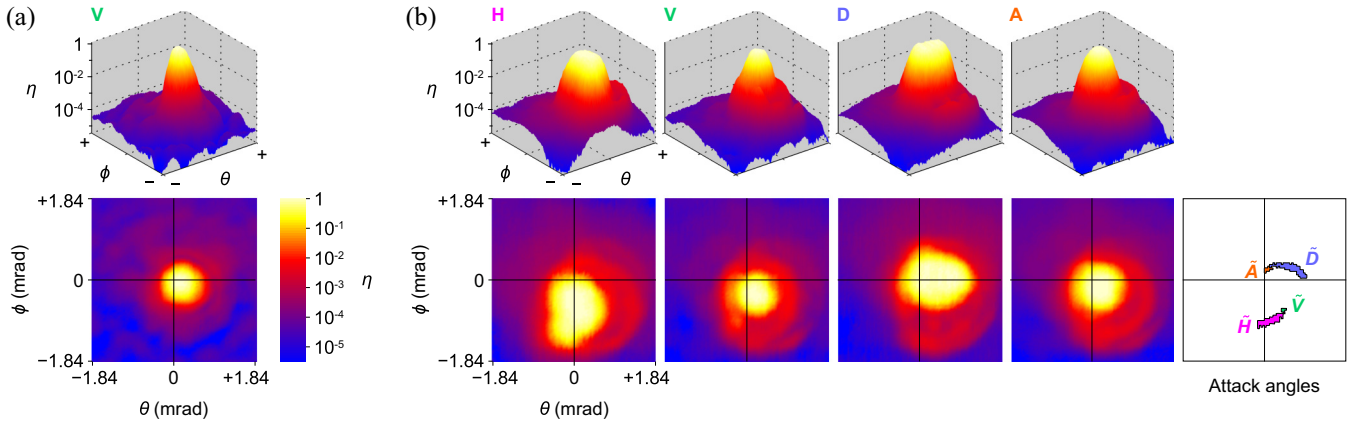

FIG. 3. Efficiency-mismatch side channel opened after laser damage in free-space QKD system. Each pair of 3D-2D plots shows normalized photon detection efficiency η in a receiver channel versus illuminating beam angles ϕ and θ. (a) Before laser damage, the angular dependence is essentially identical between the four channels [16]. Plot for one channel (V) before damage is shown. (b) After the laser damage, the four receiver channels H, V, D, and A exhibit unequal sensitivity to photons outside the middle area around $\phi = \theta = 0$. The last plot shows angular ranges for targeting the four detectors that satisfy conditions for the faked-state attack.

Eve no longer has access to the target angles required to have control over Bob. As was demonstrated in Ref. [16], a pinhole of 25 μm diameter eliminates this side channel by making the angular efficiency dependence identical between the four detectors [Fig. 3(a)].

Our testing showed that this countermeasure is destroyed by laser damage. From a distance of 26.1 m, we shot an 810 nm collimated laser beam delivering a 10 s pulse of 3.6 W cw power at the pinhole inside Bob's setup. The intensity there was sufficient to melt the material (13-μm-thick stainless steel) and enlarge the hole diameter to $\approx$150 μm. The state of the pinhole before and after damage is shown in Fig. 2(b), and a real-time video of the damage process is shown in Supplemental Material Sec. IV [37]. Although Bob was up and running in photon counting mode during the test, none of his other components were damaged. With this larger pinhole opening, it was again possible to send light at angles that had relatively higher mismatches in efficiency, as shown in Fig. 3(b). This enabled a faked-state attack under realistic conditions of channel loss in 1–15 dB range with quantum bit error ratio $< 6.6\%$. Thus laser damage completely neutralizes this countermeasure, and makes this free-space QKD system insecure (see Supplemental Material Sec. II [37] for details).

Discussion. The crucial step of the attack, creating the deviation in device characteristics, has thus been experimentally demonstrated for both systems tested. We repeated this step several times and confirmed that laser power above a certain value (1.7 W in a fiber-optic system and 3.6 W in a free-space one) always destroys the security-critical component, without inflicting any collateral damage that could result in the denial-of-service. After this, building a complete eavesdropper would be a realistic if time-consuming task [39].

