Security Evaluation of Practical Quantum Communication Systems

by

Shihan Sajeed

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering (Quantum Information)

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Name	Hugo Zbinden
	Title	Associate Professor
Supervisor	Name	Vadim Makarov
	Title	Research Assistant Professor
Supervisor	Name	Christopher Wilson
	Title	Professor
Internal Member	Name	Michal Bajcsy
	Title	Assistant Professor
Internal Member	Name	Guo-Xing Miao
	Title	Assistant Professor
Internal-external Member	Name	Michele Mosca
	Title	Professor

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Modern information and communication technology (ICT), including internet, smart phones, cloud computing, global positioning system, e-commerce, e-Health, global communications and internet of things (IoT), all rely fundamentally – for identification, authentication, confidentiality and confidence – on cryptography. However, there is a high chance that most modern cryptography protocols will be annihilated upon the arrival of quantum computers. This necessitates taking steps for making the current ICT systems secure against quantum computers. The task is a huge and time-consuming task and there is a serious probability that quantum computers will arrive before it is complete. Hence, it is of utmost importance to understand the risk and start planning for the solution now.

At this moment, there are two potential paths that lead to solution. One is the path of post-quantum cryptography: inventing classical cryptographic algorithms that are secure against quantum attacks. Although they are hoped to provide security against quantum attacks for most situations in practice, there is no mathematical proof to guarantee unconditional security ('unconditional security' is a technical term that means security is not dependent on a computational hardness assumption). This has driven many to choose the second path: quantum cryptography (QC).

Quantum cryptography – utilizing the power of quantum mechanics – can guarantee unconditional security in theory. However, in practice, device behavior varies from the modeled behavior, leading to side-channels that can be exploited by an adversary to compromise security. Thus, practical QC systems need to be security evaluated – i.e., scrutinized and tested for possible vulnerabilities – before they are sold to customers or deployed in large scale. Unfortunately, this task has become more and more demanding as QC systems are being built in various style, variants and forms at different parts of the globe. Hence, standardization and certification of security evaluation methods are necessary. Also, a number of compatibility, connectivity and interoperability issues among the QC systems require standardization and certification which makes it an issue of highest priority.

In this thesis, several areas of practical quantum communication systems were scrutinized and tested for the purpose of standardization and certification. At the source side, the calibration mechanism of the outgoing mean photon number – a critical parameter for security – was investigated. As a prototype, the pulse-energy-monitoring system (PEMS) implemented in a commercial quantum key distribution (QKD) machine was chosen and the design validity was tested. It was found that the security of PEMS was based on flawed design logic and conservative assumptions on Eve's ability. Our results pointed out the limitations of closed security standards developed inside a company and highlighted the need

for developing – for security – open standards and testing methodologies in collaboration between research and industry.

As my second project, I evaluated the security of the free space QKD receiver prototype designed for long-distance satellite communication. The existence of spatial-mode-efficiency-mismatch side-channel was experimentally verified and the attack feasibility was tested. The work identified a methodology for checking the spatial-mode-detector-efficiency mismatch in these types of receivers and showed a simple, implementable countermeasure to block this side-channel.

Next, the feasibility of laser damage as a potential tool for eavesdropping was investigated. After testing on two different quantum communication systems, it was confirmed that laser damage has a high chance of compromising the security of a QC system. This work showed that a characterized and side-channel free system does not always mean secure; as side-channels can be created on demand. The result pointed out that the standardization and certification process must consider laser-damage related security critical issues and ensure that it is prevented.

Finally, the security proof assumptions of the detector-device-independent QKD (ddiQKD) protocol – that restricted the ability of an eavesdropper – was scrutinized. By introducing several eavesdropping schemes, we showed that ddiQKD security cannot be based on post selected entanglement. Our results pointed out that testing the validity of assumptions are equally important as testing hardware for the standardization and certification process.

Several other projects were undertaken including security evaluation of a QKD system against long wavelength Trojan-horse attack, certifying a countermeasure against a particular attack, analyzing the effects of finite-key-size and imperfect state preparation in a commercial QKD system, and experimental demonstration of quantum fingerprinting. All of these works are parts of an iterative process for standardization and certification that a new technology – in this case, quantum cryptography– must go through before being able to supersede the old technology – classical cryptography. I expect that after few more iterations like the ones outlined in this thesis, security of practical QC will advance to a state to be called unconditional and the technology will truly be able to win the trust to be deployed on large scale.

Acknowledgements

I would like to thank a few people for making my time at the Institute for Quantum Computing in University of Waterloo beautiful. First and foremost, I owe my deepest gratitude to my supervisor, Dr. Vadim Makarov, who have supported me always. Since, my arrival here in waterloo back in 2013, till finishing of this thesis, he was the one person that was always there. I thank him and wish him the very best of luck.

I would also like to thank Prof. Christopher Wilson for being my co-supervisor. I would also like to extend my thanks to Prof. Michele Mosca, Marcos Curty, and Thomas Jennewein for providing me suggestions when I needed them. Thanks to Prof. Michal Bajcsy, Michele Mosca and Guo-Xing Miao for being my committee member. I am also thankful to Dr. Hugo Zbinden for agreeing to be the external examiner of my thesis. I would also like to extend my thanks to Norbert Lütkenhaus. He has helped me every single time when I required his help and was never bothered. His composure and willingness to help people (like me) always leaves me in awe.

Pursuing Ph.D. is a long and hard task, and often it feels boring. That is the time when you need people around you to spend time with to run away from research to recharge. I was fortunate to have few such people around me. First of all, I would mention members from my group: Poompong Chaiwongkhot, Anqi Huang, Jin Gyu Lim, Hao Qin, Elena Anisimova, Paulo Vinicius Pereira Pinheiro, Shaeera Rabbanee Shuvra. Without you guys, my time at the quantum hacking lab would have been pretty boring. Then there are other IQC members: Rubayet, Eima, Hemant, Jean-Philippe Bourgoin, Yanbao with whom I have spent countless number of hours. I would specially thank Jean-Philippe Bourgoin, who helped me a lot throughout my Ph.D. career. Outside IQC, I had the privilege to collaborate with Feihu Xu, Shihai Sun, Nitin Jain and would like to thank them all.

I would also like to thank my father, mother, mother-in-law, sister, sister-in-law and my whole family back in Bangladesh for their endless encouragement and support.

Last but not the least, I would like to thank my wife Nigar Sultana. While I was putting most of my times into my work, she was the one that took care of all other things – works at home, looking after our daughter, taking care of our families back at home etc. – at the same time continuing her Ph.D. I do not have words enough to appreciate the efforts she put into easing my life during the last five years. I feel lucky to have her beside me.

Shihan Sajeed
29 September 2017

Dedication

This thesis is dedicated to my daughter, Sabriyah Sajeed Eira, for being the central source of happiness in my life.

Table of Contents

List of Figures

Chapter 1

Introduction

What is cryptography? In short, the answer to this question is: cryptography is the art of secure communication. In this field, a sender, Alice, needs to send a message to a receiver, Bob, over a communication channel, in such a way that an eavesdropper, Eve, having full access and control over that channel, cannot gain any knowledge on the message. This is typically done by first encrypting the message ‘m’ at Alice with a secret key ‘k’ that Alice and Bob share beforehand, and then sending the encrypted cryptogram ‘C’, over the communication channel. Bob, receiving ‘C’, uses the secret key ‘k’ to decrypt the message. It is assumed that the eavesdropper generally knows all the rules of the protocol and other relevant information about Alice and Bob’s system; except the key ‘k’ [1]. Thus, the science of cryptography evolves around making rules, creating algorithms, and setting protocols that makes it impossible or at least very difficult for Eve, to decrypt the cryptogram without the key.

It is well-known that as long as Alice and Bob share a secret key, secure cryptography is possible using the one-time-pad (OTP) protocol [2]. In this protocol, Eve has no way to gain any information about the message unless she has information on the key. This renders the task of secure communication equivalent to that of distributing a key securely between the two communicating parties. It is also one of the main challenges for today’s cryptography.

In classical cryptography, this key exchange is achieved via protocols with unproven mathematical assumptions that rely on either algorithmic or computational complexity [3]. As a result, if the eavesdropper keeps a copy of the cryptograms and wait for a sufficiently long time until her computational power has increased sufficiently; then it will be possible for her to decode her stored messages in the future. Also, quantum mechanics have shown a way by which a quantum computer, capable of performing quantum algorithms, can solve certain hard problems exponentially faster than their classical counterparts [4]. When these computers are functional, the security of classical cryptography, based on algorithmic or computational complexity, will be under great threat. In other words, the closer quantum computers come to reality, the nearer current information age heads towards being endangered.

Although recent advances indicate that quantum computers, capable of performing quantum algorithms [4], might become a reality one day, there is a high chance that it will take more than a decade or two. This might lead one to take the optimistic route and stay indifferent to the inevitable threat. This is not a well-thought-out strategy for such a hugely important issue. If certain new cryptographic schemes, resistant against the attacks from quantum computers, need to replace today's market, it will not be able to do so overnight. History shows us that whenever a new technology tries to replace an existing technology, it takes years for the governments, industry and general customers to settle into it. Any new replacement, however obvious, needs to be scrutinized both from theoretical and practical point of view and must withstand multiple challenges before it can be considered reliable and win the trust of mass consumer. At this moment, it is not known how long it will take for a new cryptographic solution to achieve that level of trust. We also do not know whether it can happen before the arrival of functioning quantum computers. But one thing is sure that the community needs to start acting now in order to give it the best shot to prevent the eminent danger.

At this moment, there are two potential solutions at hand. One is switching to post-quantum cryptography: cryptographic algorithms that are secure against quantum attacks. These algorithms rely on hard problems not expected to yield to quantum computers. Some variants of these are the Hash-based cryptography: Merkle's hash-tree public key signature system [5], code-based cryptography: McEliece's hidden-Goppa-code public-key encryption system [6], Lattice-based cryptography: Hoffstein-Pipher-Silverman 'NTRU' public-key-encryption system [7], Multivariate-quadratic-equations cryptography: Patarin's HFE^{v-} public-key-signature system [8] etc. Although it is hoped that they are secure enough to protect secrets from quantum attacks for most situations in practice, there is no mathematical proof that they are unconditionally secure. This is where quantum mechanics comes to the rescue by providing a second solution.

The idea that quantum mechanics can be used for cryptographic purposes was first proposed by Stephen Wiesner in 1970 [9]. However, the idea was not accepted at that time and appeared as a publication more than a decade later in 1983 [10]. The term 'quantum cryptography (QC)' was first coined by Bennett and Brassard in 1984 [11] where they utilized two quantum mechanical properties – no cloning theorem [12] and uncertainty principle – into public key distributions and coin tossing. Since then QC has started to draw attention of a much larger community and that seminal protocol became known as the Bennett and Brassard 1984 (BB84) protocol. In 1991, while elaborating on a suggestion from Deutsch [13], Artur Ekert in Oxford independently discovered another version of QC [14]. His proposed scheme was based on the Bohm's version of the Einstein-Podolsky-Rosen gedanken experiment [15, 16] and used the Bell's theorem [17, 18] to test for security. The great significance of this work was that it took, what was initially thought as 'philosophical debate or idea' [15, 17], and turned it in clever way to find applications in the field of applied science: quantum cryptography.

With time, more variants of quantum cryptography – in different flavours and forms – had been reported in the literature. However, the protocols required the two parties to have a pre-shared secret for authentication purposes and to be more specific, were actually key expansion protocols. Nowadays, the term 'quantum key distribution (QKD)' is more popularly used to refer these key expansion protocols and the term 'quantum cryptography' has taken a broader and more general meaning which includes protocols other than QKD such as quantum coin tossing [19, 20], quantum fingerprinting [21, 22, 23, 24], quantum teleportation [25, 26, 27, 28], blind quantum computing [29] etc. Since, the majority of this thesis is focused on QKD protocols, I will provide the history and development of QKD protocol in detail. The other primitives will be introduced later in brief.

1.1 History and development

The first experimental implementation of QKD was reported in 1992 [30] which was an implementation of BB84 protocol. In 1992, the known QKD protocols either considered four non-orthogonally polarized single-photon or low-intensity light pulses or polarization-entangled two-photon states. Bennett in 1992 showed that in principle only two non-orthogonal quantum states suffice for QKD, and proposed a scheme, the Bennett 1992 (B92) protocol, to realize QKD with it [31]. While two states were sufficient and four states were standard, QKD employing six states were also proposed [32, 33]. Another version where encoding of the qubits was performed in bases rather than in quantum states was proposed by Scarani-Acín-Ribordy-Gisin in 2004 (SARG2004) protocol [34]. Eventually, the number of possible protocols were divided into three main families: discrete-variable (DV) protocols, continuous-variable (CV) protocols and distributed-phase-reference (DPR) protocols. DV protocols, in principle, uses discrete quantum degrees of freedom for the qubits and their detection mechanism are based on photon counting technology. The protocols mentioned so far, i.e., BB84 [11], Ekert [14], B92 [31], SARG [34], six state [33, 32], all fall within this family of QKD protocols. At present, the longest achieved distance with DV QKD employing fibers is 404 km [35]. QKD schemes over free space promise to achieve much larger distance and already entangled photons have been distributed over a distance of 1200 km via satellite [36] which is still the record for the highest distance.

A problem with the discrete variable QKD protocols is that they require photon counters that typically have low efficiency, high dark count rates and long dead times. To overcome these limitations, an alternative approach utilizing standard telecom p-i-n photo diodes and homodyne detection was proposed. This was the Continuous variable (CV) QKD. The first homodyne detection based QKD was proposed in Ref. [37, 38, 39]. These protocols were based on a discrete modulation of Gaussian states. In 2001, a CV protocol was introduced by Cerf et. al., that was based on the continuous modulation of squeezed states [40]. This idea was soon extended to coherent states with the introduction of the Gaussian modulated coherent state protocols [41, 42]. Now CV QKD is well-established from both theoretical and practical point of view with implementations like Refs. [43, 44, 45, 46, 47, 48, 49, 50].

The third family of the QKD protocols is the distributed-phase-reference protocol family. They lie in the middle of the DV and CV protocols. Like the DV protocols, the raw key is composed of discrete degrees of freedom of a quantum state and are perfectly correlated in the absence of errors. But the quantum channel is monitored using the phase coherence of subsequent pulses. The first of such protocols was the differential phase shift (DPS) protocol [51, 52] which had already been implemented [53, 54, 55]. Another variant is the coherent one way (COW) protocol [56] which has a full implemented system prototype [57, 58, 59]. A version of this protocol is even being commercialized [60].

As different versions of QKD were being reported, so was the security proofs for unconditional security using different techniques [61, 62, 63, 64, 65, 66, 67, 68, 69]. However, device models used in the security proofs have often differed from the properties and behavior of the actual equipment. This opened exploitable security loopholes like the time shift attacks [70, 71], detector control attacks [72, 73, 74, 75, 76], dead-time attack [77], phase-remapping attack [78], detector efficiency mismatch attacks [79, 80, 71], Trojan-horse attacks [81, 82, 83], attacks involving system calibration [84], photon number splitting attacks [85, 86], intercept-resend attacks [87, 88, 89, 90], laser damage attacks [91] etc. Attacks were also reported for the case of CV QKD [92, 93, 94, 95].

Most of the attacks were counter-acted by either physical countermeasures [96, 59], modified QKD protocols [97, 98, 99], or modified security proofs incorporating the imperfection of the device into the model [100, 101]. Over the years, these approaches have led to three different types of solutions. The first one was to precisely model the behavior of all the devices that are used in the implementation [101, 102]. Unfortunately, this approach is very difficult to realize in practice as behavior of practical devices are very difficult to characterize precisely and the characterization complexity increases with increasing number of devices.

The second approach is designing countermeasures to known attacks. Fortunately, when a side channel is discovered, it is often relatively easy to come out with a countermeasure to close it. However, for attacks or vulnerability that has not yet been discovered, this approach still raises a deep security concern.

The third solution is the so called device-independent QKD schemes [103, 104, 105, 106]. In this case Alice and Bob can treat their devices as black boxes and rely on the loophole free violation of Bell test [107, 108, 109] for the security. If Alice and Bob can somehow satisfy certain assumptions, this scheme is immune against all the known and unknown side-channels and can provide unconditional security. However, a shortcoming of this scheme is its practical difficulty and the comparatively low key rate. This opened up a way for schemes like measurement-device-independent QKD (mdiQKD) [110] that used a time-reversed version of EPR scheme [111, 112] and detector-device-independent QKD (ddiQKD) [113, 114, 115, 116] that used the idea of two-qubit single-photon Bell state measurement [117]. Among them mdiQKD promises to guarantee security based on post selected entanglement and provides security against all side-channel attacks in the detector. It has also been implemented in the labs [118, 119, 120, 121, 122, 123, 124, 118, 125, 126] and the reported key rates and distances are getting higher.

Other than QKD, there are other primitives of quantum cryptography that has attracted much attentions. Among them, one is quantum coin flipping: where two distrustful parties separated by a large distance wish to agree on a random bit. It has already been experimentally demonstrated in the lab [19, 20]. Another primitive, Quantum oblivious transfer – the task of transmitting one of potentially many information to the receiver with the latter being oblivious about it - has also been implemented [127]. Quantum teleportation – transmitting an unknown quantum state from one location to another – was first proposed in Ref. [25]. Since then, it has been demonstrated experimentally numerous times [26, 27] and now a teleportation distance of over 1400 km [28] has been achieved. Another primitive is quantum fingerprinting. Fingerprinting is the task of assigning two long and distinct strings with a shorter string - their fingerprints - so that the two long strings can be compared with high accuracy with the fingerprints alone. This has been achieved using quantum resources: quantum fingerprinting [21, 22, 23, 24]. Besides offering exponential computing power, quantum computers are also expected to offer the privacy of the computation and the users - a cryptographic task known as the blind quantum computation (BQC). It is already experimentally demonstrated in Ref. [29] and is crucial for real life applications in the form of quantum cloud computing.

At present, QC is not only being implemented on the lab but also started to leap out of it with efforts from both government and industry. We will discuss the globalization of QC next.

1.2 Globalization of quantum cryptography

The Swiss company ID Quantique, that originally started as a small spin off by four researchers from University of Geneva in 2001, has been able to achieve the world's first ever deployment of quantum key distribution in a live network environment, the first deployment of QKD over a multiplexed network for commercial clients and the longest-running quantum crypto installation (since 2007) [128]. At present, they are providing quantum-safe cryptographic solutions including quantum key generation [129], quantum key distribution [130, 96, 60] and quantum-safe network encryption [131] to customers that include government, industry and academia.

A major research effort from 41 research and industrial organizations from the European Union, the 'Development of a Global Network for SEcure COmmunication based on Quantum Cryptography (SECOQC), was initiated in 2003 with the aim of establishing practical applications of QKD. A summary of the SECOQC approach to QKD networks can be found in Ref. 132.

China is becoming a leading player in the globalization of quantum communication. It has already launched a quantum-enabled satellite Micius in 2016 which is the first of its kind. The satellite has now been used for implementation of several quantum cryptography primitives: entanglement distribution [36], QKD [133], quantum teleportation [28]. She is also building a 2000 km link between Beijing and Shanghai containing 32 nodes that will also connect Hefei and Jinan. The project is well underway and when built, it will the longest of its kind [134].

As part of the €270 million investment of the UK national quantum technologies program, the UK quantum technology hub for quantum communications is formed. It is a partnership of eight UK universities, several private companies (BT, Toshiba research Europe Ltd, the National Physical Laboratory etc.) and public sector bodies (Bristol City Council and the National Dark Fiber Infrastructure Service). The aim of this huge collaboration is to work for the secure communication technologies and provide secure quantum encryption systems to government, industry and wider public community.

Telecom giant BT Group and Toshiba Research Europe are also collaborating on a major project to build a Quantum Communication network (the 'UK Quantum Network'). This initiative will facilitate quantum secured communications between Cambridge, Bristol, London and Adastral Park. A link connecting BT's Labs at Adastral Park and the Cambridge Science Park is expected to be completed within year 2017.

In Canada, the government has recently proposed to provide $80.9 million over five years to the Canadian Space Agency. A part of this funding is to support demonstration of applications of quantum technologies in space. The institute for Quantum Computing (IQC) at University of Waterloo is highly involved with this project, and this project aims to put Canada as one of the leaders in quantum communication and encryption.

Europe is also joining the globalization of quantum communication by announcing a €1 billion project to boost the research on quantum technologies. The initiative – EU quantum technologies flagship – is scheduled to launch in 2018 and aims to exploits the power of quantum mechanics to develop new technologies including super-secure communication systems and miniature, ultra-accurate sensors [135].

In Asia, the Korean company SK telecom announced that it has successfully developed the country's first ever long-distance quantum repeater and completed a field test on optical fiber network across three cities in Gyeoggi achieving a distance of 112 km. The company also applied its quantum cryptography technologies to the commercial long term evolution (LTE) network in the city of Sejong and plans to expand over other areas [136]. In Japan, Toshiba is applying secure quantum encryption to transmit genome data.