In our testing, we have not done anything that Eve could not do in the real world. She could buy a copy of each system, rehearse her attacks, then attack an installed system of the same model. By Kerckhoffs' principle [40], Eve is assumed to know the system characteristics and results of damage precisely. In practice when attacking installed devices, if she needs to measure their characteristics, she may probe them remotely by imaging, reflectometry [27], and watching public communication Alice-Bob [38,39].

At present, no quantum communication system has countermeasures specifically designed to stop laser-damage attack, neither do they have a mechanism to check all possible deviations in device characteristics from the modeled values. Countermeasures to other attacks do not prevent this attack, in fact they become weak points as our experimental study shows. Development of necessary countermeasures is complicated by the fact that Eve can use a laser with different characteristics: power, timing (e.g., short-pulsed laser induces different damage mechanisms than cw thermal damage we have observed [41]), wavelength, polarization. Eve can attack the systems in different phases of their operation including the powered-off state, which can control what component is damaged. We have experimentally observed the dependence of damage on the laser timing profile, as detailed in Supplemental Material Sec. III [37], where we show that some profiles have resulted in the denial-of-service but some in a successful attack. We stress that Eve will select the illumination regime that results in the successful attack, if such regime exists at all. Any countermeasure must thus be tested in all possible illumination regimes. Possible directions of development include a passive optical power limiter [42], a single-use "fuse" that permanently breaks the optical connection if a certain power is exceeded, battery-powered active monitoring supplemented with wavelength filtering, or an optical isolator (for Alice that uses one-way light propagation [6,7,20–23,31,39]). Hardware self-characterization may be promising [43]; however, to protect from arbitrary damage it must monitor a potentially large number of hardware parameters.

It is an interesting question if risk for untested system designs can be estimated. As we have discussed, any given system design contains many optical components with unknown damage characteristics. The outcome of damage (denial-of-service or successful attack) is thus impossible to predict prior to testing. Then, if some of the system designs chosen at random were tested, the risk for the remaining untested designs could be calculated by Bayesian statistics [44]. Unfortunately,

truly random choice is impractical to implement with the current state of quantum communications research and limited sample availability. We have instead tested the two system designs that were available in our laboratory. This biased the choice towards more mature and older designs. Although this unknown bias makes the Bayesian analysis inapplicable, we find it illustrative to consider the risk figure that would have applied if the choice were random. With zero systems tested, the Bayesian probability that at least 20% of the untested system designs (assuming at least 50 of them exist) are vulnerable to this attack is 70.4% (80%), assuming a Jeffreys (uniform) prior. If two randomly chosen system designs were tested with two positive outcomes, this probability would have increased greatly to 98.9% (98.6%). Note that the security risk is generally high, which is in stark contrast with the very low expected theoretical risk [4,5,17].

We have experimentally demonstrated laser damage as an eavesdropping tool that alters parameters of a well-characterized quantum communication system. Any modification of system characteristics might compromise the security either directly by leading to an attack as we have demonstrated, or indirectly by shifting some parameter in the security proof so it would no longer apply. Existing security proofs do not accommodate this; neither do existing systems have any countermeasure implemented against this. Our results thus reveal the potential security risk for other existing systems, which should be tested against this attack.

Acknowledgments. We thank Q. Liu, E. Anisimova, and O. Di Matteo for early experimental efforts, and S. Todoroki, N. Lütkenhaus, M. Mosca, Y. Zhang, L. Lydersen, and S. Lydersen for discussions. This work was supported by the U.S. Office of Naval Research, Industry Canada, CFI, Ontario MRI, NSERC, Canadian Space Agency, ID Quantique, European Commission's FET QICT SIQS prACoject, EMPIR 14IND05 MIQC2 project, and CryptoWorks21. We acknowledge using University of Waterloo's Quantum NanoFab. P.C. was supported by a Thai DPST scholarship. J.-P.B. was supported by FED DEV.

Author contributions. V.M. conceived and led the study. S.K. implemented the fiber-optic experiment. S.S. implemented the free-space experiment and contributed to the fiber-optic experiment. P.C. contributed to the free-space experiment. M.G. contributed to the fiber-optic experiment. C.M. made minor contributions to the free-space experiment. M.L. provided and supported the fiber-optic QKD system under test. T.J. and J.-P.B. provided the free-space QKD receiver under test and contributed to the free-space experiment. R.K. provided the fiber laser facility and cosupervised the fiber-optic experiment. S.S. and V.M. wrote the article, with contributions from all authors.

[1] S. Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography* (Fourth Estate, London, 1999).
[2] P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).
[3] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)* (IEEE Press, New York, 1984), p. 175.
[4] H.-K. Lo and H. F. Chau, Science **283**, 2050 (1999).
[5] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **4**, 325 (2004).
[6] S. Barz, E. Kashefi, A. Broadbent, J. F. Fitzsimons, A. Zeilinger, and P. Walther, Science **335**, 303 (2012).
[7] R. J. Collins, R. J. Donaldson, V. Dunjko, P. Wallden, P. J. Clarke, E. Andersson, J. Jeffers, and G. S. Buller, Phys. Rev. Lett. **113**, 040502 (2014).
[8] T. Lunghi, J. Kaniewski, F. Bussières, R. Houlmann, M. Tomamichel, A. Kent, N. Gisin, S. Wehner, and H. Zbinden, Phys. Rev. Lett. **111**, 180504 (2013).
[9] W. P. Grice, P. G. Evans, B. Lawrie, M. Legré, P. Lougovski, W. Ray, B. P. Williams, B. Qi, and A. M. Smith, Opt. Express **23**, 7300 (2015).
[10] C. H. Bennett, F. Bessette, L. Salvail, G. Brassard, and J. Smolin, J. Cryptology **5**, 3 (1992).
[11] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006); **78**, 019905 (2008).
[12] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **7**, 73 (2007).
[13] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics **4**, 686 (2010).
[14] F. Xu, B. Qi, and H.-K. Lo, New J. Phys. **12**, 113026 (2010).
[15] S.-H. Sun, M.-S. Jiang, and L.-M. Liang, Phys. Rev. A **83**, 062331 (2011).
[16] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, Phys. Rev. A **91**, 062301 (2015).
[17] C.-H. F. Fung, K. Tamaki, B. Qi, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **9**, 131 (2009).
[18] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, New J. Phys. **4**, 41 (2002).
[19] Clavis2 specification sheet, http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf, visited 20 March 2016.
[20] J.-P. Bourgoin, N. Gigov, B. L. Higgins, Z. Yan, E. Meyer-Scott, A. K. Khandani, N. Lütkenhaus, and T. Jennewein, Phys. Rev. A **92**, 052339 (2015).
[21] N. Walenta *et al.*, New J. Phys. **16**, 013047 (2014).
[22] Y.-L. Tang, H.-L. Yin, S.-J. Chen, Y. Liu, W.-J. Zhang, X. Jiang, L. Zhang, J. Wang, L.-X. You, J.-Y. Guan, D.-X. Yang, Z. Wang, H. Liang, Z. Zhang, N. Zhou, X. Ma, T.-Y. Chen, Q. Zhang, and J.-W. Pan, Phys. Rev. Lett. **113**, 190501 (2014).
[23] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, Nat. Photonics **7**, 378 (2013).
[24] A. Pappa, P. Jouguet, T. Lawson, A. Chailloux, M. Legré, P. Trinkler, I. Kerenidis, and E. Diamanti, Nat. Commun. **5**, 3717 (2014).
[25] S. Félix, N. Gisin, A. Stefanov, and H. Zbinden, J. Mod. Opt. **48**, 2009 (2001).
[26] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Phys. Rev. A **91**, 032326 (2015).
[27] A. Vakhitov, V. Makarov, and D. R. Hjelme, J. Mod. Opt. **48**, 2023 (2001).

[28] 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).

[29] F. Xu, K. Wei, S. Sajeed, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H.-K. Lo, Phys. Rev. A **92**, 032305 (2015).

[30] A. N. Bugge, S. Sauge, A. M. M. Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Phys. Rev. Lett. **112**, 070503 (2014).

[31] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. **97**, 120405 (2006).

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

[33] C. Kurtsiefer, P. Zarda, S. Mayer, and H. Weinfurter, J. Mod. Opt. **48**, 2039 (2001).

[34] P. V. P. Pinheiro *et al.* (unpublished).