As quantum cryptography systems are being pushed from the lab to practical deployments in different parts of the globe, a number of security, compatibility and connectivity issues need to be solved – demanding the need for standardization and certification. A common platform to address and work on these issues was needed. As a result, the European Telecom Standard Institute (ETSI) has formed a group: the ETSI industry specification group for QKD (ISG-QKD). ISG-QKD provides a platform for creation of universally accepted standards and promotes coordination, cooperation and standardization of research for QKD. Their published paper [137] aims to raise awareness of the potential impacts of quantum computing on information security globally and contains a survey of current cryptographic principles, the possible impact of quantum computing on their effectiveness, and what can be done to mitigate the risks in an economically and technically practical manner [138].

All these are exciting news for the globalization of QKD. As government and industry become more aware of the importance of quantum-safe solutions to cryptographic problems, there is a high chance that more participation will occur, and more funds will be allocated for furthering the technology. This will surely help in globalization of the technology.

1.3 Motivation

My Ph.D. research was motivated by the following question: how can practical quantum cryptography systems be ready to provide the unconditional security claimed by its theories? At this moment it cannot as practical device behavior often cannot follow the exact modeled behavior required by the security proofs, leading to exploitable side-channels. More worrying issue is that more often than not, before a side-channel was discovered, it was not even predicted that such a side-channel could exist.

The problem can be looked from another point of view: the unconditional security of practical quantum cryptographic schemes is based on three aspects - laws of physics, security proofs and correct modeling of the implemented devices. The laws of physics that are used in the theory are correct. Security proofs have been scrutinized and are found solid. The only vulnerability left is on the correct modeling of the implemented devices: either their behavior cannot exactly follow the expected behavior or there are other side-channels that leak security-critical information to the eavesdropper to compromise the security.

Thus, practical systems should undergo thorough testing, before they are labeled with a tick mark. But the problem is: what part of the system needs to be tested? Who is going to do the test? How to make sure the testing methodology is correct? If someone performs a test on a system and label it with a tick mark, will it be trusted by others? Even if we assume that we have the correct testing methodology, there are still problems: so many quantum cryptographic solutions are being built at different parts of the globe with different components, protocols and manufacturers – how to even test so many systems? All these demands a common platform for the standardization and certification for quantum cryptographic systems. As mentioned in section 1.2, ISG-QKD group is already formed and has started working on the standardization and certification of the QKD systems. However, the task is too big and the time is short. Thus, it is required by independent research groups to contribute to the testing, standardization and certification. This motivated me to examine practical quantum communication systems and contribute to the standardization and certification process by identifying areas of vulnerability, demonstrating the existence of side-channels, showing the correct testing methodology and proposing countermeasure and testing it.

The first step of the standardization and certification process is to identify the possible areas of vulnerability and perform tests to find the existence of loopholes. For this purpose, I have investigated several quantum communication systems and looked into several parts of the system including source, detectors and assumptions in the protocols. At the source side, I investigated the calibration mechanism of the outgoing mean photon number because if it deviates from the expected value, security might be compromised. As a prototype, I chose the commercial QKD system Clavis2 from ID Quantique [96]. Although Clavis2 implements a pulse-energy-monitoring system (PEMS) to monitor the mean photon number, the design validity of the systems was untested. Thus, a security evaluation of the PEMS along with a study of the effect of mean photon number deviation on the system security was necessary which provided me the necessary motivation.

As my second project, I focused on the long-distance free-space quantum communication systems that are currently a hot topic [139, 140, 28, 133]. As a prototype, I chose the free space QKD receiver prototype designed for long-distance satellite communication [140]. An assumption of security proofs is that the measurement outcome should not depend on the measurement basis and Eve does not have any ability to control or force a particular detection outcome over another. Due to a particular side-channel – spatial-mode-efficiency-mismatch – the validity of this assumption for many receivers might be called into question. The test to identify this vulnerability and evaluating the security of the receiver against this particular side-channel was never done before. This motivated me to start a security evaluation of this receiver.

There is an assumption with practical QKD systems that if the implemented devices are properly characterized and all known side channels are closed, the implemented system is secure. However, in Ref. [141], it was shown that the application of high power laser changed several properties of the avalanche photodiode-based detector: dark count rate, detector efficiency, quantum efficiency and dark current. What if the eavesdropper can change the properties of some other security critical components? Are the changes made on-demand adequate to compromise the security? Is it possible for an eavesdropper to create a new side-channel by the application of high power laser? In order to answer these questions, thorough scrutiny is required. This motivated me undertake the project of analyzing the effect of laser damage on the security of practical quantum communication systems.

Finally, I looked at the detector-device-independent QKD protocol. It was proposed with the promise of being easily implementable along with providing device independent security at the detector side. However, its security proofs were based on assumptions that restricted the ability of an eavesdropper. It was not clear how secure the protocol was, had the ability of the eavesdropper been not restricted. This motivated me to look into the protocols.

In short, the main motivation of my Ph.D. research was to evaluate the security of several aspects of quantum communication systems and contributing to the standardization and certification process.

1.4 Outline

This thesis is organized as follows. In chapter 2, the necessary backgrounds to understand later parts of the thesis are provided. In chapter 3, effect of deviation of mean photon number on the security of quantum communication has been studied along with the security evaluation of a pulse energy monitoring scheme implemented in a commercial system. Chapter 4 describes the security evaluation of free space QKD system. In chapter 5 we introduce laser damage as a new tool for the eavesdropper to break the security of practical QC systems and experimentally demonstrate its practicality on two different systems. In chapter 6 it is shown that the newly proposed ddiQKD scheme is not as secure as was initially thought and several practical eavesdropping schemes to support this claim are presented. I have briefly discussed several other projects in which I participated during my Ph.D., (but did not lead) in chapter 7. The conclusions are given in chapter 8.

1.5 List of contributions

A number of projects are included in this thesis. All of them were collective efforts from a number of researchers. Below I outline my contributions in those projects along with others and indicate the chapters where the each of the projects are described.

1. **Effect of deviation of mean photon number on the security and security evaluation of pulse energy monitoring system:** The security of many quantum communication protocols require that the outgoing mean photon number μ be characterized and bounded. Hence, many practical quantum communication systems employ a 'pulse energy monitoring system (PEMS)' [96, 19] to constantly monitor μ. As my first project, I scrutinized the practical security of such a pulse energy monitoring system employed in the commercial QKD system Clavis2 [96]. I, along with Igor Radchenko, found three design flaws in the PEMS that allows an eavesdropper to inject extra photons into the system without triggering any alarm. We experimentally demonstrated the feasibility of it. Then I participated with Jean-Philippe Bourgoin in modeling several attacks and studying the effect of deviation of μ on the

security of two quantum key distribution protocols. I extended the theoretical analysis to quantum coin-tossing protocols with help from Anna Pappa and studied how deviation of μ can affect the cheating probability in quantum coin tossing protocol. The descriptions are included in chapter 3 and the results are published in Ref. 142. I am the first author of this paper.

2. **Security evaluation of free space QKD systems:** In free space quantum communication, the sensitivity of receiver's detectors may depend on the spatial modes of the incoming photon. An eavesdropper may take advantage of this and render the system insecure. The feasibility of this attack was identified, analyzed, and experimentally tested during my second project. I performed the initial investigation on the free space QKD receiver brought from Thomas Jennewein's lab and planned the initial design of the experiment. The experiments to demonstrate the feasibility of the attack and countermeasure testing were carried out in Vadim Makarov's lab by myself, Poompong Chaiwongkhot and Jean-Philippe Bourgoin. The attack modeling, theoretical analyses and all other simulations were carried out by me with help from Poompong Chaiwongkhot. The details of this project are given in chapter 4 and the results are published in Ref. 143. I am the first author of this paper.

3. **Security evaluation of quantum communication systems against laser damage:** As long as implemented devices are properly characterized and side channels are closed, practical quantum communication protocols are assumed to be perfectly secure. In my third project, we have shown that this is not entirely true and demonstrated that laser damage can be a potential tool at eavesdropper's disposal to modify the device behavior on demand. The project involved testing the feasibility of laser damage on two extensively characterized, completely different and widely used implementations: a commercial fiber-optic system for QKD [144, 96] and coin-tossing [19] with phase-encoded qubits, and a free-space system for QKD with polarization-encoded qubits [139]. I, Poompong Chaiwongkhot and Jean-Philippe Bourgoin carried out the free space experiment in Vadim Makarov's lab in University of Waterloo. Sarah Kaiser, me and Mathieu Gagné performed the fiber optic experiment in Raman Kashyap's lab in École Polytechnique de Montréal. The simulations and theoretical analyses for both the tests were performed by me. The details of this project are given in chapter 5 and the results are published in [91]. The author list is alphabetical in this paper and among the students involved in the project, my contribution is the largest according to the author contribution statement of Ref. 91.

4. **Insecurity of Detector-Device-Independent Quantum Key Distribution** Detector-device-independent-QKD (ddiQKD) was proposed with the promise of be-

ing robust against detector-side-channel-attacks. However, contrary to the claim, Me (Shihan Sajeed) along with Anqi Huang, developed several attack strategies that showed ddiQKD is in fact insecure against detector side channel attacks. I conceptualized the attacks exploiting trigger-pulse-energy-threshold difference under different blinding conditions and imperfect phase modulation. Anqi Huang tested detector efficiency mismatch under detector blinding attack and proposed the attack of exploiting imperfect beam splitter. I and Anqi Huang contributed equally to this work. The details are provided in chapter 6 and the results are published in Ref. 145.

1.6 List of publications

The following journal papers have resulted from my Ph.D. research:

[1] **S. Sajeed**, C. Minshull, N. Jain, and V. Makarov. Invisible trojan-horse attack. *Sci. Rep.*, 7:8403, 2017.

[2] P. Chaiwongkhot, **S. Sajeed**, L. Lydersen, and V. Makarov. Finite-key-size effect in a commercial plug-and-play qkd system. *Quantum Science and Technology*, 2(4):044003, 2017.

[3] **S. Sajeed**, A. Huang, S. Sun, F. Xu, V. Makarov, and M. Curty. Insecurity of detector-device-independent quantum key distribution. *Phys. Rev. Lett.*, 117:250505, 2016.

[4] A. Huang, **S. Sajeed**, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov. Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption. *IEEE Journal of Quantum Electronics*, 52(11), 2016.

[5] V. Makarov, J-P. Bourgoin, P. Chaiwongkhot, M Gagné, T. Jennewein, S. Kaiser, R. Kashyap, M. Legré, C. Minshull, and **Sajeed, S.** Creation of backdoors in quantum communications via laser damage. *Phys. Rev. A*, 94:030302, 2016.

[6] **S. Sajeed**, I. Radchenko, S. Kaiser, J-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov. Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing. *Phys. Rev. A*, 91:032326, 2015.

[7] **S. Sajeed**, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov. Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch. *Phys. Rev. A*, 91:062301, 2015.

[8] F. Xu, K. Wei, **S. Sajeed**, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H-K Lo. Experimental quantum key distribution with source flaws. *Phys. Rev. A*, 92:032305, 2015.

[9] F. Xu, J. M. Arrazola, K. Wei, K. Wang, P. palacios Avila, C. Feng, **S. Sajeed**, N. Lütkenhaus, and H-K. Lo. Experimental quantum fingerprinting with weak coherent pulses. *Nat. Commun.*, 6:8735, 2015.

1.7 List of conference presentations

The following conference presentations have resulted from my Ph.D. research:

1.7.1 Delivered by me

1. **S. Sajeed,** A. Huang, S. Sun, F. Xu, V. Makarov, and M. Curty, Insecurity of detector-device-independent quantum key distribution, 6th International Conference on New Frontiers in Physics, Crete, Greece, August 17-29, 2017 (**3rd prize for best poster**).

2. **S. Sajeed,** C. Minshull, N. Jain, and V. Makarov, Invisible Trojan-horse attack, 6th International Conference on New Frontiers in Physics, Crete, Greece, August 17-29, 2017 (**contributed talk**).

3. **S. Sajeed,** C. Minshull, N. Jain, and V. Makarov, Invisible Trojan-horse attack, QCrypt 2017, Cambridge, UK, 18-22 September 2017 (**accepted as a poster**).

4. **S. Sajeed,** P. Chaiwongkhot, M. Gagné, J.-P. Bourgoin, C. Minshull, M. Legré, T. Jennewein, R. Kashyap, and V. Makarov, Laser damage creates backdoors in quantum cryptography, SPIE Laser Damage Symposium, Boulder, Colorado, USA, September 2016 (**contributed talk**).

5. **S. Sajeed,** P. Chaiwongkhot, M. Gagné, J.-P. Bourgoin, C. Minshull, M. Legré, T. Jennewein, R. Kashyap, and V. Makarov, Laser damage creates backdoors in quantum cryptography, presented at QCrypt 2016, Washington DC, USA, September 12-16, 2016 (**contributed talk**).

6. **S. Sajeed,** P. Chaiwongkhot, M. Gagné, J.-P. Bourgoin, C. Minshull, M. Legré, T. Jennewein, R. Kashyap, and V. Makarov, Laser damage creates backdoors in quantum cryptography, presented at Trustworthy Quantum Information (TyQI) Workshop, Shanghai, China, June 27-30, 2016 (**poster**).

7. **S. Sajeed**, I. Radchenko, S. Kaiser, J-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Securing two-way quantum communication: the monitoring detector and its flaws, QCrypt 2014, Paris, France, September 1-5, 2014 **(contributed talk)**.

1.7.2 Delivered by my coauthors

1. P. Chaiwongkhot, K. B. Kuntz, A. Huang, J.-P. Bourgoin, **S. Sajeed,** N. Lütkenhaus, T. Jennewein, and V. Makarov, Effect of atmospheric turbulence on spatial-mode detector-efficiency mismatch, QCrypt 2017, Cambridge, UK, 18-22 September 2017 **(accepted as a poster)**.

2. S. Sajeed, C. Minshull, **N. Jain,** and V. Makarov, Insecurity of practical quantum key distribution against long-wavelength Trojan-horse attacks, presented at Central European Workshop on Quantum Optics 2017, Lyngby, Denmark, 26-30 June 2017 **(poster)**.

3. **P. Chaiwongkhot,** S. Sajeed, L. Lydersen, and V. Makarov, Finite-key-size effect in commercial plug-and-play QKD system, presented at QCrypt 2016, Washington DC, USA, September 12-16, 2016 **(poster)**.

4. **A. Huang,** S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, An advanced Eve of QKD: breaking a security assumption and hacking a black box, presented at QCrypt 2016, Washington DC, USA, September 12-16, 2016 **(poster)**.

5. **A. Huang,** S. Sajeed, S. Sun, F. Xu, V. Makarov, and M. Curty, Insecurity of detector-device-independent quantum key distribution, presented at QCrypt 2016, Washington DC, USA, September 12-16, 2016 **(contributed talk)**.

6. **P. Chaiwongkhot,** S. Sajeed, L. Lydersen, and V. Makarov, Finite-key-size effect in commercial plug-and-play QKD system, presented at Trustworthy Quantum Information (TyQI) Workshop, Shanghai, China, June 27-30, 2016 **(poster)**.

7. **A. Huang,** S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, Gaps between industrial and academic solutions to implementation loopholes in QKD: testing random-detector-efficiency countermeasure in a commercial system, presented at Trustworthy Quantum Information (TyQI) Workshop, Shanghai, China, June 27-30, 2016 **(poster)**.

8. **P. V. P. Pinheiro,** P. Chaiwongkhot, S. Sajeed, R. T. Horn, J.-P. Bourgoin, and V. Makarov, Measurements of light emission from silicon avalanche photodetectors, presented at QCrypt 2015, Tokyo, Japan, September 28 - October 2, 2015 **(poster)**.

9. **P. Chaiwongkhot,** S. Sajeed, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, Spatial-mode detector efficiency mismatch security loophole in free-space QKD, presented at QCrypt 2015, Tokyo, Japan, September 28 - October 2, 2015 **(contributed talk)**

10. **A. Huang,** S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, Gap between industrial and academic solutions to implementation loopholes: testing random-gate-removal countermeasure in commercial QKD system, presented at QCrypt 2015, Tokyo, Japan, September 28 - October 2, 2015 **(poster)**.

11. **F. Xu,** S. Sajeed, S. Kaiser, Z. Tang, V. Makarov, and H.-K. Lo, Experimental quantum key distribution with source flaws and tight finite-key analysis, presented at QCrypt 2014, Paris, France, September 1-5, 2014 (contributed talk), **(best student presentation award)**.

Chapter 2

Backgrounds

The main results of this thesis is presented in chapters 3 to 6, each of which contains its own more specific theory section. However, there are some theories, protocols, attacks, and schemes that are common in more than one chapter. They are presented in this chapter. We start this chapter by introducing the first QKD protocol in section 2.1 along with describing the steps to calculate the key rates. Next we present several eavesdropping schemes that are relevant to this thesis.

2.1 The first QKD protocol: Bennett-Brassard 1984 (BB84) protocol

The Bennett-Brassard 1984 (BB84) protocol was first proposed in 1984 [11]. It uses two properties of quantum mechanics: uncertainty principle and no-cloning theorem to provide a cryptographic scheme that provides unconditional security. The steps of the protocol are described below:

1. **Preparation:** Alice randomly selects a basis out of a set of two predefined bases, and randomly selects an eigenstate of that basis. In other words, she randomly prepares a sequence of quantum states drawn from a set of four predefined states. For example, she can choose to encode her bit in the polarization degree of freedom of a photon. In that case, the bases can be horizontal-vertical or diagonal-antidiagonal corresponding to X and Z basis in the Bloch sphere respectively; the signal set contains horizontal

(H), vertical (V), +45° (D), and −45° (A) polarization of a photon. After preparation, Alice's state can be described as,

$$|\psi\rangle_A = (|\mathrm{i}\rangle + e^{i\theta_\mathrm{A}}\,|\mathrm{j}\rangle)/\sqrt{2}, \tag{2.1}$$

where $|\mathrm{i}\rangle$ ($|\mathrm{j}\rangle$) denotes two orthogonal vectors forming the Y basis in the Bloch sphere and the phase $\theta_\mathrm{A} \in \{0, \pi/2, \pi, 3\pi/2\}$ provides the mapping into four polarization states.

2. **Transmission:** Alice sends her prepared states to Bob through a quantum channel. Physically, the channel can be any fiber optic or free space channel.

3. **Measurement:** For each incoming signal, Bob makes a measurement in one of the two predefined bases chosen randomly and his result can be described as $|\psi\rangle_B$. This ends the quantum part of the protocol.

4. **Sifting:** For each of the exchanged signals, both party disclose to each other the corresponding basis used for that signal. The signals of the slots for which bases matched are kept, and called 'sifted key'. Signals from the remaining slots, for which the bases did not match are discarded. Note that, from this step onwards, all the communications are performed in a classical channel and an authenticated classical channel is mandatory [30]. Otherwise, Eve can sit in the middle of the line and impersonate each of them to the others; resulting in no security.

5. **Parameter estimation:** Both party chooses a fraction of the sifted key and disclose the bit values to each other over the authenticated classical channel. Ideally, the values should match. However, if a fraction of them do not match, it is considered as error. If the error fraction is below an acceptable limit, the protocol is continued. Otherwise, the protocol is aborted.

6. **Error correction:** In this step, Alice and Bob perform a classical error correction protocol to correct for the errors [30]. For this, they have to exchange additional information on their data. At the end of this protocol, both party end up with identical bit strings K_{EC} but on the process the eavesdropper acquired some additional information $I(E.C)$

7. **Privacy amplification:** Having their keys reconciled, the next step is to decouple the eavesdropper from any information she has on the reconciled key. This process is called privacy amplification (PA) at the end of which the initial key K_{EC} is mapped

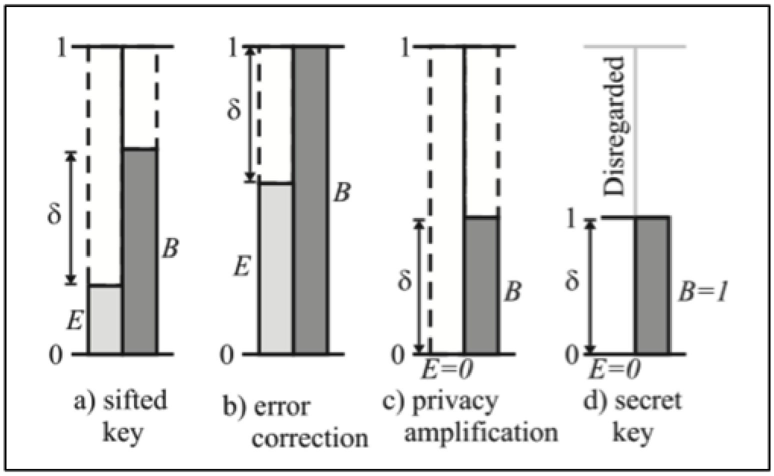


Figure 2.1: The relative key length after sifting, error correction and privacy amplification. Here, Alice's information has been normalized to 1. Bob's and Eve's information have been denoted by B and E respectively. Figure reprinted from [9]

into a shorter, final key K using special families of functions (universal functions). The key size after every steps described in this section is shown in figure 2.1.

2.2 Calculation of key rate

The secret key rate is given by,

$$K = I(A,B) - I(A,E) \tag{2.2}$$

Here, $I(A,B)$ and $I(A,E)$ are the mutual information between Alice–Bob and Alice–Eve, respectively. For BB84, $I(A,B)$ can be given as [146],

$$I_{A,B} = [1 - H(Q)] \tag{2.3}$$

where Q is the error rate and H(Q) is the Shannon entropy,

$$H(Q) = -Q\log_2 Q - (1-Q)\log_2(1-Q) \tag{2.4}$$

$I(A,E)$ is calculated based on the assumption on Eve's strategy. It has been classified into three groups.

2.2.1 Individual attack

- **IA1:** Eve probes each signal going from Alice to Bob independently and her probing strategy stays same throughout the attack.
- **IA2:** Eve must make a measurement before the classical post processing starts. This constraint ensures that before the post processing starts, the symbols at Alice, Bob and Eve are classical and can be described by a joint probability distribution $P(A, B, E)$.

In this case, the bound for Eve's information is given by [146]

$$I(A, E) = max_E I(A, E) \tag{2.5}$$

and similarly for Bob. The notion max_E is used to signify that one has to maximize Eve's mutual information over all of her possible strategies and choose the bound according to that. Some examples of individual attacks are intercept-resend, beam splitting, photon number splitting attack etc.

2.2.2 Collective attack

The main assumptions are:

- **CA1** Eve probes each signal going from Alice to Bob independently and her probing strategy stays same throughout the attack.
- **CA2** Eve can keep her ancillas in a quantum memory and delay her measurement until a later time that is most optimum for her. She can then perform the most optimum measurement that maximizes her knowledge.

In this case, Eve's knowledge bound is calculated by [146],

$$I(A, E) = max_E \ \chi(A.E) \tag{2.6}$$

and I_{BE} is defined similarly. Here, $\chi(A.E)$ is the Holevo quantity defined by,

$$\chi(A.E) = S(\rho_E) - \sum_a p(a) S(\rho_{E|a}) \tag{2.7}$$

where S is the Von Neumann entropy, a is a member from Alice's symbol set having probability $p(a)$, $\rho_{E|a}$ is the state of Eve's ancilla given a and $\rho_E = \sum_a p(a)\rho_{E|a}$.

2.2.3 General or coherent attack

Eve's most general strategy can include anything that is permitted by the laws of quantum mechanics. However, it turns out [146] that the bound is the same as that for collective attacks.

Finally, the secret key rate is given by,

$$K = I(A,B) - min\{I(A,E), I(B,E)\} \tag{2.8}$$

2.3 Attacks in QKD

Several eavesdropping schemes are presented below that are required to understand the latter parts of this thesis.

2.3.1 Intercept-resend attack

Intercept-resend attack (IRA) was first proposed in Ref. [30]. In this attack, Eve sits between Alice and Bob and measures each signal going from Alice to Bob. After her measurement, she prepares a signal identical to her measurement and resend it to Bob. Since her measurement result is classical, she can prepare the signal to compensate for the line loss, efficiency and delay. Assuming BB84 protocol, after raw key exchange and sifting, considering no other sources of error, Eve's measurement basis coincides with that of Alice and Bob during for half of cases. Eve knows full information for this portion of the data and there will be no error in it. However, half of the times, Eve's basis will not coincide with that of Alice and Bob and she will get random results. When she resends her random result to Bob, half of the time Bob will get the correct result and the other half of the time his result will be wrong. Hence, in total, there will be 25% error in Alice and Bob's sifted data and 50% of this information will be known to Eve.

Although, it seems harmless from Alice and Bob's point of view as the resultant error rate is too high, but Eve can be clever and choose a smarter approach. Instead of measuring each of the signals individually, she can measure only 10% of the sifted key. In this case, there will be an error of only 2.5% in the sifted key and Eve will have knowledge of 5% of them. Alice and Bob then must make sure that this partial information is removed during the privacy amplification step.

IRA belongs to the class of individual attacks since Eve interacts with each signal going from Alice to Bob individually. However, there are other known individual attacks that perform better than IRA [67, 147, 148, 149].

2.3.2 Faked-state attack

This attack was first proposed in [150] and later combined with other vulnerabilities in [79, 80, 73, 75, 151, 143, 145]. It is a modified version of IRA where Eve measures each signal going from Alice to Bob. However, she does not try to reconstruct the original signal to send to Bob. Rather, she prepares light pulses that are specially crafted to exploit one or more flaws of Bob's detector scheme to have some kind of control over the detection outcome. An example of faked-state attack can be as follows:

Let us assume that Alice and Bob have finished sifting, i.e., they contain bits for which they have used the same bases. We will now consider Eve in this analysis. We assume, Eve measured each signal and based on her measurement outcome, she sent a faked-state to Bob. We consider two cases: First, when Eve's measurement bases coincide with that of Alice-Bob. In this case, she obtains the right results and wants her faked state to be detected by Bob. On the other cases, when Eve's bases do not match with that of Alice-Bob, she wants her faked state to be undetected. Thus, the success or failure of a faked-state attack relies on whether Eve can control the detection outcomes based on the matching-mismatching of her bases with Alice-Bob; or more specifically with Bob. In other words, whenever Bob's basis matches (does not match) that of Eve, a detection (no detection) is required. By satisfying this, the resultant error can ideally be made to zero and Eve can have full information of the key. In practice, if there is no way to achieve the desired outcome for a perfect faked-state attack, Eve can perform some kind of partial faked-state attack and obtain partial information about the key. Many of the attacks, mentioned later in this thesis, uses faked-states for better attack performance (see chapters 3 to 6).

2.3.3 Detector-efficiency mismatch attack

The detectors employed in the detection assembly are assumed to be identical in characteristics; in practice however, they are not. There can be slight mismatch in their actual characteristics that can lead to a compromise in security. This type of attack was first reported in Ref. [79]. The attack exploited mismatch in detector efficiency characteristics as described next.

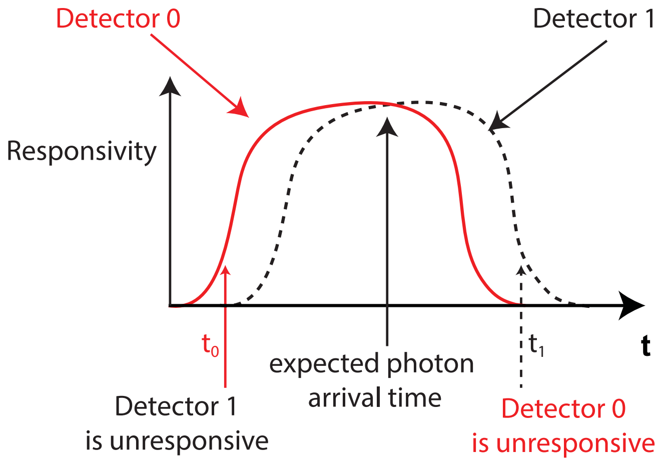

Detector 0
Detector 1
Responsivity
t
t_0
t_1
expected photon arrival time
Detector 1 is unresponsive
Detector 0 is unresponsive

Figure 2.2: Efficiency mismatch among two detectors. The solid (dashed) curve shows the responsivity of detector 0 (1) as a function of time. Even though the gate pulse is applied at the same time, the efficiency versus time curves are different for the two detectors. In practice, the photons are expected to arrive at a time when the responsivities are maximum and equal.

Most of the single photon avalanche photodiodes (APDs) used today are operated in gated-mode. This means they are normally reverse biased just below the breakdown voltage V_{br} and are not sensitive to incoming single photons. When photons are expected to arrive, a pulse is applied to the detectors that increases the (reverse) bias voltage above V_{br} and makes the APDs sensitive to incoming single photons. This mode of operation is called 'Geiger mode'. Thus, APDs are sensitive to single photons during the time window when gate pulses are applied and are practically insensitive to photons outside the gate-pulse window. The sensitivity of the gate window of two detectors are shown in fig. 2.2.

Normally, the timing is synchronized in such a way that incoming photons hit the detector at the middle of the gate-window where the sensitivity of the two detectors are maximum and approximately same. However, if Eve shifts the arrival time of the photon at a time when one detector is relatively well responsive than the other, it is possible for Eve to control the detection outcome. For example, in fig. 2.2, at time t_0 detector 1 is almost unresponsive to incoming photons while detector 0 is relatively well responsive. Thus, if Eve sends a photon at this time, and if it goes towards detector 0, it has a much higher chance to be detected compared to that if it goes towards detector 1. The opposite case happens at time t_1. In this example, the efficiency mismatch in the time domain has been shown. However, it can also happen in the spatial domain or any other domain depending on the nature of the implementation. If efficiency mismatch exists in the implementation, Eve can perform a cleverly staged faked-state attack and compromise the security as pointed out in Ref. [79].

2.3.4 Time-shift attack

Time shift attack exploits the efficiency mismatch present in the detector assembly. It is based on the proposal from Ref. [70] and was experimentally demonstrated in [71] (which was the first quantum attack against a commercial QKD system). The basic principle of time shift attack can be described with the help of fig. 2.2 and is very similar to the attack described in section 2.3.3. Instead of performing a faked-state attack, Eve just randomly shifts the arrival time of the incoming photon to t_0 or (t_1). When Eve shifts the arrival time to t_0 and it results in a detection, it is highly likely that it was detected in detector 0 as detector 1 was relatively less responsive at that time. The amount of mismatch at the two time instants can be defined as [70],

$$\begin{aligned} r(t_0) &= \frac{\eta_0(t_0)}{\eta_1(t_0)} \\ r(t_1) &= \frac{\eta_1(t_1)}{\eta_0(t_1)} \end{aligned} \tag{2.9}$$

Here, $\eta_i(t)$ is the efficiency of detector i at time instant t. The amount of mismatch determines the amount of information leaked to Eve and based on the type of the protocol an analytical expression can be derived.

2.3.5 Photon number splitting attack

Single photon sources are difficult to implement in practice. Instead, weak coherent sources such as lasers are often used as the source of quantum communication. Let us consider, Alice is using weak coherent pulses (WCP) for encoding the qubits with a mean photon number μ. The photon number distribution at Alice is described by a Poisson distribution:

$$P_A(n|\mu) = \frac{(\mu)^n}{n!}\exp(-\mu) \tag{2.10}$$

Here, $P_A(n|\mu)$ is the probability that the laser emits n photons given the mean photon number is μ. In this circumstance, Eve can utilize a powerful attack scheme called the photon number splitting attack [30, 85]. The steps are as follows: [86]:

- **QND measurement** For each pulse flying towards Bob, Eve performs a quantum non demolition measurement (QND) to measure the number of photons contained in that pulse. This measurement does not change the signal state and neither induces any errors.
- **Multiphoton pulses** Whenever she finds a pulse contains $n \geq 2$ photons, she splits a photon to store it in her quantum memory, and let the other photons go towards Bob. This also does not disturb the signals neither induces any error.
- **Single photon pulses** Whenever she finds a pulse contains $n = 1$ photon, this photon is blocked.
- **Line loss compensation** As splitting photons out from the pulses is equivalent to loss, she compensates for by replacing the original line with a lossless line.
- **Measurement**Eve waits until Alice and Bob declare the basis for each pulse and measure the photons stored in her quantum memory in the same basis. In this way, Eve can have a copy of each of the multi photon pulse without introducing any error.

Note that, when the loss of the line is high enough that detection probability in Bob P_{det} is lower than the multi photon probability P_{multi}, PNS attack compromises the whole key without introducing any error However, when $P_{multi} < P_{det}$, she needs to pass some of the single photon pulses to Bob to match the rate and can also perform some optimal eavesdropping strategy that will optimize her knowledge [65]. However, if the photon number statistics is monitored at Bob, Eve runs the risk of disclosing her intervention. The expected photon number distribution at Bob is a Poisson distribution with a changed mean photon number (μt_{AB}):

$$P_{Bob}(n) = \frac{(\mu t_{AB}\eta_{Bob})^n}{n!} e^{-\mu t_{AB}\eta_{Bob}} \tag{2.11}$$

After Eve's intervention the actual photon number statistics is no longer Poisson and in fact becomes [86]:

$$P_{PNS}(n) = \begin{cases} (1+\lambda\mu)exp(-\mu t_{AB}), & \text{if} n = 0 \\ ((1-\lambda)\mu + \frac{\mu^2}{2})\exp(-\mu), & \text{if} n = 1 \\ \frac{\mu^{n+1}}{(n+1)!}\exp(-\mu), & \text{if} n > 1 \end{cases} \tag{2.12}$$

Clearly the two distributions in eq. (2.11) and eq. (2.12) are not same and if bob checks the coincidence probability at his end, Eve's stealth is disclosed. In order to stop this, Eve can perform the extended PNS attack [65] and can replicate the whole probability distribution expected by Bob in a practical parameter region.

2.3.6 Trojan-horse attack

The so called Trojan-horse attack [82] was first introduced as 'large pulse attack' in 2001 [81]). The basic principle of this attack can be understood from fig. 2.3 which is a reprint from Ref. [152]. In this attack, Eve tries to extract security critical information from a component inside Alice or Bob by sending in a bright pulse and then measuring a back-reflected part of it as shown in fig. 2.3. The thick arrow represents the bright Trojan-horse pulse from Eve. The wall represents lossy components (filter, isolator and attenuators) and might constitute other defensive structures to keep the system secure. The Trojan-horse pulse picks up the encoding information from the encoding device, then gets back-reflected into the channel containing the security critical encoding information. Then Eve measures them to extract the encoding information and compromise the security. This attack has already been experimentally demonstrated in Ref. [83, 153, 154]

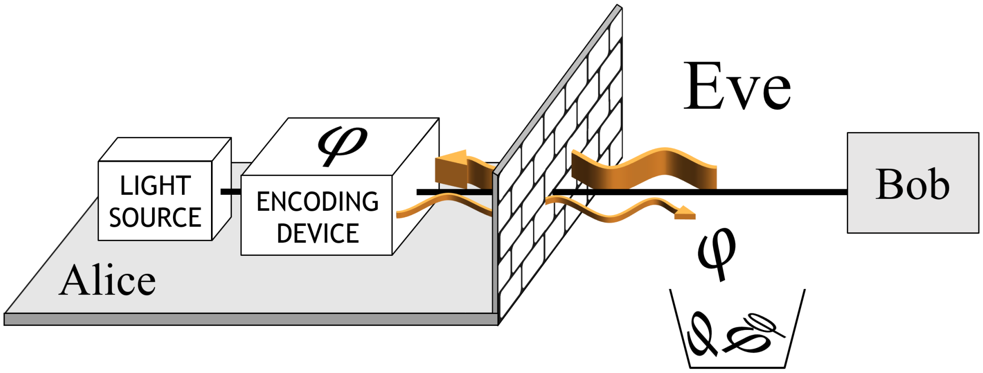


Figure 2.3: Trojan-horse attack against a practical Alice. The thick arrow from Eve to Alice represents the bright pulse from Eve. The wall represents the lossy mechanisms (isolator, attenuator, filter) in the system. Due to losses, an attenuated portion of the Trojan-horse pulse is reflected back to the channel from the encoding device, containing the encoding information. This is represented by the thin arrow from Alice to Eve. If Eve is able to measure this attenuated signal, she can have the secret encoding information and the security can be compromised. Figure reprinted from [152].

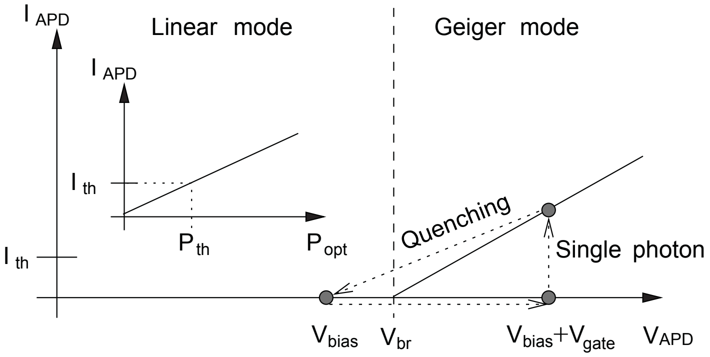


Figure 2.4: Linear mode and Geiger mode operation of an APD. In the Geiger mode, the APD is biased over the breakdown voltage V_{br} and are sensitive to single photons. After a single photon detection event the avalanche starts and a quenching mechanism is required to force it to stop. The quenching mechanism brings the bias voltage below breakdown, and the diode enters the linear mode of operation. At this mode, the diode is no longer sensitive to single photon and the generated photo current I_{APD} is proportional to the incident optical power P_{opt}. Figure reprinted from [73].

2.3.7 Detector control attack

Most single photon detectors used in QKD are avalanche photodiodes (APDs) that operate in the Geiger mode [155], *i.e.,* they are (reverse) biased above their breakdown voltage V_{br}. In this mode, an incident single photon triggers a large avalanche current I_{APD} that, when exceeds a threshold I_{th}, causes a detection event or 'click'. After a click, to stop the avalanche, a quenching mechanism in employed that brings the diode into linear mode of operation as shown in fig. 2.4. In this mode, the generated photo current I_{APD} is proportional to the incident optical power P_{opt}. It was shown in [73] that at the linear mode, if APDs are illuminated with continuous bright light, it is possible to force them stay in that mode. This is called 'detector blinding' and in this case, they are not sensitive to single photons anymore and only produce a click when the incident mean photon number μ goes above a certain threshold intensity μ_{th}. This behavior has been experimentally confirmed in many detectors types [73, 151, 156, 76, 75, 157, 158, 159]. After the detectors are blinded, Eve can perform a faked-state-attack. She measures each of Alice's signals in one of the two BB84 bases chosen randomly, and sends her measurement result to Bob with a mean photon number satisfying,

$$\frac{\mu}{2} < \mu_{th} < \mu. \tag{2.13}$$

As a result, whenever, Bob's basis matches that of Eve, all the photons end up in the single detector and $\mu > \mu_{th}$ ensures that a click is registered. On the other hand, if Bob's basis does not match with Eve's, the photons are divided equally into the two detectors and $\frac{\mu}{2} < \mu_{th}$ ensures none of the detectors register a click. This is shown in fig. 2.5. The net result is, whenever, Bob has registered a click, Eve knows that his measurement was the same as she sent at no extra error. Eve then listens to the public communication between Alice and Bob and performs the same error correction and privacy amplification procedures as them, to obtain an identical copy of their secret key [73]. This is the called the detector control attack.

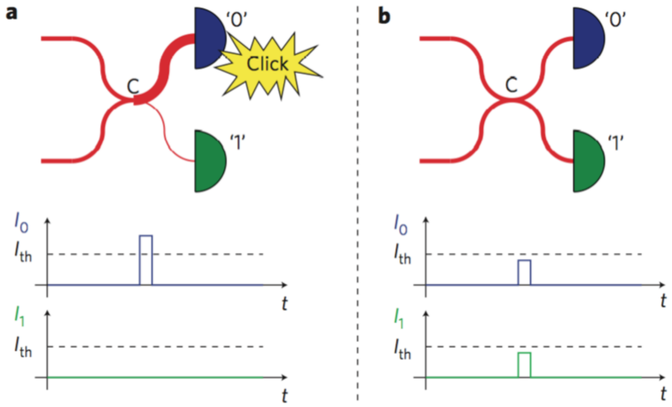


Figure 2.5: Detector control via blinding. If the measurement basis matches that of the incoming light, the deterministic detector receives all the light intensity $I > I_{th}$ and results in a detection event. However, if the bases do not match, light intensity is split between the two detectors. In this case, each of the detectors receives an intensity $I/2 < I_{th}$ and none of them results in a detection. This allows the eavesdropper to successfully carry out the faked-state attack. Figure reprinted from [73].

Chapter 3

Security evaluation of a pulse energy monitoring system

This chapter is based on a published paper [142].

3.1 Motivation

In quantum communication systems employing weak coherent pulses, the sender, Alice needs to choose an optimum value of her mean photon number μ to maximize the secure key rate [160, 65, 97, 161]. If the actual μ emitted by Alice is larger than this optimum value without her knowledge, it can lead to a side channel with severe security consequences. This can happen either because of an active manipulation by Eve, or because Alice underestimates μ owing to a finite precision of her calibration. Regardless of the reason, its effect is significant in any quantum communication scheme that uses weak coherent states and most severe in case of two-pass schemes.

Two-pass optical schemes have significant practical advantages and are widely used today, e.g., in plug-and-play quantum cryptography [124, 162, 163, 144], relativistic quantum cryptography [164], and coin-flipping [19]. In any two-pass scheme, bright light is generated at one party and sent towards the other. The other party makes the necessary encoding, attenuates the light and then reflects it back to the first party. The risk with this scheme is that the eavesdropper has access to the light in the channel during the first pass and she can add extra photons called Trojan-horse photons with this light. When this light is reflected back, the Trojan-horse photons carry the same information as the actual photons and the eavesdropper can split it off from the line and get the key information. This attack is called Trojan-horse attack (THA) [81, 82].