[35] A. Meda, I. P. Degiovanni, A. Tosi, Z. L. Yuan, G. Brida, and M. Genovese, arXiv:1605.05562.

[36] *National Security Telecommunications and Information Systems Security Advisory Memorandum (NSTISSAM) TEMPEST/2-95, Red/Black Installation Guidance* (US National Security Agency, 1995), declassified in 2000 [http://cryptome.org/tempest-2-95.htm].

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.94.030302 for details of the laser damage experiment on fiber and free space QKD system and other relevant details.

[38] V. Makarov and D. R. Hjelme, J. Mod. Opt. **52**, 691 (2005).

[39] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nat. Commun. **2**, 349 (2011).

[40] A. Kerckhoffs, J. Sci. Mil. **9**, 5 (1883).

[41] R. M. Wood, *Laser-Induced Damage of Optical Materials* (CRC Press, Boca Raton, FL, 2003).

[42] L. W. Tutt and T. F. Boggess, Prog. Quantum Electron. **17**, 299 (1993).

[43] L. Lydersen, V. Makarov, and J. Skaar, Phys. Rev. A **83**, 032306 (2011).

[44] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. (Chapman and Hall, London, 2004).

Supplemental material: Creation of backdoors in quantum communications via laser damage

Vadim Makarov, Jean-Philippe Bourgoin, Poompong Chaiwongkhot, Mathieu Gagné, Thomas Jennewein, Sarah Kaiser, Raman Kashyap, Matthieu Legré, Carter Minshull, and Shihan Sajeed

I. LASER-DAMAGE EXPERIMENT ON FIBER-OPTIC SYSTEM

In our experiment, we damaged D_{pulse} during QKD operation, trying not to interrupt it. The system was allowed to start up and produce a secret key for several QKD cycles, using BB84 protocol [1]. To perform laser damage, we disconnected the channel for 2–3 min, giving us enough time to apply high power to Alice, and then reconnected the channel. We tried this at different points in the QKD operation cycle. Sometimes the software recovered and resumed QKD, and sometimes it got stuck in recalibration routines. In the latter case, a manual software restart resumed QKD. Owing to a limited number of trials, we did not perfect this timing aspect.

We tested a total of 6 photodiode samples. We damaged each of them by applying high power laser light at Alice's entrance. We then used the manufacturer's factory-calibration software to measure how much extra signal power (compared to the pre-calibrated power level) could be injected without triggering the alarm [2]. This quantified the reduction in sensitivity due to the damage. Three samples were exposed twice to a progressively higher power. For example, one sample was first exposed to 0.5 W power at Alice's entrance that reduced its photosensitivity by 1 dB, then to 0.75 W power that reduced its photosensitivity by 6 dB. For the other two samples these numbers were 0.75 W with no change in sensitivity then 1.0 W, 1.6 dB (shown in 2nd microphotograph in Fig. 1(b) in main text); 1.0 W, 5 dB then 1.5 W, 5.5 dB (shown in 3rd microphotograph in Fig. 1(b) in main text). For the remaining three samples, 1.7 W was applied at Alice's entrance, and D_{pulse} completely lost photosensitivity, becoming electrically either a large resistor (shown in 4th microphotograph in Fig. 1(b) in main text). or an open circuit. After we were done with each sample, we used the same manufacturer's factory-calibration software to pre-calibrate the sensitivity of the next undamaged D_{pulse} sample, following the factory procedure.

No other component in Alice was damaged during these trials. We also tested some components separately. FC/PC and FC/APC optical connectors used in Alice and in the channel withstood 3 W c.w., while copies of Alice's 10:90 fiber beamsplitters (AFW Technologies FOSC-1-15-10-L-1-S-2) withstood up to 8 W c.w. with no damage.

Figure 4 summarizes a system operation log when it recovered automatically after the damage that made the photodiode an open-circuit with no photosensitivity. In the current system implementation, this represents an ideal outcome for an attacker.

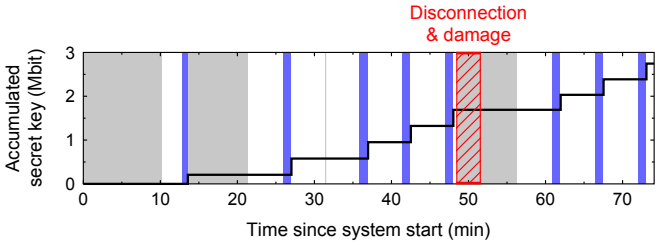

FIG. 4. **Fiber-optic QKD system operation during laser damage.** The plot shows accumulated secret key amount versus time. Grey bands denote the system performing recalibration routines, white bands denote the quantum bit sending and receiving, and blue (darker) bands denote classical post-processing. All this information was extracted from the QKD system log files after the experiment. The band hatched in red denotes the time when the fiber channel Alice–Bob was temporarily disconnected and the laser damage to Alice was done by 1.7 W laser power, resulting in D_{pulse} becoming an open circuit with no photosensitivity.

For damaging and component tests, Eve used an erbium-doped fiber amplifier seeded from a 1550.7 nm laser source (EDFA; IPG Photonics ELR-70-1550-LP). She injected 0–2 W c.w. power at Alice's entrance. The injected power was monitored with a 1:99 fiber beamsplitter tap and a power meter (see Fig. 1(a) in main text). A manually operated shutter at the output of EDFA allowed to ramp the power up and down smoothly between 0 and the target level, with tens of milliseconds transition time. The spectral characteristics of EDFA's built-in seed laser did not precisely match the passband of the BPF at Alice's entrance (1551.32–1552.12 nm passband at −0.5 dB level, $<$ 0.7 dB insertion loss; AFW Technologies BPF-1551.72-2-B-1-1). We therefore removed the BPF for the duration of experiment. The BPF was separately tested in-passband using a different EDFA (PriTel LNHPFA-37) with a narrowband seed laser, and passed more than 1 W c.w. with no damage.

The system QKD software ('QKD Sequence' application [3]) set the variable attenuator VOA2 at 2 dB. Thus, 44% of Alice's incoming light impinged D_{pulse}, while smaller fractions impinged D_{sync} and D_{cw}. The alarm threshold of D_{pulse} is calibrated when the system is assembled at the factory, and is not changed after that [2]. VOA3 introduced channel loss of 1.87 dB, to simulate the effect of ≈9 km long fiber line Alice–Bob.

The QKD system Clavis2 normally operates automat-

ically in cycles consisting of sending and receiving quantum states until either the memory buffer is full or photon detection efficiency has dropped significantly. It then uses the classical link Alice–Bob to post-process the detected data and distill the secret key [4]. Each cycle takes several minutes. If the last QKD cycle was interrupted because the detection efficiency was too low, or the key distillation failed, the system returns to start-up routines such as timing recalibration [5] before it resumes sending quantum states. This happens often in normal operation, because of naturally occurring drift of hardware and channel parameters. The software generally tries to recover automatically from various error conditions, to provide long-term unattended operation [6].

Predicted attacks on fiber-optic system with damaged pulse-energy-monitoring photodiode. As modeled in Ref. 2, for BB84 QKD protocol Eve can eavesdrop partial or full key information using today's best photonics technologies when the sensitivity of D_{pulse} has dropped by 4.3–5.6 dB, given that communication channel loss Alice–Bob is in a 1–7 dB range. (This corresponds to a multiplication factor x in the range of 2.7–3.6, see Fig. 11 in Ref. 2.) If we assume that Eve's equipment is only limited by the laws of quantum mechanics, then she can extract the full key information after only 0.4–0.8 dB reduction in sensitivity (x of 1.1–1.2). Similarly, for QCT with a dishonest Bob only limited by the quantum mechanics, all the quantum advantages of the protocol are eliminated if sensitivity reduction of 2.6 dB is obtained in Alice ($x = 1.805$), for a 15 km long communication channel. For a 10 dB sensitivity reduction, Bob's cheating probability approaches unity [2]. Since we have surpassed the above sensitivity reduction thresholds in our laser damage experiment, we consider the security of both QKD and QCT implementation compromised.