Hence, in a two pass scheme, it is necessary to monitor the incoming light during the first pass to prevent the THA. Also, in other schemes employing week coherent sources, monitoring is required to ensure that mean photon number going out or coming in is within the allowable bound. However, implementation of the monitoring detector has largely been ignored in experimental realizations so far. The first implementation has been done in ID Quantique's commercial QKD device Clavis2 [96]. In this chapter we use this particular system as a prototype and evaluate the security.

Since the system under consideration is a plug-and-play system, we first describe in section 3.2 the principle of operation of the plug-and-play system and the pulse energy monitoring system (PEMS) employed in Clavis2. In section 3.4 we identify three flaws in the design of the PEMS and show experimentally that each of these flaws can be exploited to inject extra photons into the system without triggering an alarm. We model two attacks in section 3.5 and simulate their performance for both BB84 and SARG protocols for a range of system parameters in section 3.6. Our results confirm that even a practical attack implementable today would breach security of this implementation. In section 3.7, we discuss the applicability of our attacks to the case of practical quantum coin-tossing. We discuss how to redesign the pulse-energy-monitoring system correctly in section 3.8, and conclude in section 3.9.

3.2 Theory

3.2.1 Operation of a plug-and-play system

Most fiber-based implementations of quantum key distribution (QKD) systems encode a qubit value by either different photon polarizations (H, V, D or A) or by different phases ($0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$). However, keeping the polarization stable over a long distance in fiber is difficult due to fiber's birefringence property that changes the polarization states of the light going through it. As a solution to this, a phase-based plug-and-play QKD system was proposed in [165]. Since, in our subsequent works, we focus on the commercial plug-and-play system Clavis2 [166], we summarize its operation and major points of implementation in the following.

The schematic of the commercial plug-and-play system Clavis2 is shown in fig. 3.1. Pulses originate in Bob's laser at a fixed frequency of 5 MHz. They pass through an unbal-

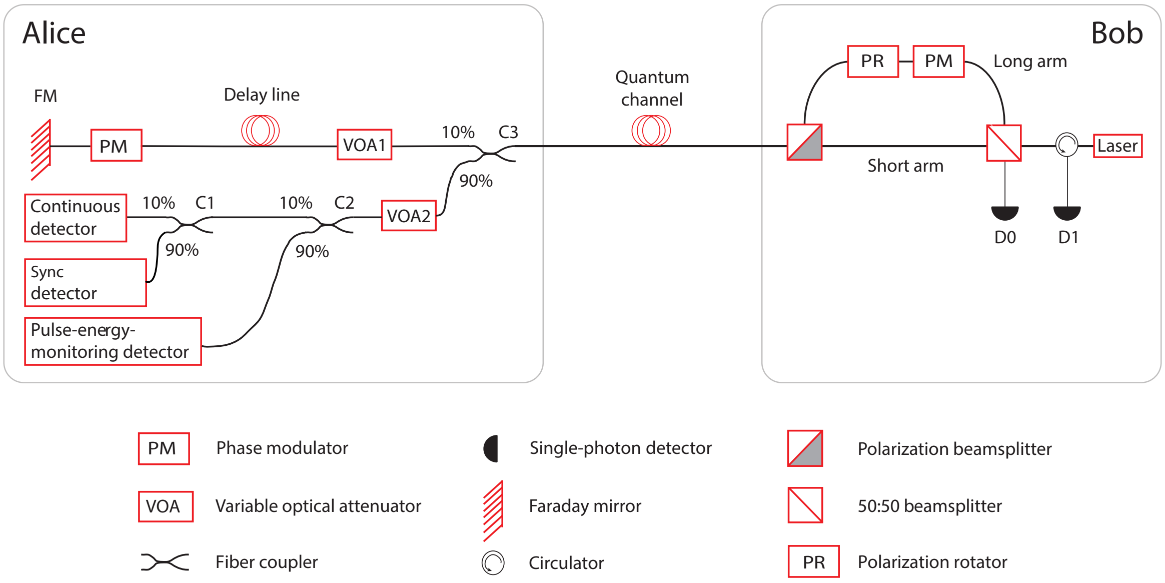


Figure 3.1: Plug-and-play system, as implemented in Clavis2. Figure reprinted from [142].

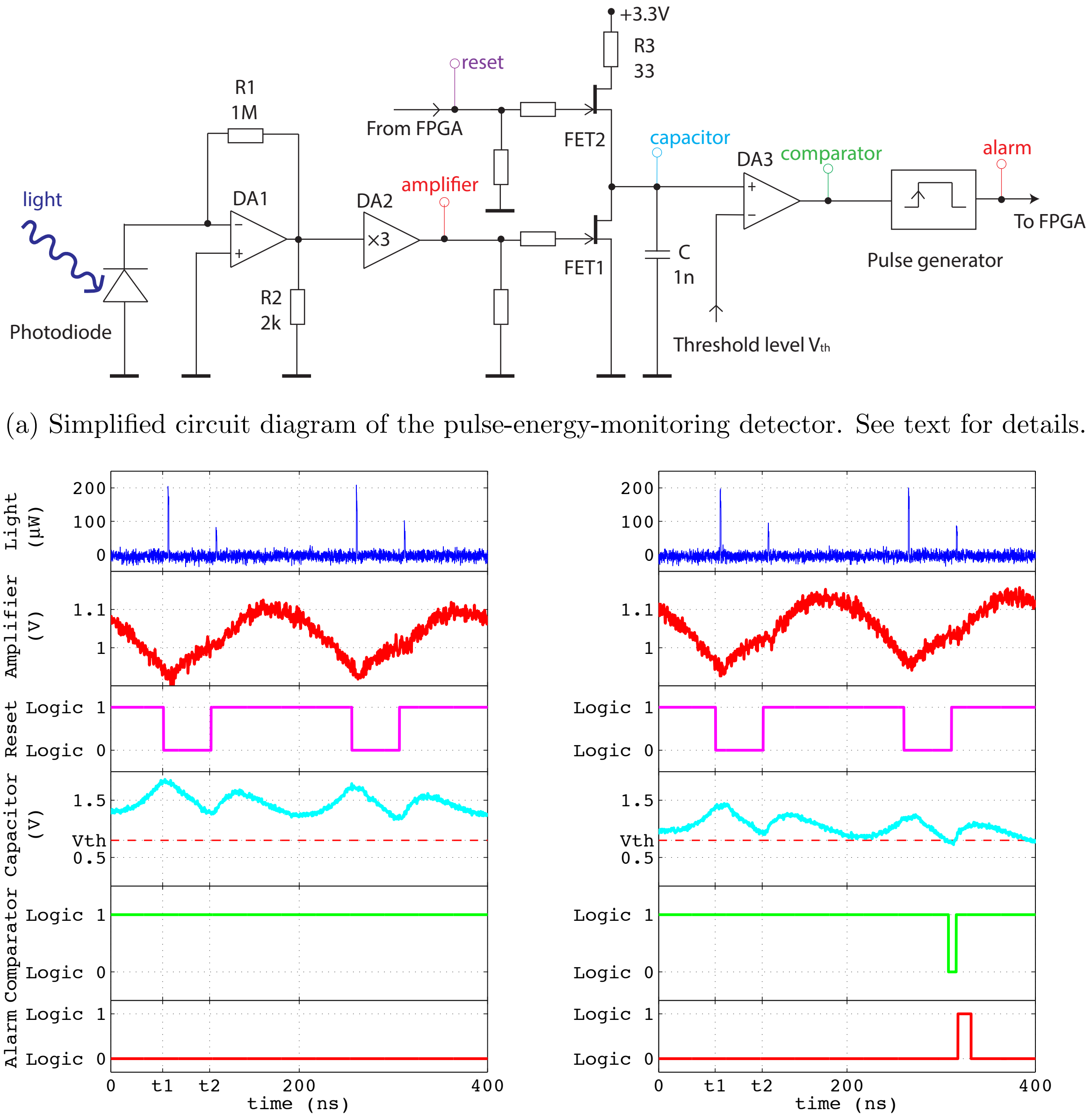
anced Mach-Zehnder interferometer (MZI) where one arm is intentionally made 1 m longer than the other arm. The longer arm has a polarization rotator to rotate the polarization by 90 °. It also contains a phase modulator that chooses the measurement basis for Bob. The phase modulator is off during the first pass when light is traveling from Bob to Alice. The pulses from the two arms are combined by the polarization beam splitter (PBS) and goes out into the quantum channel. Thus, for each laser pulse generated by Bob, there are two orthogonally polarized pulses in the quantum channel going towards Alice with a fixed delay of 50 ns between them corresponding to the arm length difference of 10 m in the MZI. The first pulse is traditionally called 'reference pulse', while the second is called 'signal pulse'. The signal pulse has lower energy than the reference pulse since it passes through the longer arm containing the phase modulator and suffers additional loss.

Alice's attenuator VOA1 attenuates the signals, her phase modulator (PM) applies random phase ϕ_A $(0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$ on the signal pulse, and the Faraday mirror (FM) reflects and rotates the polarization orthogonally for both pulses. Thus, when the two pulses go out of Alice towards Bob, the signal pulse, attenuated to single photon level, contains the encoded bit value from Alice. After the two pulses arrive at Bob, they take the opposite arms of the MZI than the ones they took during first pass. The PM in the long arm is now 'on' and applies a random phase ϕ_B (either 0 *or* $\frac{\pi}{2}$) to the reference pulse. As a result of the combination of FM and unbalanced MZI, the two pulses have the same polarization, path difference and arrive at Bob's 50:50 beam splitter (BS) at the same time. Hence, the choice of the output BS path depends only on their relative phase difference ($\phi = \phi_A - \phi_B$). Two detectors D_0, D_1 and a circulator are used in the configuration shown in fig. 3.1 to collect the light after the BS. If $\phi = 0$ ($\phi = \pi$), the pulses emerge at the same (different) path from which they came, and are collected by D_1 (D_0). This is a measurement in the compatible basis. However, if Alice and Bob choose different bases (such that $\phi = \frac{\pi}{2}$ or $\frac{3\pi}{2}$), then the photons are split with equal probability between D_0 and D_1. These are measurements in the incompatible basis and these results are discarded during sifting. In this way, the raw key exchange is performed in Clavis2 system.

3.2.2 Operation of the pulse-energy-monitoring system

The function of the pulse-energy-monitoring system (PEMS) is to integrate the incoming pulse energy and trigger an alarm when the energy exceeds a predefined threshold value. A simplified diagram of the PEMS employed in Clavis2 is shown in fig. 3.2. A fiber-pigtailed p-i-n photo diode (JDSU EPM 605LL) is used to detect the light. Its photo current is processed by an electronic circuit shown in fig. 3.2a. Signals at six test points marked in the circuit are shown in fig. 3.2b and fig. 3.2c. At the front-end of the circuit there is a two-stage transimpedance amplifier, converting photo current into voltage signal. Owing to insufficient bandwidth of the amplifier first stage (opamp DA1; Texas Instruments OPA380), it outputs slow-rising electrical pulses that extend to the next few bit slots and interfere with the signals from those slots. The amplifier's second stage is a wideband current-feedback opamp DA2 (Analog Devices AD8009) that does not further distort the signal. Its output acts as a gate pulse for an N-channel field-effect transistor FET1 that is a part of an integrator circuit.

In theory, the operation of the integrator circuit should be the following. The gate pulse for FET2 (reset signal) is applied by the field-programmable-gate-array (FPGA) system controller. This reset signal is normally high, keeping FET2 in a conductive state such that current flows through it to charge an integrating capacitor C. At time t_1, the reset signal switches FET2 into high-impedance state for 50 ns, and the capacitor starts to discharge through FET1 (see capacitor signal). The amount of discharge is proportional to the power of the incoming light. At time t_2, reset signal switches FET2 into conductive state again and stops the discharging. This happens for each bit slot, and a negative spike proportional to the incoming light energy is generated at the capacitor. The negative spike is compared to a predefined threshold level V_{th}, whose value is calibrated at the factory in such a way that during normal operation, the negative spike amplitude is very close but almost never



(a) Simplified circuit diagram of the pulse-energy-monitoring detector. See text for details.

(b) Signals during normal operations.

(c) Generation of an alert signal

Figure 3.2: Pulse-energy-monitoring circuit and oscillograms. The six test points are marked 'light', 'amplifier', 'reset', 'capacitor', 'comparator' and 'alarm' in (a), and the oscillograms at these points are shown (b) for normal operation and (c) for the case when light power is increased by 0.1 dB (i.e., by ≈ 2%) above normal operation. During normal operation, when light pulses arrive with expected energy, the capacitor voltage always stays over the threshold level V_{th}. However, when the pulse's energy is higher than expected, due to higher gate pulse to FET1, deeper discharge of the integrating capacitor results. This causes its voltage go below V_{th}, which in turn creates an alarm. Figure reprinted from [142].

goes below V_{th}. However, when there is an extra light, this negative voltage spike crosses the threshold causing the output of comparator DA3 to go low triggering an alarm signal.

In actual operation of the practical implementation, when the reset signal from the FPGA goes back in the normal high state while the amplifier output is also high, both FET1 and FET2 are in the conductive state simultaneously. As a result, current from the +3.3 V supply flows through both the FETs into the ground and the integrating capacitor cannot be charged instantly by the positive reset signal. This produces the capacitor signal seen on the oscillogram that does not quite match the expected ideal circuit behavior. Nevertheless, the capacitor signal's lowest level during the cycle strongly depends on the light energy, allowing the circuit to detect a small excessive amount of light in a single pulse when tested to ID Quantique's specification.

The comparator signal is fed to a pulse generator that produces fixed-width pulse on the low-to-high logic level transition. This is the alarm signal fed to the FPGA that indicates the excess of incoming light. The system software discards all detections in the frame if one or more pulses inside the frame have triggered alarm in Alice. Thus any attempt by Eve to inject brighter pulses in a frame should lead to that frame being dropped from QKD.

3.3 Requirements for successful eavesdropping

We have chosen the following criteria for Eve that she must satisfy in order to remain inconspicuous while performing the attack:

1. **Maintain synchronization:** The synchronization between Alice and Bob must not break.

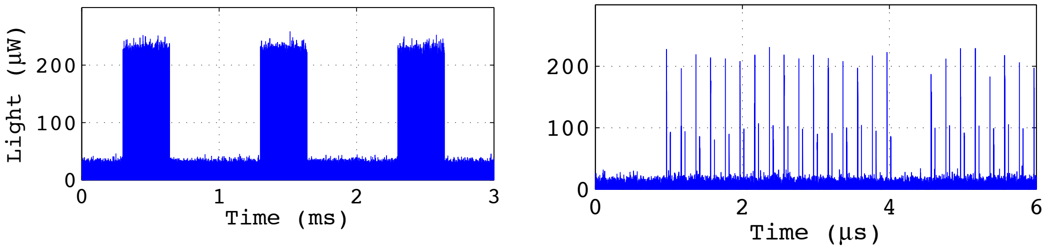
2. **Inject Trojan photons without alarm:** The Trojan photons must not trigger the alarm.

3. **Maintain same detection rate:** Detection rate at Bob must not be altered.

Next we discuss how each of the issues can be maintained during the successful attack.

3.3.1 Maintaining synchronization

In Clavis2 system, the synchronization of Alice's clock to Bob's clock is maintained by the sync detector (see fig. 3.1). Pulses are sent from Bob in packets called *frames* generated



(a)

(b)

Figure 3.3: Optical pulses coming from Bob to Alice. (a) Trains of pulses (frames) generated by Bob. The frames are generated every 1 ms, are 340 μs long, and contain 1700 pulse pairs. (b) Beginning of the frame showing a synchronization pattern. The synchronization circuit checks for this specific pattern in every frame. Figure reprinted from [142].

every 1 ms as shown in fig. 3.3a. Each frame is 340 μs long and contains 1700 pulse pairs with 200 ns period. Each of the 200 ns intervals containing one of these pulse pairs is called a *slot*. Only the first 20 slots of each frame are used for the synchronization of Alice's clock, i.e., the timing of Alice's modulator to Bob's laser modulator and detector. The synchronization pulses are shown in fig. 3.3b where at the beginning of each frame, Bob first sends 16 pulses, then skips the next two (17^{th} and 18^{th}) intentionally, and then sends the rest of the pulses of the frame. Alice's synchronization detector checks for this pattern in the first 20 slots with an avalanche photo diode receiver (Fujitsu FRM5W232BS). Upon detection of the correct pattern, Alice's electronics clock is synchronized to the frame and this is done separately for each frame coming from Bob. During the attack, we placed Eve in the quantum channel and generated synchronization pulses in the same pattern as expected by Alice to maintain the synchronization. In this way, synchronization was maintained.

3.3.2 Injecting Trojan photons

As mentioned in section 3.2.2, for each laser pulse generated by Bob, there are two signals in quantum channel as shown in fig. 3.4. For the system under test, for a 1 m fiber line between Alice and Bob, the energies of the calibrated reference and signal pulses are 150 fJ and 73 fJ respectively at the output of Bob. Energy values in the rest of this work are measured at the same point. We remark that as Alice applies her random encoding only at the (second/trailing) signal pulse, only this pulse contains the quantum information.

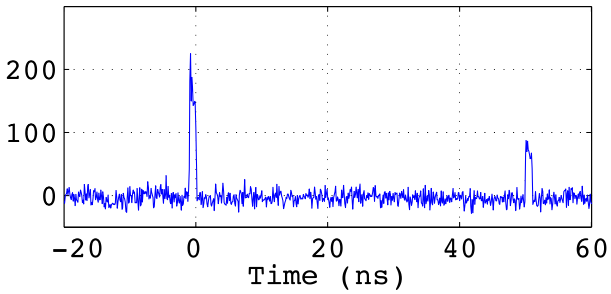


Figure 3.4: Pulse pair per slot in the optical link. The energy of the first/leading reference pulse is 150 fJ and the energy of the second/trailing signal pulse is 73 fJ. Figure reprinted from [142].

Hence, it is called 'signal pulse' while the first/leading pulse is called 'reference pulse'. In order to do a successful Trojan-horse attack, the Trojan photons need to be injected in a time overlapping with the signal pulse, so that they can come out of Alice with the desired encoding information. In section 3.4, we experimentally demonstrate three different strategies, that achieve this goal by experimentally exploiting imperfections present in the design of the PEMS.

3.3.3 Maintaining detection rate

During the attack, Eve should not change the probability of detection at the receiver; otherwise, the receiver can easily monitor the detection rate and discover Eve's presence. After experimentally demonstrating the feasibility of injecting Trojan photons into the system, we also modeled two different attacks to study the effect of increased mean photon number on security. In each of these models, we assumed that the detection rate at the receiver with and without the attack are same. We describe them in detail in section 3.5.

3.4 Experimental demonstration of attacks

In this section, we experimentally demonstrate three attack strategies to inject extra light into the system at desired time slot without triggering any alarm. Each of the strategies exploits imperfections of the PEMS. We named the first 'bandwidth attack' that exploited the insufficient electrical bandwidth of the front end amplifier. The second attack was 'saturation attack' that took advantage of the reverse saturation and long recovery time of the same amplifier. The third attack was 'edge-trigger attack' that exploited a design flaw of the electronics. They are discussed in more detail next:

3.4.1 Bandwidth attack

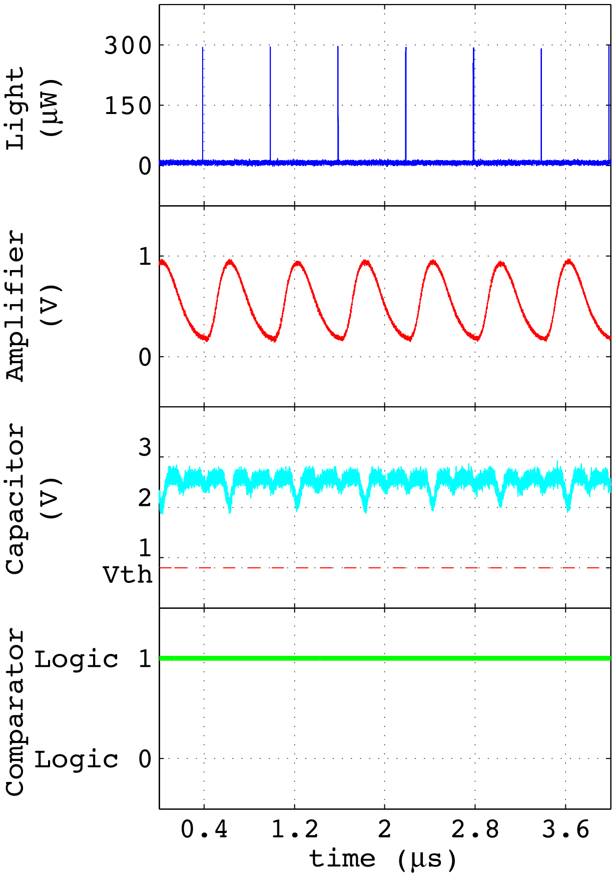
Ideally, the output of the front-end amplifier corresponding to the reference and signal pulse should be two non-overlapping electrical signals separated by 50 ns. But the 3 dB bandwidth of this amplifier in the current configuration was about 1 MHz [167]. This caused it to output slowly rising electrical pulses (as mentioned in section 3.2.2) that overlapped with the next signal or slot.