II. LASER-DAMAGE EXPERIMENT ON FREE-SPACE QKD SYSTEM

In order to neutralize the effect of the pinhole and reproduce the side-channel of spatial-mode detector-efficiency mismatch, our experiment consisted of three steps. Firstly, we performed scanning to certify that the system is secure against this side-channel. Secondly, we laser-damaged the pinhole to open the side-channel. Finally, we performed scanning again to demonstrate that the system's security has been compromised. In all three steps, Eve was placed at a distance of 26.1 m away from Bob and the steps were performed in sequence without making any interactions with Bob.

The first step involved changing the outgoing beam's angle (ϕ, θ) emitted from Eve's scanning setup (Fig. 2(a) in main text), then recording the corresponding count rate at all four detectors in Bob. This step is identical to that in Ref. 7. The scanning result is shown in Fig. 3(a) in main text, where a pair of 3D–2D plots shows the normalized photon detection efficiency in one receiver channel versus the illuminating beam angles ϕ and θ. With the pinhole in place, the angular dependence of efficiency is essentially identical between the four channels, hence only a plot for channel V is shown. No measurable amount of efficiency mismatch was found and no attack angles existed [7].

Then as the second step, Eve's scanning setup was replaced with the damaging setup. The latter contained a 810 nm laser diode (Jenoptik JOLD-30-FC-12) pumped by a current-stabilized power supply and connected to 200 μm core diameter multimode fiber. It provided continuously adjustable 0 to 30 W c.w. power into the fiber. An almost-collimated free-space beam was subsequently formed by a plano-convex lens L5 (Thorlabs LA1131-B; Fig. 2(a) in main text). The beam's intensity was nearly uniformly distributed across Bob's L1 (50 mm diameter achromatic doublet, Thorlabs AC508-250-A), with less than ±10% intensity fluctuation across Bob's input aperture. Transmission of L1 was about 82%, owing to its antireflection coating being designed for a different wavelength band. In the test detailed here, the power delivered at the pinhole plane was 3.6 W, sufficient to reliably produce a hole of ≈150 μm diameter in less than 10 s in a standard stainless-steel foil pinhole (Thorlabs P25S). We tested several pinholes and found that this power always made the hole. We also tested that power decreased to 2.0 W still produced a hole. No other component in Bob was damaged during the tests. Bob's lenses L3 received ∼1 μW power each, and single-photon detectors only received on the order of a few nW each, mainly owing to the presence of BPF after the pinhole. The BPF was used by Bob to increase the signal-to-noise ratio during QKD by heavily attenuating all light outside the 531–533 nm passband (it consisted of two stacked filters, Thorlabs FESH0700 followed by Semrock LL01-532-12-5) [8]. While the damaging beam was on, the detectors counted at their saturation rate of ∼35 MHz, which did not look abnormal to Bob as this sometimes occurs naturally owing to atmospheric conditions (during sunset, sunrise, fog). We remark that this type of detector usually survives tens of mW for a short time [9, 10]. Even if we had to use a wavelength within the BPF's passband, detector exposure to higher power could likely be avoided by shaping Eve's damaging beam.

After the damage, as the third step we replaced the damaging setup with the scanning setup again, and performed the final scanning of Bob's receiver with the damaged pinhole. The results are shown in Fig. 3(b) in main text. Now, the four receiver channels H, V, D, A exhibited unequal sensitivity to photons outside the middle area around $\phi = \theta = 0$. These efficiency plots were different from those measured in Ref. 7 without the pinhole, because of extra scattering at the edges of our laser-enlarged pinhole.

Predicted attack on free-space QKD system with damaged pinhole. We model a practical faked-state attack as described in Ref. 7. We assume a part of Eve is

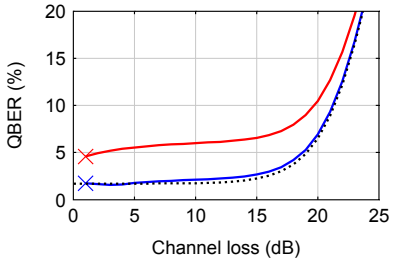

FIG. 5. **Modeled QBER observed by Bob in free-space QKD system.** The dotted curve shows QBER without Eve. At lower channel loss, the QBER is due to imperfect fidelity, while at higher channel loss Bob's detector background counts become the dominant contribution. The lower solid curve (blue) shows QBER under our attack when only Bob's sifted key rate is kept the same as before the attack. The upper solid curve (red) additionally keeps the same sifted key rates conditioned on each polarization sent by Alice, which more closely mimics a realistic system operation (see Ref. 7 for details).