Since only the 'signal pulse' is encoded by Alice, only its photon number μ is significant for the security [1]. However, because the two pulses largely overlap at the amplifier output, the electronics is mainly monitoring the total energy of the pulse pair and not the 'signal pulse' only. Thus, we began by redistributing energy between the signal and reference pulses (shown in fig. 3.4) by suppressing the reference pulse and making the signal pulse proportionally brighter.

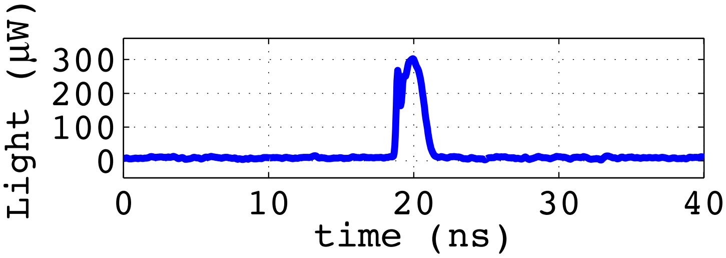
In this way, without triggering an alarm, we were able to increase the energy of the signal pulse over the calibrated value by a multiplication factor of $x = 3.1$. The value of $x = 3.1$ means that our injected photon number was 3.1 times higher than the calibrated value. This breaks the security in theory, but is only sufficient for a partial information leak of 49% with BB84 protocol (80% with SARG04 protocol) when using an attack implementable with today's technology (analyzed in section 3.6). To increase x further, we started to suppress additional pulses as described next.

For every four pulses, we suppressed the first three and injected at the fourth slot a bright pulse which we call the 'probe pulse' (see fig. 3.5). Due to the three blocked pulses, the voltage level at the output of the front-end amplifier is lower than normal (compare fig. 3.5a with fig. 3.2b). As a result, when the brighter probe pulse arrives at the fourth slot, it does not increase the voltage enough to trigger the alarm. In this case, we were able to inject a probe pulse with a maximum energy of 623 fJ (shown in fig. 3.5b) without triggering and alarm. This corresponded to a multiplication factor of approximately $x = 8.5$. We also experimentally performed blocking two out of three and one out of two pulses and were able to inject a probe pulse with $x = 7.3$ and $x = 5.4$. Higher values of x could be achieved by blocking more than three pulses, but in that case the negative

[1] This has been shown to be incorrect [168], however the current Clavis2 software assumes μ is the mean photon number of the second pulse, disregards the mean photon number of the first pulse, and performs QKD according to these assumptions.



(a) Suppression of three pulses out of four and the corresponding effect on the amplifier, capacitor and comparator output. The pulse energy has been increased 8.5 times from 73 fJ to 623 fJ.



(b) Pulse shape carrying the maximum injected energy using this method (623 fJ).

Figure 3.5: Exploiting the low bandwidth of the front-end amplifier. Figure reprinted from [142].

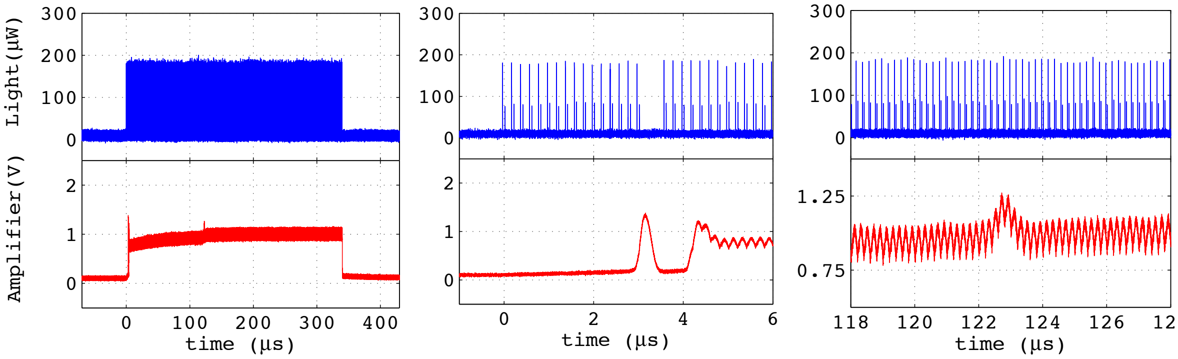


Figure 3.6: Recovery of the front-end amplifier from the negative saturation to the normal operation. (a) Entire 340 µs long frame. A minor peak is visible in the amplifier output at ∼ 123 µs, marking the recovery of the amplifier from the negative saturation. (b) Initial part of the recovery from the negative saturation. Even though light pulses are arriving at the input of the amplifier, no output is produced for ∼ 3 µs. (c) A transient at ∼ 123 µs is the last irregularity, after which the amplifier fully recovers from the saturation. Figure reprinted from [142].

saturation of the amplifier became the dominant factor, as discussed and generalized in section 3.4.2.

3.4.2 Saturation attack

Although the first 20 slots in each frame bear the synchronization pattern, data-carrying pulses (which we will henceforth call *data pulses*) start from slot 701 (140 µs) and continue to the last slot 1700 (340 µs) of the frame. The slots 21 to 700 are kept idle, which is a work-around for an engineering mistake: The output of opamp DA1 enters negative saturation when there is no light coming in [167] at the end of the frames. Once pulses start to appear at the beginning of the next frame, recovery from this saturation state takes a relatively long time, approximately 123 µs or 615 slots, with a bump at the end of the recovery (see fig. 3.6). Hence, no data are sent during this unstable time. Pulse energy is only monitored during the data pulses (slots 701–1700) [2].

[2] The frame structure described is taken from ID Quantique's factory calibration utility for their commercial encryption products. The current version of QKD software distributed with the research system Clavis2 (as of December 2014) does not perform pulse energy monitoring, and has no idle pulses.

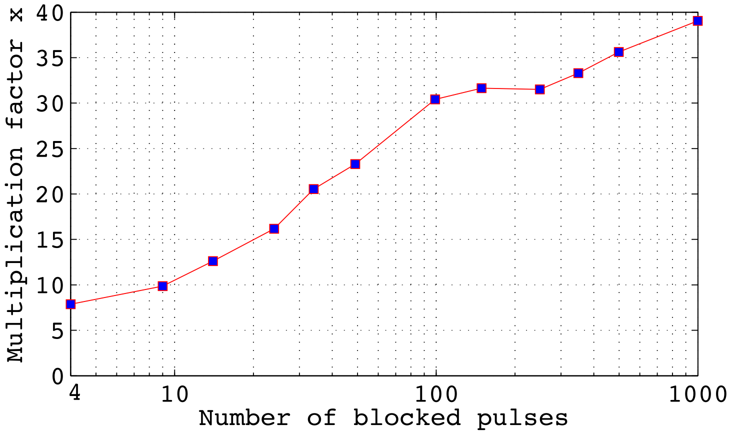


Figure 3.7: Multiplication factor x for $(n + 1)$st pulse vs. number of blocked pulses n. Figure reprinted from [142].

To perform the attack, we removed all the pulses starting from 100 µs till the start of the monitoring period (140 µs), which forced the amplifier to re-enter the negative saturation. Then, starting at 140 µs, for every $n + 1$ pulses, we blocked the first n pulses and sent a bright probe pulse at $(n + 1)$st slot. We continued to increase the energy of this probe pulse until an alarm was generated. The multiplication factor achieved versus n is plotted in fig. 3.7. We see that the curve rises steeply for up to 100 pulses blocked, then starts to saturate. By blocking 250 pulses, Eve can achieve multiplication factor $x = 31.5$, while by blocking 100 pulses she can have $x = 30.4$. Thus, to avoid a reduction of the key rate under attack, it is likely more efficient to block 100 or fewer pulses. As an example, we show the 100 pulse blocking case in fig. 3.8. The signal at the amplifier output became smaller as we went further into the frame, vanishing in the last part of it. This is because the longer the amplifier stayed into saturation, the more energy it needed to recover. While we have entered 9 probe pulses each with 2220 fJ energy $(x = 30.4)$, no alarm was generated during the 140–340 µs monitoring period.

3.4.3 Edge-trigger attack

As mentioned in section 3.2.2, the output from the comparator is applied to a pulse generator that produces a fixed-width alarm pulse on the low-to-high transition of its input. In addition, the integrator is unable to reset the capacitor voltage if the amplifier output is high. These particular design choices pose the biggest loophole in the system, which we have confirmed experimentally. Before the start of the monitoring period, at around

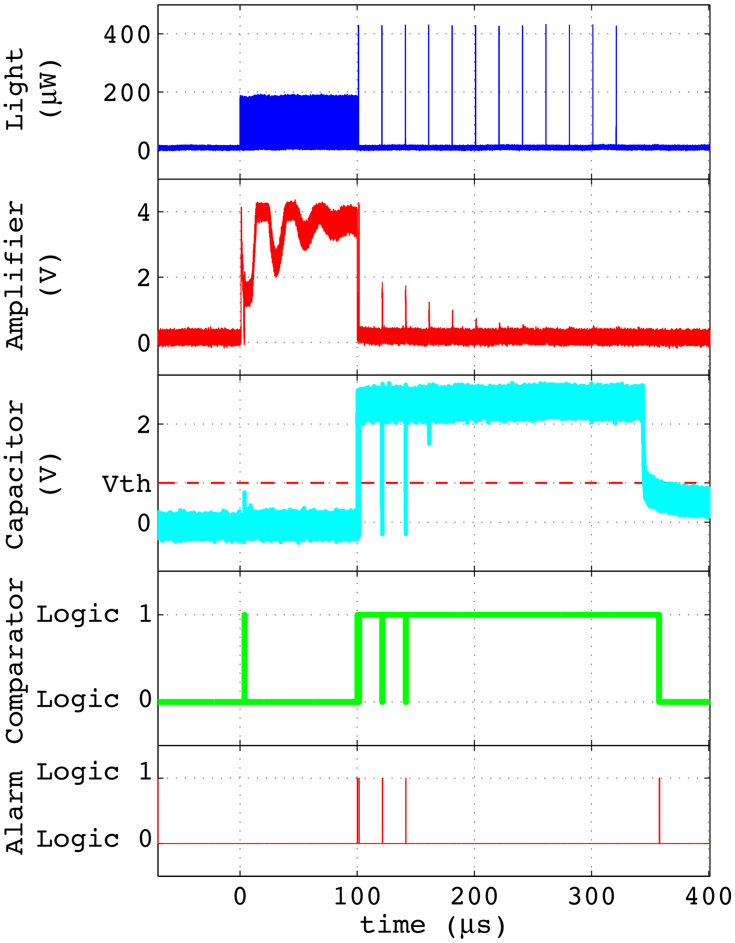
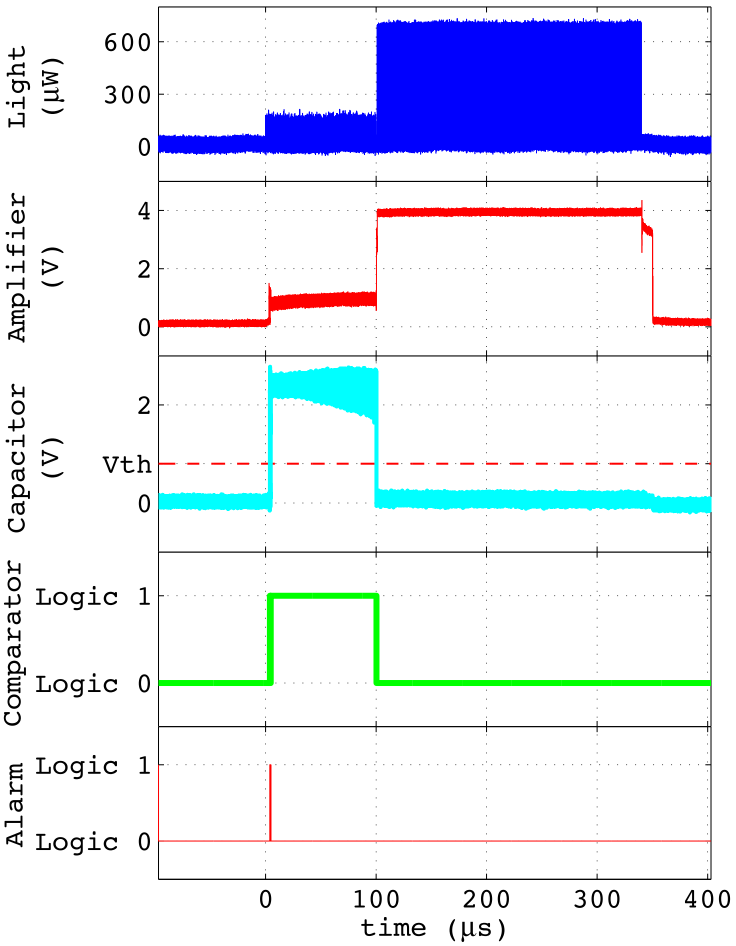
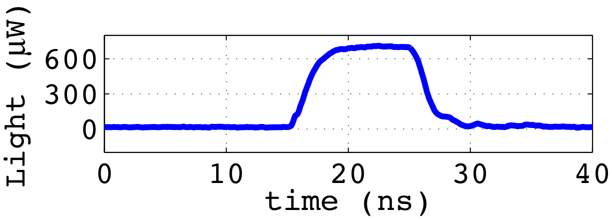


Figure 3.8: Attack exploiting the saturation effect of the front-end amplifier, blocking 100 pulses. The further into the frame the probe pulses are injected, the smaller the amplifier output becomes, because the amplifier stays into saturation for a longer period and more energy is required to bring it out of it. In the alarm plot, the first three pulses occurred because the energy of the probe (light) pulses was enough to produce an amplifier output strong enough to result in an alarm (as it has not yet been into a deep saturation). However, they occurred before the monitoring period and were not counted as an alarm signal by the FPGA. Similarly, the last pulse in the alarm plot occurred when the integrator was reset after the frame (after the end of monitoring period) and was not counted as an alarm. Figure reprinted from [142].



(a) Bright pulse injection and its effect on the circuit. At the end of the frame when the amplifier output became zero, the capacitor voltage was still low. The reason is because after the end of the frame, the FPGA no longer generated the reset signal and hence the integrator did not reset. It reset at the beginning of the next frame after the reset signal was produced.



(b) Pulse shape carrying the maximum injected energy using this method (7150 fJ or 97 times more than the calibrated signal pulse).

Figure 3.9: Exploiting edge-triggered alarm monitoring. Figure reprinted from [142].

100 µs, we started injecting bright probe pulses at each slot in order to push the capacitor voltage completely below the threshold (see fig. 3.9). As long as the bright pulses were sent (in our case until the end of the frame), the comparator output remained low and there was no low-to-high transition for the pulse generator to produce the alarm. After the end of the frame, when we stopped sending the bright pulses, the amplifier output went low as seen from fig. 3.9a but the capacitor voltage was still below the threshold as there was no reset signal to reset the integrator at the end of the frame. Using this method, we were able to inject probe pulses with a maximum energy of 7150 fJ (limited by our available source power) corresponding to a multiplication factor $x = 97$ (fig. 3.9b). Note that the attack takes place in every bit slot, and no pulses needed to be blocked. Intuitively, at such a high μ this attack shifts Alice's operation close to a classical regime, and no security can be maintained.

3.5 Modeling of attack

The calculation of secret key rate in CLavis2 is based on the following assumptions:

1. Eve has no control over Bob's detectors, i.e., she cannot change Bob's detector efficiency and detection probability.
2. Bob expects a certain count rate and Eve should not change it.
3. Eve performs individual attacks. More specifically, her attack is a combination of photon-number-splitting (PNS) (see section 2.3.5) and cloning (see [169]) attack.
4. Events containing more than two photons are ignored since they occur too infrequently to make a significant contribution.

Considering the above assumptions, the mutual information between Alice and Eve becomes [170]

$$I_{A:E} = \frac{1}{2}\mu\eta(tt_b - \frac{\mu}{2})I_1(D_1) + \frac{1}{2}\mu\eta\frac{\mu}{2}, \tag{3.1}$$

where μ is the average photon number per pulse set by Alice, η is Bob's detector efficiency, t is the measured channel transmission efficiency, t_b is the transmission in Bob's interferometer, and $I_1(D_1)$ is the information gathered by Eve when she performed cloning attack that introduces a disturbance D_1 on the state. The first term in the equation comes from the cloning attack, where Eve obtains partial information, and the second term comes from the PNS attack which gives Eve full information. The mutual information between Alice and Bob $I_{A:B}$ is defined as [170]

$$I_{A:B} = \frac{1}{2}[\mu t t_b \eta + 2p_d][1 - f_{ec}H(Q)]. \tag{3.2}$$

Here, p_d is Bob's detector dark count probability, f_{ec} is the error correction efficiency, H is the binary entropy function, and Q is the measured QBER. The term $f_{ec}H(Q)$ accounts for the information revealed during error correction, which must be discarded. The final key is calculated as:

$$S = I_{A:B} - I_{A:E}, \tag{3.3}$$

Our two modeled attacks follow the first three assumptions but includes multi-photon events (with more than two photons) which become significant as μ is increased. The attacks also assume that Bob does not monitor double clicks, and instead implements the squashing model [171] (implemented by ID Quantique in a recent software update to Clavis2), where double clicks are assigned a random bit value, therefore contributing to an average 50% QBER. We named our two modeled attacks as 'strong attack' and 'realistic attack' that are described in more details next.

3.5.1 Strong attack

In this attack, we assume that Eve is limited by the laws of quantum mechanics only. We assume she is doing a combination of PNS and cloning attacks [170] (as assumed in the key-rate calculation of Clavis2 system). However, in this case μ is being manipulated by a multiplication factor of x and at higher values of x, the contribution of multi photon pulses are no longer negligible (as assumed in the key-rate calculation of Clavis2). So, actual information available to Eve is higher than what was calculated in the privacy amplification step and in this way Eve can extract extra information about the keys. The actual mutual information between Alice and Eve then becomes:

$$I'_{A:E} = R_1 I_1(D_1) + R_{multi}. \tag{3.4}$$

Here R_1 (R_{multi}) is the contribution to Bob's detection rate from the single-photon (multi-photon) pulses where Eve implements the cloning (PNS) attack.

$$R_1 = \frac{1}{2} p^1_{attack} \eta x \mu e^{-x\mu} \tag{3.5}$$

$$R_{multi} = \frac{1}{2}\sum_{n=2}^{\infty} p^n_{attack}[1-(1-\eta)^{n-1}]\frac{(x\mu)^n}{n!}e^{-x\mu}, \tag{3.6}$$

with p^n_{attack} the probability of performing the attack on the n-photon pulse. When Eve does not attack a pulse, this pulse is blocked and does not contribute to Bob's detection rates. To ensure that expected count rate at Bob remains unchanged, the rates must follow

$$R_1 + R_{multi} = \frac{1}{2}[1-\sum_{n=1}^{\infty}((1-tt_b\eta)^n)\frac{\mu^n}{n!}e^{-\mu}]. \tag{3.7}$$

Here, the right-hand side is the detection rate when there is no attack. When x is small, p^n_{attack} is always 1 for $n \geq 2$. As x increases, the probability of cloning attacks (p^1_{attack}) decreases. If x is large enough for eq. (3.7) to be satisfied with $p^1_{attack} = 0$, Eve stops performing cloning attacks and begins blocking the pulses with lower photon number to satisfy eq. (3.7), i.e., first p^2_{attack} is reduced, then p^3_{attack} and so on until the equation is satisfied. In this way, Eve can maintain the detection rate at Bob.

3.5.2 Realistic attack

Our modeled realistic attack assumes that Eve is limited by present day technologies. In a realistic attack, Eve cannot alter the transmission of the channel, the alignment of the system or characteristics of Bob's detectors. In addition, she must use realistic beam splitters and optical switches that have non-zero insertion loss.

Eve's realistic attack strategy is to implement an unambiguous state discrimination (USD) attack [87] with a certain probability p^{USD}_{attack} while doing nothing with a probability $(1-p^{USD}_{attack})$.[3] Eve's measurement apparatus, shown in fig. 3.10, consists of a 50:50 beam splitter followed by two receiver units Bob$'$ (one for each measurement basis) with two detectors each. We assume Eve is placed immediately outside Alice's system (before any transmission losses in the fiber) as this gives Eve the maximum detection probabilities. Eve also has a source Alice$'$, placed just before Bob. This source emits attenuated-laser quantum states with an average photon number μ_e. Using this source, Eve sends a pulse whenever her detections allow her to unambiguously discriminate the state (i.e., when she measures photons in three different detectors, ensuring the correct state is the one

[3] We also analyzed the beam-splitting attack strategy [30, 88], but it performed significantly worse than the USD attack. Hence we only present here the results from the USD attack. In addition, the USD attack has the advantage of producing no extra errors (which could be monitored and used to detect Eve).

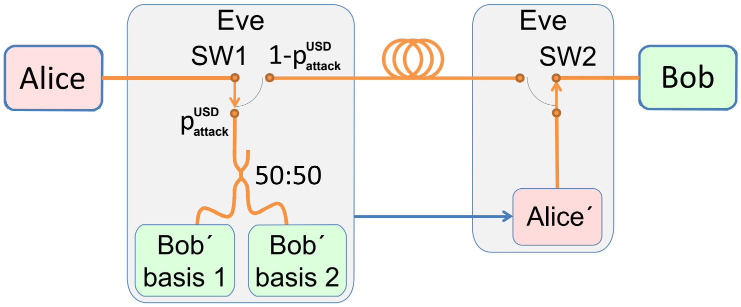


Figure 3.10: Realistic attack scheme. With a probability p^{USD}_{attack}, pulses from Alice are measured by Eve using a 50:50 beam splitter followed by two copies of Bob's setup Bob′ that use different measurement bases. When the USD measurement is successful, Eve sends a pulse in the measured state using a source Alice′ placed next to Bob. SW1 and SW2 are optical switches. SW2 can in practice be replaced by an asymmetric beam splitter. Figure reprinted from [142].

measured in the basis with only one detector click). When the state discrimination is ambiguous (measurement in only one or two detectors), she sends nothing. We assume that Eve's alignment is as good as Alice's and Bob's (same fringe visibility V), giving Eve's QBER [170]

$$Q_e = \frac{1}{2}\left(1 - \frac{V}{1 + 4p_e/(\mu t_s t_{BS} \eta_e)}\right), \tag{3.8}$$

where t_{BS} (t_s) is the insertion loss of Eve's imperfect beam splitter (optical switch), η_e is the total detection efficiency of Bob′ (including its internal losses), and p_e is the detector dark count probability in Bob′. The mutual information between Alice and Eve is then:

$$I''_{A:E} = R_{USD}(1 - H(Q_e)), \tag{3.9}$$

where R_{USD} is the contribution to Bob's detection rate when Eve successfully performs the USD attack. The rate is given by the probability that Eve's measurement is unambiguous multiplied by the probability that Bob registers a measurement in the right basis:

$$R_{USD} = p_{USD}\frac{1}{2}(2p_d + 1 - e^{-\mu_e t_s t_b \eta}), \tag{3.10}$$

where p_{USD} is the probability of an unambiguous measurement by Eve given by the probability of three-detector click:

$$p_{USD} = (1 - e^{-x\mu t_{BS} t_s \eta_e/2})(1 - e^{-x\mu t_{BS} t_s \eta_e/4})^2 \tag{3.11}$$

In order for Eve to not be detected, she must maintain the expected rate at Bob:

$$\begin{aligned} p_{attack}^{USD} R_{USD} + (1 - p_{attack}^{USD}) \frac{1}{2}(2p_d + 1 - e^{-x\mu t_s^2 t t_b \eta}) \\ = \frac{1}{2}(2p_d + 1 - e^{-\mu t t_b \eta}). \end{aligned} \tag{3.12}$$

As x increases, p_{attack}^{USD} will increase, allowing Eve to perform her attack more often. If x is large enough, Eve can perform the attack on every pulse ($p_{attack}^{USD} = 1$) without reducing the rate, giving her maximum information.

3.6 Simulation and results

We simulated our attacks using parameters extracted from experimental runs of the Clavis2 system. For several values of channel transmission t we extracted QBER Q, fringe visibility V, average photon number at Alice's output μ, Bob's detector efficiency η and dark count rate p_d (averaged between Bob's two detectors). We used the factory-calibrated value for Bob's interferometer short-arm transmission t_b. The number of data pulses sent by Alice was extrapolated based on the number of detections at Bob, t, μ, t_b, η and p_d, allowing us to ignore detector deadtime by giving us a number of pulses for which Bob's detectors were sensitive.

In the strong attack, Eve uses lossless lines, perfect efficiency detector with no dark counts and perfect alignment, and has access to perfect-efficiency quantum memory and the quantum non-demolition photon-number measurement. In the realistic attack, we assume commercially available fiber beam splitters that can achieve insertion loss as low as 0.3 dB [172] (in addition to splitting loss), and optical switches which can achieve insertion loss of < 1 dB [173, 174]. The best detectors that would currently be available for Eve are superconducting nanowire detectors, which are commercially available and have shown both high efficiency ($> 90\%$) and very low dark count rate ($< 100\,\mathrm{s}^{-1}$) [175, 176]. We assume the total detection efficiency of Bob$'$ $\eta_e = 80\%$, to further account for minor losses in his optical scheme. We measured the QBER of our Clavis2 system without Eve (for example, in BB84 at 3.4 dB line loss, it was 1.34%). In both of our attacks, this measured QBER is used as the minimum QBER for Bob. We allow Eve to increase the QBER in the strong attack to a maximum of 8%, which is near the limit where Clavis2 can (sometimes) extract secure key [84]. The realistic attack does not cause any increase in QBER because Eve will block all pulses where she does not unambiguously determine the state.

Of the three strategies to inject Trojan photons presented in this thesis, the first two require Eve to suppress a certain number of pulses. This limits the information that Eve can gather because she has to maintain the rate at Bob by decreasing the probability of her attack (p_{attack}^{USD}). In the third attack, Eve can increase the energy of all pulses, which allows her to get the most information. We used numerical simulation to compute the performance of the attacks.

3.6.1 Attack performance on BB84

The fraction of secret key leaked to Eve with the attacks is shown in fig. 3.11. In the bandwidth and saturation attacks, Eve must increase μ sufficiently to compensate for the suppressed pulses before the attack can be performed without her being notice. The bandwidth attack can increase μ by up to a factor $x = 7.3$ while suppressing two pulses, more than the required $x = 5$ to extract full information in the realistic attack model. The performance of the saturation attack is hindered by the large number of pulses suppressed. Nevertheless, the required $x = 6.2$ to extract full information in the realistic attack model can be achieved since suppressing four pulses allows $x = 7.87$. The edge-trigger attack, where no pulses are suppressed, allows Eve to extract information with a lower μ (starting at $x = 3$ in the realistic attack model), and can extract full information at $x \geq 3.2$. For strong attack model, full information can be extracted at $x = 1.2$, 2.1 and 2.8 for edge-trigger, bandwidth and saturation attacks respectively.

Figure 3.12 shows the dependence of minimum x on channel loss for both partial and full information leak in the edge-trigger attack. The value of μ depends on the channel loss ($\mu \approx t$ [170]), resulting in attack thresholds that only weakly depend on the channel loss, as seen in fig. 3.12. Note that commercial Clavis2 systems are only able to extract secure keys up to a certain line loss, limited by detector dark counts. BB84 protocol is more sensitive to loss than SARG04. Our system sample was able to produce secure key with BB84 at up to 6.7 dB line loss. Beyond this loss, BB84 was never able to extract secure key and thus there was no key information for Eve to gain.

3.6.2 Attack performance on SARG04

In the SARG04 protocol [34], keys are encoded in the basis instead of in the state. This lowers the sifting factor to 1/4 (from BB84's 1/2) but makes the protocol more robust to PNS attacks. To properly identify the encoded bit, Eve's measurement must return the same outcome as Bob's measurement. Each photon measured by Eve thus has a

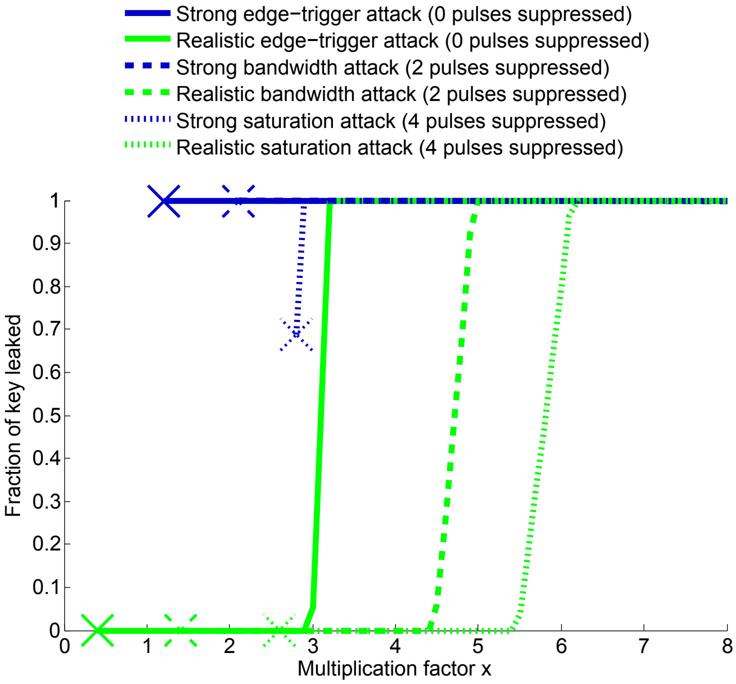


Figure 3.11: Fraction of secret key leaked to Eve in the BB84 protocol. The edge-trigger attack, which can increase μ in all pulses, allows Eve to gain full information with lower multiplication factor x than the attacks that require suppression of pulses. At low x (where the curve stops, marked by the crosses), Eve is unable to maintain the expected count rate at Bob (in the realistic attack), or induces too high QBER (in the strong attack), resulting in disclosing her presence. When the ratio is 0 (realistic attack), Eve is able to maintain the rate but cannot extract sufficient information to overcome privacy amplification. Channel loss is 3.4 dB and, in the strong attack model, Eve is restricted to a maximum QBER of 8% to avoid suspicion. This maximum QBER value was chosen because it is near the limit where Clavis2 can (sometimes) extract secure key [84]. Figure reprinted from [142].

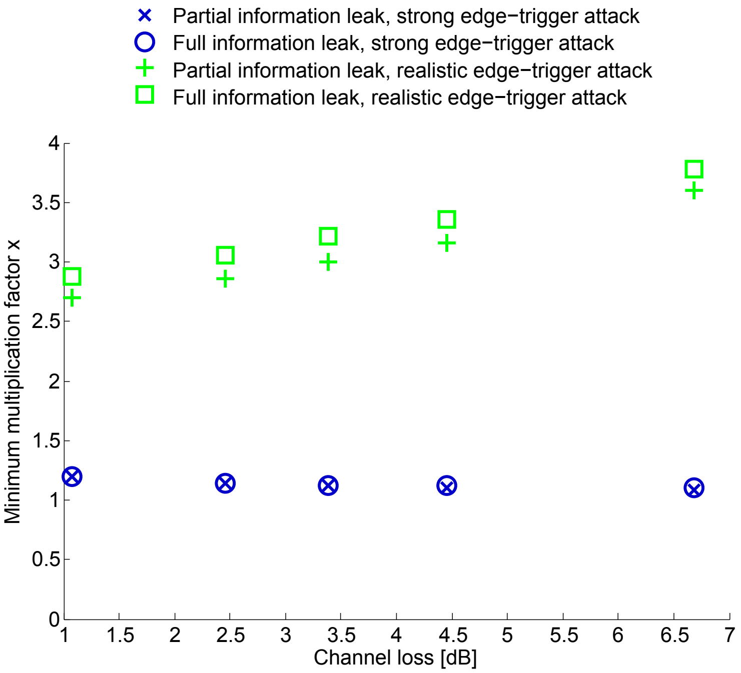


Figure 3.12: Minimum x to obtain partial and full information on the secret key in the edge-trigger attack (i.e., with no pulses suppressed) on the BB84 protocol. For the strong (realistic) attack model, Eve is able to extract partial information when between the saltire (cross) and circle (square), and full information above the circle (square). Again, Eve is restricted to a maximum QBER of 8% to avoid suspicion. Figure reprinted from [142].

probability 1/4 of giving the desired outcome. The probability that Eve fails to gain the right information when measuring n photons is then

$$E_n = \left(\frac{3}{4}\right)^n . \tag{3.13}$$

In addition, because the basis is never revealed in the analysis, Eve gains no advantage in waiting until sifting to perform her measurement. We extended both the strong and the realistic attack models to this protocol using Eve's modified probability of failure. The results are shown in fig. 3.13.

While SARG04 is more resistant to the PNS attack than BB84, it's also less resistant to the USD attack. This is because the SARG04 protocol performs privacy amplification based on the photon-number-splitting attack in which, for one measured photon, Eve extracts only 1/4 of the information. In comparison, Eve could extract full information in BB84 for one photon measurement using photon-number-splitting attack. However, the information extracted by the USD attack is the same for both SARG04 and BB84, allowing partial key extraction at lower x owing to the reduced privacy amplification performed by the SARG04 protocol. As with BB84, the attacks requiring fewer blocked pulses perform better.

3.7 Attack on quantum coin-tossing

Quantum coin tossing (QCT) allows two distrustful parties (Alice and Bob) that are separated by distance to agree on a bit value, while providing security guarantees that are stronger than classically possible. Loss-tolerant strong QCT protocol was first proposed in [177] and implemented in [20] with the use of an entangled source. The protocol was slightly modified in [178] to account for noise in the system, and enabled the implementation of QCT using a plug-and-play system [19]. The two implementations [20, 19] expanded the applicability of quantum information processing beyond QKD. Their results confirmed that using today's technology, QCT can provide a lower cheating probability than its classical counterpart. In this section we demonstrate how a deviation of μ from the ideal value can affect the performance of the QCT protocol presented in [178]. In order to take into account all standard experimental imperfections, including channel noise, multi photon pulses, losses and dark counts, Pappa *et al.* introduced an honest abort probability H, which is the probability that the protocol is unsuccessful when both parties are honest. For a desirable value of H, the two players can agree on the value of the protocol parame-

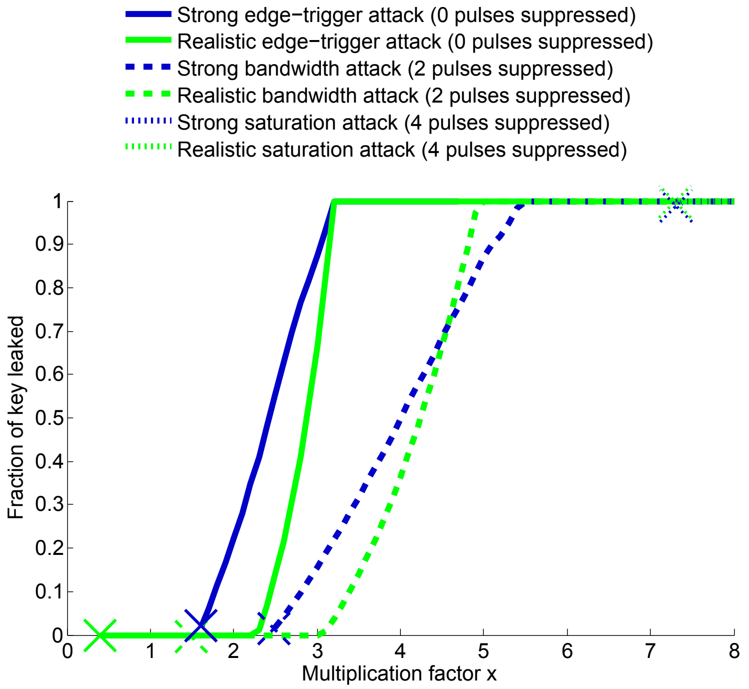


Figure 3.13: Fraction of secret key leaked to Eve in the SARG04 protocol. As with BB84, the attack that does not require Eve to suppress pulses performs better than the attacks that require pulse suppression. Once again, the missing points in the curve at low x (marked by the crosses) occur when Eve is unable to maintain the expected count rate at Bob (in the realistic attack), or induces too high QBER (in the strong attack), resulting in her presence being noticed and the key aborted. When the ratio is 0 (realistic attack), Eve is able to maintain the rate but cannot extract sufficient information to overcome privacy amplification. Channel loss is 3.4 dB and, in the strong attack model, Eve is restricted to a maximum QBER of 8% to avoid suspicion. Figure reprinted from [142].

ters, namely the number of protocol rounds K, the mean photon number μ and the state coefficient y of the (rotated) Bell states used by the protocol [19].

Alice's cheating probability only depends on the coefficient y of the quantum states, therefore a deviation of the mean photon number will not improve her strategy. However, Bob's cheating probability is a function of μ and can be upper-bounded [178, 19]

$$p^B_{cheat} \leq \sum_{i=1}^{4} P(A_i)P(\text{cheat}|A_i) + [1 - \sum_{i=1}^{4} P(A_i)]. \tag{3.14}$$

Here, $P(A_i)$ (for $i = 1, \ldots, 4$) is the probability of the four possible events where Bob receives at most one two-photon pulse in the K protocol rounds, and $P(\text{cheat}|A_i)$ is the maximum cheating probability given that event A_i occurred. For the remaining events, we consider that the cheating probability is 1.

We use the data obtained from the plug-and-play implementation of QCT over 15 km of optical fiber using Clavis2 [19], to demonstrate how a malicious Bob, having the ability to increase μ by a factor x without being detected, can increase his cheating probability. In fig. 3.14, we show the effect of the three attacks presented in this chapter, on Bob's cheating probability in comparison with the ideal case where μ does not deviate from its ideal value (in this case $\mu = 0.0019$)[4]. Using the bandwidth attack for two-pulse blocking case, the mean photon number increases to $7.3\,\mu$ while the protocol rounds decrease to $K/3$. For the saturation attack with four-pulse blocking, we have mean photon number $7.87\,\mu$ and rounds $K/5$. Finally, for the edge-triggered attack we have used $x = 10$ while keeping the number of protocol rounds the same (i.e., no pulses suppressed), resulting in unity Bob's cheating probability. We note that our modeling here upper-bounds Bob's cheating probability, considering that he has perfect equipment, controls the losses of the channel, and also has the ability to perform quantum non-demolition measurements.

We observe that, if Bob uses any of the three attacks to increase the mean photon number, and is not detected by Alice's pulse-energy-monitoring system, then there is no provable quantum advantage for coin tossing. This means that, similar to QKD, QCT is also vulnerable against the inability to maintain a constant mean photon number. In fig. 3.14, we also show how much manipulation of μ is required from a quantum Bob in order to increase his cheating probability to the classical limit. For example, for honest abort probability 0.014, if Bob is able to increase μ by $x = 1.389$ from the ideal value of 0.0019, then

[4] Note that in [19] the authors also considered errors during the state preparation (Alice), the choice of measurement basis and bit value, as well as differences in detector efficiencies (Bob). For simplicity, here we assume uniform distribution for Alice's state preparation and Bob's basis, bit choice, as well as equal detector efficiencies

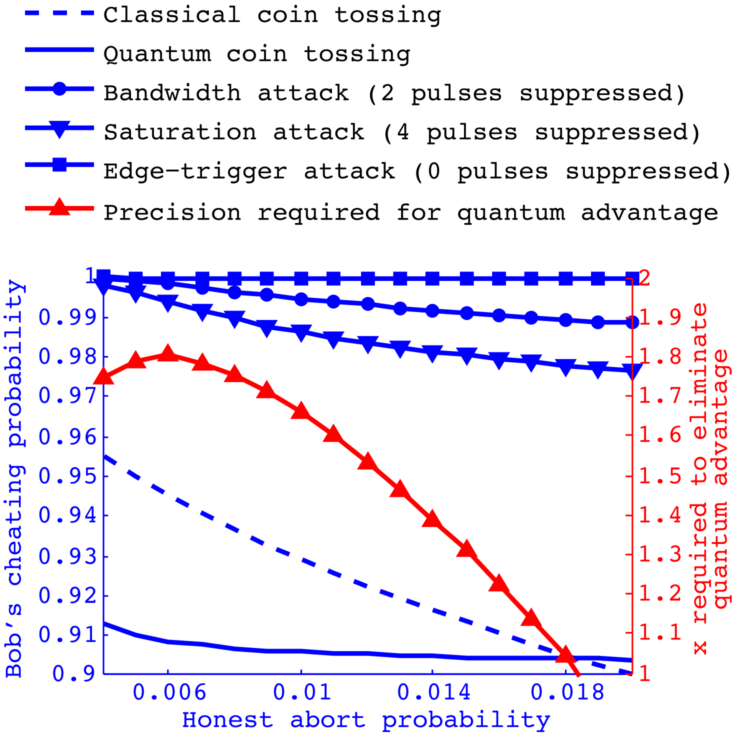


Figure 3.14: Bob's cheating probability versus honest abort probability in the coin-tossing protocol. The plot shows limits for the classical coin-tossing and QCT (for 15 km and $\mu = 0.0019$ [19]), as well as limits for the three attacks on QCT. We observe that all three Bob's attacks beat the classical limit, and QCT can therefore provide no provable advantage. Also plotted is factor x required to reduce security of QCT to that of its classical counterpart. Figure reprinted from [142].

his cheating probability becomes the same as the classical cheating probability. Equivalently, this means that for this specific honest abort probability, Alice needs to have a measurement precision of 38.9% on the value of μ, if she wants to make her protocol at least as secure as its classical counterpart. So, even if measures are taken to prevent an adversary from manipulating μ, limited experimental precision for setting the exact security parameters inherently affects the protocol performance, and can even make it insecure.