situated outside Alice and measures the quantum states coming out. Then, another part of her regenerates the measured quantum states as attenuated coherent pulses and sends them to Bob, tilting her beam at an angle such that it has a relatively higher probability of being detected by the desired detector. Eve has information about Bob's receiver characteristics after the laser damage, and only uses devices available in today's technology [7]. For example, let's assume Eve sends a horizontally polarized light pulse. In this case, she should choose her tilt angle (ϕ, θ) from a subset $\tilde{H}$ selected in such a way that the efficiency $\eta_h(\tilde{H})$ of Bob's horizontal channel in $\tilde{H}$ is as high as possible, in order to maximize mutual information Eve–Bob. On the other hand, if Bob measures in the opposite (DA) basis, the detection probabilities in the D and A channels $\eta_d(\tilde{H})$ and $\eta_a(\tilde{H})$ should be as low as possible, to minimize QBER. Thus, to find attack angles for the horizontally polarized light, we choose $\tilde{H}$ that satisfies $\eta_h(\tilde{H}) \geq 0.6$ and $\delta(\tilde{H}) = \min\left\{\frac{\eta_h(\tilde{H})}{\eta_d(\tilde{H})}, \frac{\eta_h(\tilde{H})}{\eta_a(\tilde{H})}\right\} \geq 100$. Similarly, for V, D and A polarized pulses, we choose attack angles that satisfy $\eta_v(\tilde{V}) \geq 0.03$, $\delta(\tilde{V}) \geq 4.5$; $\eta_d(\tilde{D}) \geq 0.6$, $\delta(\tilde{D}) \geq 120$; $\eta_a(\tilde{A}) \geq 0.2$, $\delta(\tilde{A}) \geq 22$. These subsets of angles are shown in the rightmost plot in Fig. 3(b) in main text. Note that the thresholds η and δ used here are not optimal and have been picked manually. However, they satisfy the required conditions to successfully perform the faked-state attack with a resultant QBER $\leq 6.6\%$ in 1–15 dB channel loss range, as shown in Fig. 5. In the simulation, we assumed that Alice–Bob and Alice–Eve fidelity $F = 0.9831$ [7, 8], while Eve–Bob experimentally measured $F = 0.9904$. All other assumptions were the same as in Ref. 7.

III. DEVIATION THAT HAS LED TO A DENIAL-OF-SERVICE, AND HOW WE AVOIDED IT

Alice's setup in Clavis2 (Fig. 1(a) in main text) consists of more than 20 discrete optical components: over 10 fiber connectors, 3 beamsplitters, 3 detectors, 2 variable attenuators, a bandpass filter, phase modulator, and Faraday mirror. As discussed in main text, certain deviations in any of these components lead to the denial-of-service, while other deviations result in opening different security loopholes. Attacker's goal, and the goal of a thorough security tester, is to do one's best to avoid the former and demonstrate the latter. Parameters of the damaging laser illumination can and should be varied to reach this goal.

When we began testing the system components for laser damage, the synchronization detector D_{sync} initially presented an obstacle. This detector was based on an optical receiver module (Fujitsu FRM5W232BS) incorporating an avalanche photodiode biased below breakdown at > 30 V, providing an avalanche multiplication factor ≈ 6. It only took about 6 mW of optical power at the photodiode (translating to about 0.15 W at Alice's entrance) to die. It stopped providing the synchronization signal for Alice and thus broke the system, i.e., led to the denial-of-service. After an investigation, it turned out that the energy that killed it was chiefly provided by its high-voltage electrical bias circuit and not the optical signal. The bias circuit was based on a specialised integrated circuit with overcurrent protection (Maxim Integrated MAX1932ETC) followed by an LC low-pass filter with inductor L = 330 μH and capacitor C = 0.47 μF. If the optical power is applied suddenly, with sub-nanosecond rise time, it momentarily induces a large photocurrent supplied from C that destroys the avalanche photodiode. If, however, the optical power is applied gradually, with millisecond rise time, C discharges slowly and then the relatively slow overcurrent protection reacts in the integrated circuit, lowers the bias voltage and saves the photodiode. We thus added a manual shutter to the EDFA to make the damaging power rise from zero slowly, allowing D_{sync} to easily withstand the optical power used in our attack while being electrically powered up. Another solution could be to damage the system when it is without electrical power. It can also be said that we could choose to selectively damage one of two components in Alice, albeit one of them bricking the system.