3.8 Countermeasures

Although the vision to implement the PEMS is highly appreciative and the implemented strategy is generally correct, the technical realization should be revised dramatically in order to be efficient against arbitrary Trojan-horse attacks. It requires changes in many parts of the circuit: the front-end amplifier, the integrator and the alarm detector.

The negative saturation of the transimpedance amplifier OPA380 can be prevented by pulling down its output by a $2\,\mathrm{k}\Omega$ resistor to the $-5\,\mathrm{V}$ power supply, as advised in the data sheet of the opamp [167]. Nevertheless, the amplifier bandwidth choice, which has been made on a specification considering limited classes of attacks, is not sufficient for the accurate metering of the calibrated signal (second) pulse when Eve can transfer optical energy from the first pulse to the second one. To obtain precision of, say, 10%, the amplifier output after the first pulse needs to decay to 5% of its maximum value, since the first pulse is about twice as large as the second pulse. It limits the time constant of an amplifier by the value of $50\,\mathrm{ns}/(-\ln(0.05)) = 16.7\,\mathrm{ns}$, which corresponds to bandwidth of at least $9.5\,\mathrm{MHz}$ (assuming amplifier's frequency response equivalent to an RC-filter). Hence, the front-end transimpedance amplifier should be remodeled to enhance the bandwidth.

At the moment, the integrator circuit functions more like a peak detector than an ideal integrator. Square-law dependence of the FET1 source current on the gate voltage results in non-linearity. It appears that the circuit output is more sensitive to a higher level of the input signal, which is typical for peak-detecting. This way, the circuit actually measures the pulse peak intensity rather than the pulse energy. For proper implementation, the integrator should be built in such a way that the capacitor is charged by current linearly depending on the input voltage.

The edge-triggered alarm generation by means of a monostable is not needed in this circuit at all. Instead, a simple level triggering can be used. Actually, there is no risk of the FPGA missing a too-short electrical pulse at the output of the comparator, because the voltage at C cannot rise until it is reset by the FPGA through FET2. Hence, the monostable can be simply excluded, with possibly slightly delaying resetting the integrator capacitor C to ensure a minimum time to keep the comparator output in a low logic-level state. Implementing a precise high-speed analog integrator could be challenging. Alternatively, the amplifier signal could be digitized with a fast analog-to-digital converter, and the rest of processing done numerically in the FPGA.

3.9 Conclusion

We tested the commercial plug-and-play system Clavis2 [166] for possible loopholes in the implemented pulse-energy-monitoring system (PEMS). Three implementation flaws were found in the system design that permitted three distinct modes of attacks. We named them 'bandwidth attack', 'saturation attack' and 'edge-trigger attack' respectively. We experimentally exploited each of these loopholes and were able to inject extra light into Alice without triggering any alarm. The attacks changed the actual mean photon number μ' coming out of Alice than the expected value μ permitted by the security proofs. We defined a multiplication factor $x = \mu'/\mu$ that quantifies this deviation. We modeled two types of attack. 'The strong attack' was limited by the laws of quantum mechanics only. The second attack, 'the realistic attack', was modeled using components that are found in real world. We calculated the information extracted by Eve as a function of multiplication factor for both these models and quantified the attack performances. We also analyzed the implementation of a coin tossing system and showed that it is possible for an eavesdropper to achieve a cheating probability up to 1.

This work points out the limitations of closed security standards developed inside a company – ID Quantuqie – in this case. Although the company went above and beyond everyone else's prior research in this field to secure their commercial system by implementing the PEMS, it was not sufficient; our tests found that the PEMS security was based on flawed design logic and conservative assumptions on Eve's ability. Our proposal is to develop – for security – open standards and testing methodologies in collaboration between research and industry.

Chapter 4

Security evaluation of a free space QKD system

This chapter is based on a published paper [143].

4.1 Motivation

Although fiber based QKD has been demonstrated in numerous occasions [30, 179, 180, 181, 121, 182, 55, 183, 119, 118, 120, 121, 126, 123, 35] and the technological advancement has led to even commercial production [96, 60, 184], one important drawback is that its maximum distance is limited to the order of 400 km [35]. The main reason is that losses in fiber scales exponentially with distance and after certain limiting distance, QKD is no longer possible. In comparison, the losses in free space scales much slowly, which makes it a potential candidate for long distance QKD at distances beyond the limit of fiber based QKD. Because of this, the idea of using ground to satellite link for QKD, which promises to achieve much larger distance [139], is becoming a very promising and feasible proposition.

4.2 Theory

4.2.1 Operation of the free space QKD prototype

As a representative of free-space quantum communication system, we chose a long-distance satellite QKD prototype operating at 532 nm wavelength [140] employing polarization encoded Bennett-Brassard 1984 (BB84) protocol [11]. In this particular version of the protocol, at each time slot, the sender Alice randomly chooses one out of four polarizations: horizontal (H), vertical (V), +45° (D), and −45° (A) and sends it using phase randomized weak coherent laser. At the receiving end, Bob, measures the polarization of each incoming signal in either the horizontal-vertical (HV) basis or in the diagonal-antidiagonal (DA) basis. After the transmission phase is complete, both parties communicate over a classical channel and disclose their bases to each other. They store the classical measurement results for the signals for which the bases were same and discard the rest of the results. If no eavesdropper is present and all sources of errors (i.e., visibility, dark counts etc.) are well characterized, then after sifting, the quantum bit error rate (QBER) is well below a predefined threshold value. However, if there is an adversary trying to interfere with the flying qubits in the channel, the resultant QBER will go higher than the threshold and the presence of the eavesdropper is disclosed. In this way, free space QKD promises to provide unconditional security.

In practice, our device under test employs a telescope to reduce the size of the incoming collimated beam, followed by a non-polarizing beam splitter to randomly choose between two measurement bases. It is followed by polarization beam splitters and single-photon detectors to measure photons in the four states of polarization. The telescope of our receiver consisted of a focusing lens L1 (diameter 50 mm, focal length $f = 250$ mm; Thorlabs AC508-250-A) and collimating lens L2 ($f = 11$ mm; Thorlabs A397TM-A). The choice of basis was performed by the use of a 50:50 beam splitter BS (custom pentaprism [140]). In addition to polarization beamsplitter PBS1 (Thorlabs PBS121), PBS2 (Thorlabs PBS121) was also used to increase the polarization extinction ratio in the reflected arm of PBS1. Lenses L3 (Thorlabs PAF-X-18-PC-A) focused the four beams into 105 μm core diameter multimode fibers (Thorlabs M43L01) leading to single-photon detectors (Excelitas SPCM-AQRH-12-FC). The scheme and photograph of the receiver are shown in fig. 4.1(a,c).

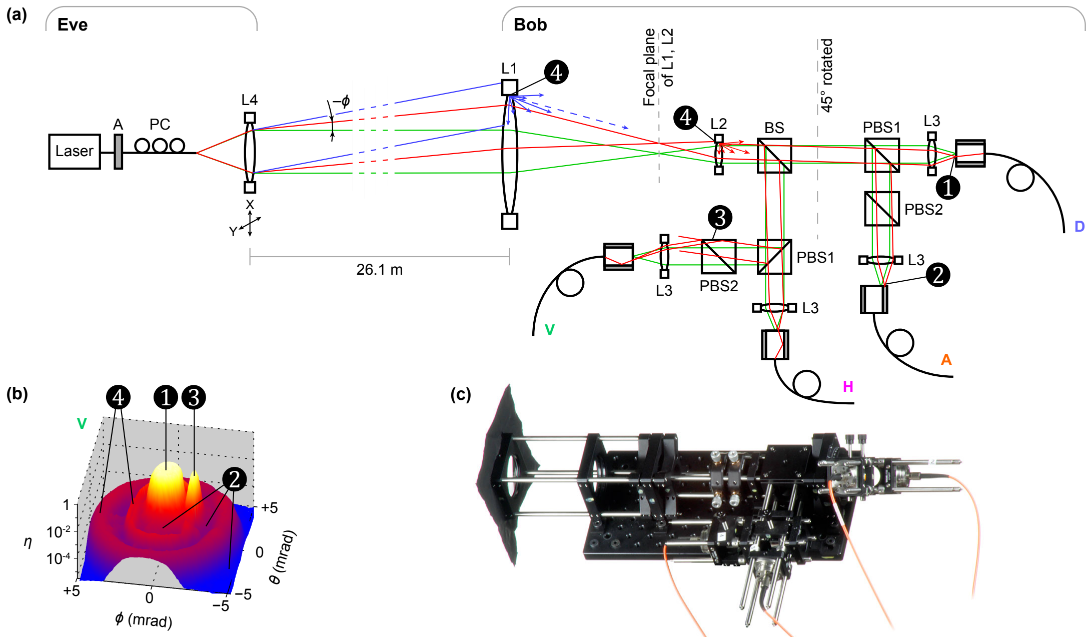


Figure 4.1: Experimental setup. (a) Scheme of the experimental apparatus, top view (drawing not to scale). Eve's source consists of a fiber-coupled 532 nm laser, attenuator A, polarization controller PC, and a collimating lens (Thorlabs C220TME-A) mounted on a two-axis motorized translation stage (Thorlabs MAX343/M). The latter allows changing the beam's incidence angle and lateral displacement at Bob's front lens L1 simultaneously. Green marginal rays denote the original alignment of Alice's beam to Bob. Red and blue marginal rays show a scanning beam from Eve tilted at an angle (ϕ, θ) relative to the original beam. Features ❶–❹ mark different transmission paths for light inside Bob. (b) Normalized detection efficiency η in channel V versus the illumination angle (ϕ, θ). This scan was taken to show the features clearly by placing Eve at a closer distance. (c) Photograph of Bob's receiver. The actual distance between facing surfaces of L2–BS is 42 mm, BS–PBS1 66 mm, PBS1–L3 31 mm, PBS1–PBS2 45 mm, PBS2–L3 10 mm in channel A and 5 mm in channel V. Figure reprinted from [143]

4.2.2 Spatial-mode-detector-efficiency mismatch side channel

One fundamental assumption of the security proofs of QKD is that the measurement outcomes must be independent of the measurement bases and not controllable by Eve. If this is not the case, the security proofs no longer holds. In theory, similar case can exist in the present scenario. While the light is traveling from Alice to Bob, an eavesdropper can, in principle, tilt the beam by an angle (ϕ, θ) such that at Bob, the beam misses, partially or fully, the cores of fibers leading to three detectors while being relatively well coupled to the core of the fourth one as illustrated in fig. 4.1a. This is probable because real-world optical alignments are inherently imperfect and manufacturing precision is finite. Thus, just by tilting the beam at a particular spatial angle, the eavesdropper may make a particular detection outcome more probable than the others and also make the outcome dependent on the choice of basis. This will allow her a control over Bob's detection outcome. In that case, it might also be possible for her to stage a faked-state attack and steal keys without introducing any error [73, 150]. This side channel is termed spatial-mode-detector-efficiency side channel and in the following, we evaluate the risk of this side channel.

4.3 Experiment

In order to test the presence of the spatial-mode-detector-efficiency mismatch side channel, we performed a 'scanning' procedure. Eve's setup consisted of a fiber-coupled 532 nm laser, attenuator A, polarization controller PC, and a collimating lens (Thorlabs C220TME-A) mounted on a two-axis motorized translation stage (Thorlabs MAX343/M) as shown in fig. 4.1. Light from Eve was circularly polarized Gaussian-shaped with 9 mm width (at $1/e^2$ peak intensity) at L1. Green marginal rays denote the initial alignment from Eve, replicating the alignment from Alice to Bob. This is the initial position of the translation stage and the corresponding angle is called the reference angle $\{\phi, \theta\} = (0, 0)$.

At first, we did a preliminary scan from a distance of 1 m. The result is shown in fig. 4.1(b). It was seen that, at the vicinity of the reference angle, light is very well coupled into the fiber core in all four channel ❶ and there is no difference in count rate among them. When the angle started to increase, the focused beam started missing the fiber core, and the detector count dropped off ❷. A region was found when the beam reflected off the polished edge of PBS2 back into the fiber core, causing the peak ❸. Increasing the angle further made the beam hit the anodized aluminum mount of L1 and possibly edges of other lens mounts and round elements in the optical assembly. It was scattered at these edges, producing two ring-like features ❹. Beyond these features, there were no noticeable

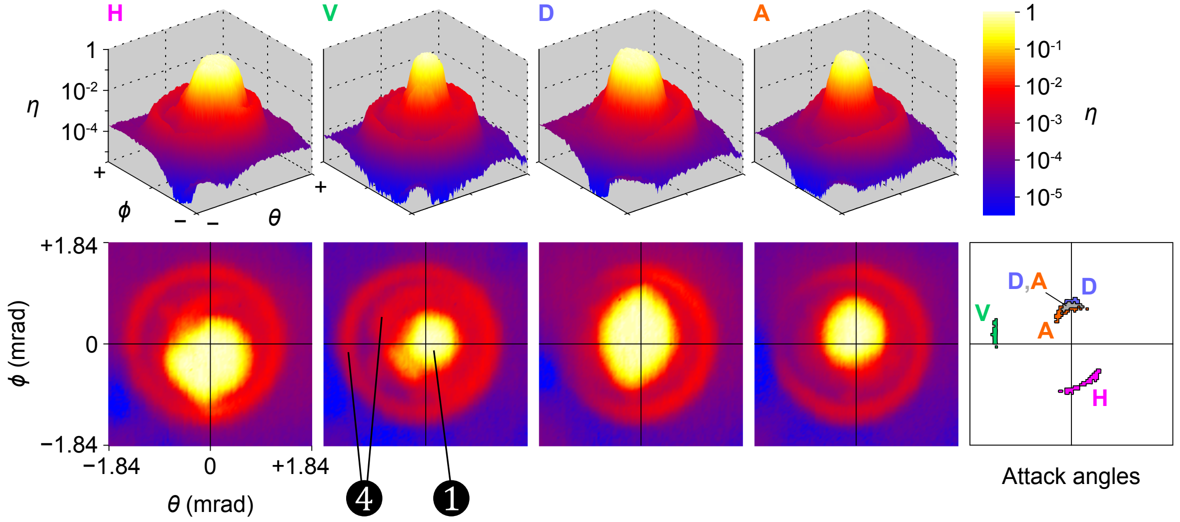


Figure 4.2: Angular efficiency scan of the receiver, and points of interest. Four pair of plots **H, V, D, A** shown in both 3D and 2D represent normalized detection efficiency in the four receiver channels versus illuminating beam angle (ϕ, θ). The angle $\phi = \theta = 0$ is the initial angle of QKD operation. The last plot shows angle ranges with a high mismatch, usable in our attack. Figure reprinted from [143]

counts other than background, as the beam completely missed the receiver aperture. We adjusted the alignment of the setup to minimize ❸ and then started the final scanning from a distance of 26.1 m.

The final scanning setup is shown in fig. 4.1. First, we moved the translation stage in the transversal plane to change the beam's incidence angle and lateral displacement at Bob's front lens L1 simultaneously. This is shown by the red marginal rays in fig. 4.1, representing a beam from Eve coming at an angle (ϕ, θ) relative to the reference beam. Then we recorded the corresponding count rate at all four detectors of Bob. For each angle, that represents a data point, we used an integration time of 1 s. Then during post-processing, for each data point from each detector, we subtracted the corresponding detector's background count rate, and normalized it after dividing by the maximum count rate in that detector. The scanning was done in 38.3 µrad steps and composed of approximately 10000 data points covering ±1.84 mrad range, corresponding to lateral displacement of ±48 mm, covering the entire clear aperture of L1.

Figure 4.2 shows the normalized detection efficiency in all four receiver channels as a function of (ϕ, θ).

4.4 Attack modeling and results

To evaluate the security risk, we modeled an attack that exploits the discovered side-channel. We modeled a practical faked-state attack using the obtained data and the following assumptions: Alice and Bob perform non-decoy-state Bennett-Brassard 1984 (BB84) protocol using polarization encoding. Alice emits weak coherent pulses with mean photon number μ equal to Alice–Bob line transmittance [65]. Whenever Bob registers a multiple click, he performs a squashing operation (double-click in one basis is mapped to a random value in that basis, while multiple clicks in different bases are discarded) [185, 186, 171]. Alice and Bob also monitor total sifted key rate, and quantum bit error ratio (QBER). Eve has information about Bob's receiver characteristics described above, and only uses devices available in today's technology. She intercepts photons at the output of Alice, using an active basis choice and superconducting nanowire detectors, with overall detection efficiency $\eta_e = 0.85$ and dark count probability $< 10^{-9}$ per bit slot [175]. Then, a part of her, situated close to Bob, regenerates the measured signal and sends to Bob. We assume that Alice–Bob and Alice–Eve fidelity $F = 0.9831$ [140], while Eve–Bob experimentally measured $F = 0.9904$. Here fidelity refers to the probability that a polarized photon will emerge from the PBS at the correct path, which is related to visibility V by $F = (1+V)/2$. We also confirmed experimentally that Eve–Bob fidelity is preserved at all illumination angles shown in fig. 4.2.

Let $\eta_i(j)$ be the normalized efficiency of Bob's i-th channel ($i \in \{h, v, d, a\}$) given that incoming light is $j \in \{H, V, D, A\}$ polarized. To maximize Eve–Bob mutual information, Eve wants to maximize the detection probability when Bob measures in the same basis as her. Thus, to find attack angles for the j-th polarization, she should pick angles that have higher values of $\eta_j(j)$. On the other hand, to reduce the QBER, she wants to minimize Bob's detection probability when Bob measures in the non-compatible basis. This requires her to pick angles for which the ratio $\delta_j(j) = \min\left\{\frac{\eta_j(j)}{\eta_{nc0}(j)}, \frac{\eta_j(j)}{\eta_{nc1}(j)}\right\}$ is maximum. Here η_{nc0} and η_{nc1} are the normalized efficiencies of the two detectors in the non-compatible basis. This ensures that when Eve's polarization basis matches that of Bob, light is detected with maximum efficiency; however, when her basis does not match that of Bob, the detection probability is very low to contribute significantly to the QBER. Our experimental attack angles are shown in the rightmost plot in fig. 4.2. In the figure, the H attack angles were composed of points for which $\eta_h(H) \geq 0.2$ and $\delta_h(H) \geq 75$. Similarly, the V, D and A attack angles were composed of points for which $\eta_v(V) \geq 0.002$, $\delta_V \geq 8$; $\eta_d(D) \geq 0.4$, $\delta_D \geq 80$; $\eta_a(A) \geq 0.1$, $\delta_A \geq 20$ respectively. Note that, the thresholds chosen here to find the attack angles were picked manually and not optimal; nevertheless, they result in successful attacks. Relaxing the thresholds to lower values will increase the area of the attack angles but may lead to higher values of QBER.