We ran our damage tests with VOA2 (OZ Optics DD-600-11-1300/1550-9/125-S-40-3S3S-1-1-485:1-5-MC/IIC) set at 2 dB, because this is what the manufacturer's QKD software available for the research system Clavis2 set it at. The support of the pulse-energy-monitoring countermeasure was not implemented in this software [2]. In contrast, the manufacturer's factory-calibration software supported it fully and set VOA2 between 2 and ≈ 15 dB, complementary to the channel loss, in order to maintain constant power at the three Alice's

detectors D_{pulse}, D_{sync}, and D_{cw}. The higher settings of VOA2 would require more laser power to damage D_{pulse}. However, D_{pulse} could also be damaged during the system start-up time, when it sends the homing command to VOA2. The homing command causes it to traverse its lowest attenuation values for a few seconds, likely being sufficient for Eve to do the damage at already demonstrated power levels.

IV. REAL-TIME VIDEO RECORDING OF LASER DAMAGE TO THE SPATIAL FILTER INSIDE BOB'S SETUP

Download the video at `http://vad1.com/pinhole-laser-damage-20140825.wmv` (Windows Media Video, 14.4 MiB) or `http://vad1.com/pinhole-laser-damage-20140825.ppsx` (PowerPoint Show, 17.0 MiB). The video shows the spatial filter (Thorlabs P20S) illuminated by 3.6 W c.w. 810 nm laser beam for 10 s, focused in a spot much wider than the original pinhole diameter of 20 μm. This is a filter sample with a slightly smaller original pinhole diameter than the one used to obtain efficiency mismatch data in this article and shown in Fig. 2(b) in main text. The samples were otherwise of the same type and damaged under the same conditions. The video was taken via a mirror lowered inside Bob's setup. The pinhole plane was imaged from the front side at an angle slightly off normal, in order for the mirror not to obstruct the damaging beam. Canon MP-E 65 mm lens was used at 2.8× magnification and f/16 lens aperture (f/60 effective aperture), with Canon EOS 7D camera body. The pinhole was brightly lit sideways with a fiber-optic illuminator bundle, in order to bring up detail. During the laser exposure, the steel foil can be seen deforming from heat, popping out of focus and apparently shifting laterally in the image; however the lateral shift is an artefact of the camera's angle of view being off-normal. After the laser is switched off, the foil cools and returns to the original position, now with about 150 μm diameter hole in it. Sound was added later for an artistic effect.

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

[2] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Phys. Rev. A **91**, 032326 (2015).

[3] Clavis2 specification sheet, `http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf`, visited 20 March 2016.

[4] C. H. Bennett, F. Bessette, L. Salvail, G. Brassard, and J. Smolin, J. Cryptology **5**, 3 (1992).

[5] N. Jain, C. Wittmann, L. Lydersen, C. Wiechers, D. Elser, C. Marquardt, V. Makarov, and G. Leuchs, Phys. Rev. Lett. **107**, 110501 (2011).

[6] D. Stucki, M. Legré, F. Buntschu, B. Clausen, N. Felber, N. Gisin, L. Henzen, P. Junod, G. Litzistorf, P. Monbaron, L. Monat, J.-B. Page, D. Perroud, G. Ribordy, A. Rochas, S. Robyr, J. Tavares, R. Thew, P. Trinkler, S. Ventura, R. Voirol, N. Walenta, and H. Zbinden, New J. Phys. **13**, 123001 (2011).

[7] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, Phys. Rev. A **91**, 062301 (2015).

[8] J.-P. Bourgoin, N. Gigov, B. L. Higgins, Z. Yan, E. Meyer-Scott, A. K. Khandani, N. Lütkenhaus, and T. Jennewein, Phys. Rev. A **92**, 052339 (2015).

[9] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, Opt. Express **19**, 23590 (2011).

[10] A. N. Bugge, S. Sauge, A. M. M. Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Phys. Rev. Lett. **112**, 070503 (2014).