To derive an expression for the key rate and QBER in Eve's presence, we start with a system with only Eve and Bob. Let's consider Eve sending an H-polarized pulse to Bob within the attack angles H. Before squashing, the raw click probability $p_i(j)$ that detector i in Bob clicks given Eve has sent j-polarized light is

$$\begin{aligned} p_h(H) &\approx c_h + 1 - \exp\left(-\frac{\mu_H F \eta_h(H)}{2}\right), \\ p_v(H) &\approx c_v + 1 - \exp\left(-\frac{\mu_H (1-F) \eta_v(H)}{2}\right), \\ p_{d(a)}(H) &\approx c_{d(a)} + 1 - \exp\left(-\frac{\mu_H \eta_{d(a)}(H)}{4}\right), \end{aligned} \tag{4.1}$$

where μ_i is Eve's mean photon number when she is sending i-polarized light at attack angle i and c_i is Bob's background click probability per bit slot in i-th channel. The probability $P_{hv}(H)$ that after squashing Bob measures in HV basis, given Eve has sent an H-polarized pulse, is composed of three events: when only detector H clicks, when only detector V clicks, or when both click. It can be written as

$$\begin{aligned} P_{hv}(H) = &\left[1 - p_d(H)\right]\left[1 - p_a(H)\right] \\ &\times \left[p_h(H) + p_v(H) - p_h(H) p_v(H)\right]. \end{aligned} \tag{4.2}$$

Let's now include Alice into the picture. Consider Alice sends an H-polarized pulse, and Eve intercepts it. Let $P_c^e \approx \frac{1}{2}(1-e^{-\mu F \eta_e})e^{-\mu(1-F)\eta_e}$ and $P_w^e \approx \frac{1}{2}e^{-\mu F \eta_e}(1-e^{-\mu(1-F)\eta_e})$ be the probability that Eve measures in the compatible basis (i.e., the same basis as Alice) and gets a click only in the correct and wrong detector respectively. Let $P_{nc}^e \approx \frac{1}{2}(1-e^{-\frac{\mu\eta_e}{2}})e^{-\frac{\mu\eta_e}{2}}$ be the probability that she measures in the non-compatible basis (different basis than Alice's) and gets a click in a single detector. The sifted key rate given Alice has sent H-polarized light is

$$\begin{aligned} R_e(H) \approx &P_c^e P_{hv}(H) + P_w^e P_{hv}(V) + P_{nc}^e \left[P_{hv}(D) + P_{hv}(A)\right] \\ &+ (1 - P_c^e - P_w^e - 2P_{nc}^e)(c_h + c_v - c_h c_v). \end{aligned} \tag{4.3}$$

An error can occur when Eve measures Alice's signal in non-compatible basis or when Eve measures in compatible basis but Bob measures a wrong value owing to imperfect fidelity or dark count. Hence, the error rate conditioned on Alice sending H-polarized light is

$$\begin{aligned} E_H \approx & P_c^e P_v(H) + P_w^e P_v(V) + P_{nc}^e \left[P_v(D) + P_v(A)\right] \\ & + (1 - P_c^e - P_w^e - 2P_{nc}^e)(c_v - \frac{c_v c_h}{2}), \end{aligned} \tag{4.4}$$

where $P_i(j)$ is the probability that Bob measures value i after squashing, given Eve has sent j-polarized light. For example,

$$P_v(H) = \big[p_v(H) - \frac{p_h(H)p_v(H)}{2}\big]\big[1 - p_d(H)\big]\big[1 - p_a(H)\big]. \tag{4.5}$$

Sifted key rates and errors in Eve's presence [eqs. (4.3) and (4.4)] conditioned on V, D, A polarizations sent by Alice can be calculated similarly. The total sifted key rate and QBER in Eve's presence become

$$\begin{aligned} R_e &= \frac{1}{4} \sum_{j=H,V,D,A} R_e(j), \\ \mathrm{QBER}_e &= \frac{1}{4R_e} \sum_{j=H,V,D,A} E_j. \end{aligned} \tag{4.6}$$

The only free parameters left for Eve to manipulate are the mean photon numbers of her signal. Knowing the angular scanning data, Eve can use a numerical optimization to find values of μ_H, μ_V, μ_D, μ_A that minimize QBER_e while keeping $R_e = R_{ab}$, where R_{ab} is Bob's sifted key rate without Eve. Our numerical optimization achieves this for Alice–Bob channel loss ≥ 3 dB if they are willing to accept a slight increase of QBER by less than 0.7% (see fig. 4.3). Here we assumed Bob's detector parameters as measured by us: efficiency at $\phi = \theta = 0$ was 0.4 in all four channels, and individual detector background count probabilities were in the range of 430×10^{-9} to 1560×10^{-9} per 1 ns coincidence window. These optimization results are realistic conditions for a successful attack on most communication channels [187, 188, 189, 190, 191, 192, 193, 140] Note that the distance Eve–Bob can be increased without affecting attack performance, by replacing Eve's illuminator with four collimators oriented at the required attack angles.

We went further and imposed an additional constraint on Eve to make $R_e(H) = R_e(V) = R_e(D) = R_e(A) = R_{ab}$. Our optimization shows that it is still possible for Eve to pick appropriate mean photon numbers and successfully attack the system with resultant QBER $< 6.82\%$ in 3–15 dB line loss range (fig. 4.3). Similar QBER values are typical for outdoor channels, because of background light. Eve could shield Bob from the

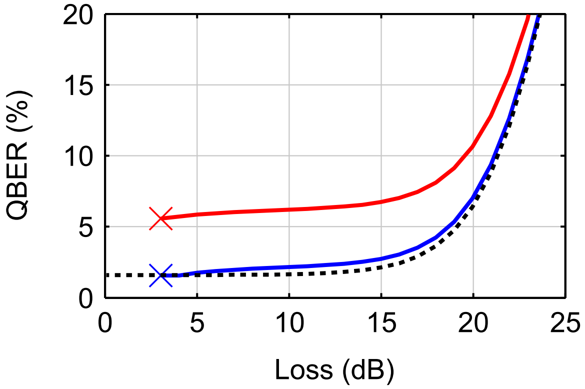


Figure 4.3: Modeled QBER observed by Bob versus line loss. The dotted curve shows QBER without Eve. At lower line loss, the QBER is due to imperfect fidelity, while at higher line loss Bob's detector background counts become the dominant contribution. The lower solid curve (blue) shows QBER$_e$ under our attack when only the total Bob's sifted key rate R_{ab} is matched. The upper solid curve (red) additionally keeps his four channel rates equal. Figure reprinted from [143]

latter to hide QBER resulting from her attack.

4.5 Countermeasure

In our attack, by sending lights at different angles, Eve has broken a fundamental assumption of security proofs that detection probabilities are independent of detection bases [194, 148]. We propose to restore this assumption by placing a spatial filter (pinhole) at the focal plane of Bob's L1 and L2 [fig. 4.1(a)]. Spatial filtering is sometimes done before the beam splitters to increase signal-to-background ratio in the channel [190, 195, 196], however it has not been characterized as a security countermeasure. We performed scanning with 100, 75, and 25 µm diameter pinholes, and found that decreasing the pinhole diameter gradually reduces the mismatch. The 25 µm diameter pinhole eliminated any visible mismatch (fig. 4.4) even though we reduced our search parameters to $\eta_i(j) \geq 0.001$ and $\delta_i \geq 4$. This pinhole provides Bob's field-of-view of 100 µrad, which does not reduce his efficiency with turbulent atmospheric channels [191]. Hence, we conclude that a 25 µm pinhole may be an efficient countermeasure for the current setup.

One may ask the validity of this attack strategy by pointing that Eve needs to stay at the line-of-sight between Alice and Bob which they can surely observe; It can also be a

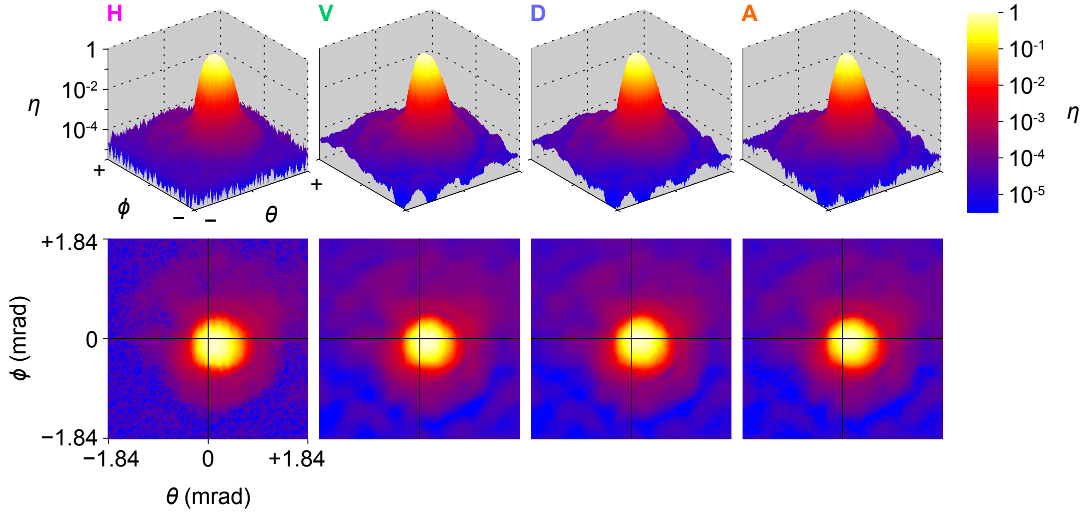


Figure 4.4: Angular efficiency scan of the receiver after a 25 μm diameter pinhole is placed in the focal plane of L1, L2 (Fig. 4.1). No detectable mismatch between channels was found under tight search conditions $\eta_i(j) \geq 0.001$ and $\delta_i(j) \geq 4$. Figure reprinted from [143]

countermeasure. However, we remark that this line of reasoning is wrong because it shifts the basis of security from quantum mechanics to the visual inspection of the channel–by some technical means like cameras or range-finders supported by image processing software. Although there is no reason why that line of solution will not work, it is not the intent of QKD to use some other mechanism to protect the channel to guarantee security. Modern practice of cryptography assumes that Eve have full access and control of the quantum channel and she can do whatever she wants (within the laws of quantum mechanics) to break the security. From this point of view, the attack proposed in this chapter is completely valid.

Usage of single-mode fibers to connect the free-space outputs of Bob's receiver to the single photon detectors can also be regarded as a countermeasure like the pinhole. However, there are two main reasons why multimode fibers are used instead of single-mode in long-distance free-space QKD receivers. First, propagation of Alice's single-mode beam through a turbulent atmosphere splits it into multiple spatial modes [197] which requires a multimode fiber for efficient collection. Second, the finite precision and speed of real-time angular tracking of Alice's beam requires that Bob accepts multiple spatial modes in a certain acceptance angle [198, 195, 191, 193]. Use of single-mode fibers under these conditions would lead to additional coupling losses $\gtrsim 10$ dB [199] if the system does not include appropriate (and often expensive) adaptive correction optics [197]. Therefore, multimode fibers and detectors with larger area are generally preferred as they allow good collection efficiency without increasing complexity and cost.

Finally, one could argue to use adaptive optics at Bob's receiver which would permit spatial mode reception and hence would be robust against any spatial-mode-detector-efficiency mismatch. Although it is a valid countermeasure, the practicality of it – considering present long distance QKD implementations are complex and lossy – is yet to be tested. Note that, in Refs. 70 and200, a detector scrambling strategy was proposed that might also be an effective countermeasure against efficiency mismatch attacks for single-photon qubits. However, it is not clear how effective that countermeasure is, when one considers that the detectors operate on optical modes, not on single-photon signals. This can be a future study.

4.6 Conclusion

In this work, we investigate a long-distance free-space polarization based QKD receiver and experimentally demonstrate efficiency mismatches that an eavesdropper can take advantage of. We identify the sources of the mismatches. We also experimentally demonstrate the feasibility of a spatial-mode-efficiency-mismatch attack and quantify the performance of the attack by modeling a faked-state intercept-resend attack. Our results show that under the assumptions made in this chapter, the security is compromised for a loss range of $3 - 20$ dB.

Although our practical attack should work, and the physical countermeasure seems promising, there is still room for improvement on both the attack scheme and countermeasures. Eve can employ more attack angles or combine this attack with some other suitable attack schemes to increase the number of her free parameters. Alice and Bob can make this harder by monitoring more parameters. Thus, it is a cat and mouse game which eventually – after several iterations – should lead to a proper guideline for the standardization and certification of the free space QKD schemes against spatial-mode-efficiency mismatch attacks.

After finishing this security evaluation, we recommend the following for the standardization and certification of quantum communication against spatial-mode-detector-efficiency side-channel. First, the user should characterize and check the receiver system for the existence of spatial-mode-detector-efficiency mismatch. This characterization method needs to be optimized, standardized and certified. Secondly, the user should choose and employ proper countermeasure such as employing the optimum-sized spatial filter, or use adaptive optics, single mode fibers, or use the detector scrambling proposals from Refs. 70 and 200. After employing the proper countermeasure, its effectiveness must be tested and its effects on the normal system operation of the system must be analyzed; this testing method also needs to be standardized and certified.

Chapter 5

Security evaluation of quantum communication systems against laser damage

This chapter is based on a published paper [91].

5.1 Motivation

'Laser damage' refers to manipulating the characteristics of a system component from its characterized behavior by exposing it to high power laser [201]. The fact that laser damage can create deviations in device characteristics in isolated system components has already been shown in [141]. There are now open questions: is it possible to do the same in a practical and running QKD system, is it feasible for an eavesdropper to do so, and how much threat laser damage possess for the security of QKD systems. To answer all these questions, there needs to be a risk evaluation for the security of QKD against laser damage.

Unfortunately, to the best of our knowledge, no such risk evaluation has been reported in the literature so far. Hence, we embark on our journey to test the security of practical quantum communication systems against laser damage. We test laser damage attack on two different implementations of quantum communication system. The first one is a fiber based plug-and-play system implementing quantum key distribution and coin tossing protocol. The second one is a long distance free space quantum key distribution system. In both cases, we aim to laser damage a security critical component while the system is up and running. Our goal is to investigate the feasibility of the attack and also evaluate the risk. The details are given in the next sections.

5.2 Theory

Practical quantum communication systems consist of many components. Ideally, the components should have well characterized properties and there should be no deviation from it. In practice, there are deviations. However, all deviations may not necessarily lead to a security vulnerability, rather they might lead to a denial of service. Below, we consider few examples of probable deviations in device characteristics, and discuss their consequences.

1. **Attenuator:** In order to set a precise value of the outgoing mean photon number μ, a calibrated optical attenuator is required in the implementations of ordinary QKD [144, 96], decoy-state QKD [140], coherent-one-way QKD [59], measurement-device-independent QKD [121], continuous-variable QKD [47], digital signature [202], relativistic bit commitment [203], coin-tossing [19] and secret-sharing [204] protocols. An unexpected increase of attenuation may make μ too low causing 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 multi-photon pulses [see section 2.3.5] [142, 45]. E.g., in QKD and secret-sharing this will allow eavesdropping of the key, and in bit commitment cheating the committed bit value.

2. **Synchronization and monitoring detectors** Some implementations use a detector for time synchronization [144, 96, 59, 121, 47, 203, 19, 204]. Desensitizing it may result in the denial-of-service. However, several implementations require a calibrated monitoring detector for security purposes [144, 96, 59, 47, 203, 19, 204]. A reduction in its sensitivity may lead to security vulnerabilities such as a Trojan-horse attack that reads the state preparation [81]. This leaks the key in QKD, increases the cheating probability in coin-tossing [142], leaks the program and client's data in quantum cloud computing [29] and allows forging of digital signatures [202].

3. **Linear optics components:** Many implementations use beam splitters and rely on their pre-characterized splitting ratio (e.g., [144, 96, 140, 59, 47, 203, 19]). A shift in the splitting ratio may lead to either the denial-of-service or security vulnerabilities (e.g., [205] or one of the above-mentioned attacks).

4. **Encoding device:** 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 [206, 78, 207].

5. **Detector properties:** If the dark count rate of single-photon detectors is increased, it may lead to the denial-of-service [141]. However, if it can be decreased, an eavesdropper might use it to her advantage.

Even in device-independent QKD (DI-QKD) [104], there are assumptions on the absence of information-leakage channels and memory [208]. Thus, there is a risk that these assumptions may be compromised by deviations in device characteristics. To give a speculative illustration, let's suppose detectors in DI-QKD emit light on detection [209, 210, 211], 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.

5.3 Laser damage in free space quantum communication system

To evaluate the risk of laser damage on free space quantum communication system, we chose the long-distance satellite QKD prototype employing Bennett-Brassard 1984 (BB84) protocol [11]. The construction and the operating principle of this system has already been outlined in section 4.2.1. As concluded in section 4.5, in order to make this system secure against the spatial-mode-efficiency-mismatch attack, the use of a spatial filter or pinhole was necessary. As a part of the risk evaluation procedure, we tested the endurance of this pinhole against laser damage.

5.3.1 Experimentation

The experiment consisted of three steps. Firstly, we performed the same scanning procedure as described in section 4.3 to certify that no spatial-mode-detector-efficiency-mismatch exists with the presence of the pinhole. Secondly, we damaged the pinhole using high power laser to increase the pinhole diameter such that the effect of the countermeasure is nullified. Finally, we performed scanning again to demonstrate that the system has become vulnerable against the spatial-mode-detector-efficiency-mismatch again and it 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 physical interactions with Bob.

The first step involved changing the outgoing beam's angle (ϕ, θ) emitted from Eve's scanning setup as shown in fig. 5.1a, then recording the corresponding count rate at all four detectors in Bob. The step is identical to that mentioned in section 4.3. The result of this step is shown in fig. 5.2a, 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 guaranteeing the security of the system against the spatial-mode-efficiency-mismatch attack as suggested and demonstrated in section 4.5.

Then in 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. 5.1a). 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) [140]. While the damaging beam was on, the detectors counted at their saturation rate of $\sim$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 [157, 141]. 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.

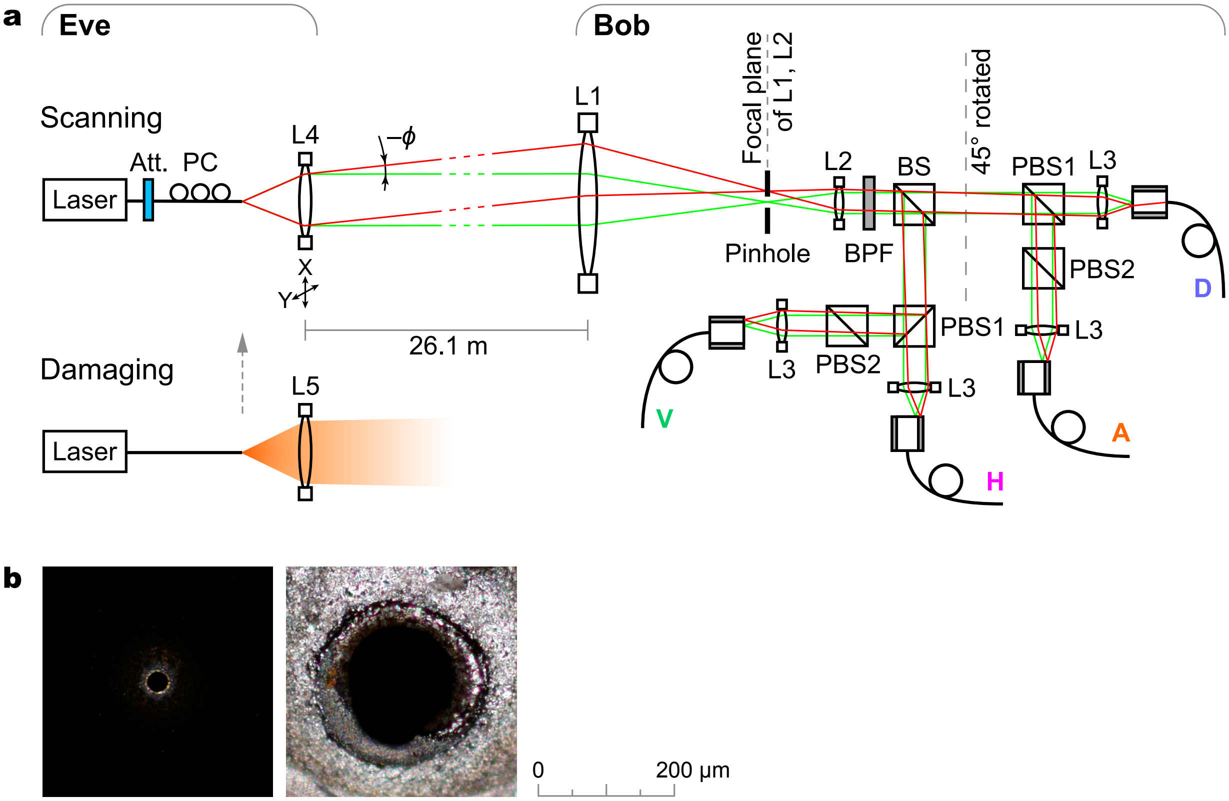


Figure 5.1: Attack on free-space QKD system. **a**, Experimental setup. QKD receiver Bob consists of two lenses L1, L2 reducing input beam diameter, 50:50 beam splitter BS, and two arms measuring photons in HV and DA polarizations using polarizing beam splitters PBS [143, 140]. Photons are focused by 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 micro-photographs show front view of the pinhole. Figure reprinted from [91]

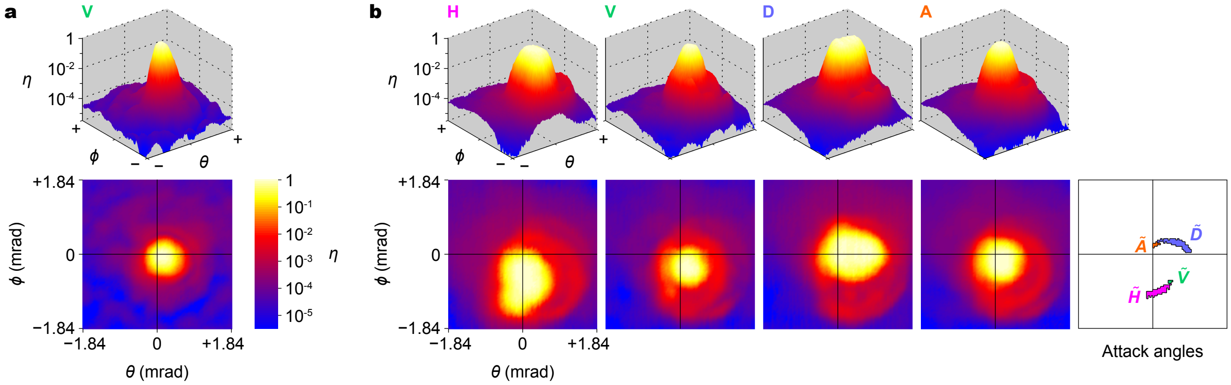


Figure 5.2: Efficiency-mismatch side-channel opened after laser damage in free-space QKD system. Each pair of 3D–2D plots shows normalized photon detection efficiency η in a receiver channel versus illuminating beam angles ϕ and θ. **a**, Before laser damage, the angular dependence is essentially identical between the four channels [143]. Plot for one channel (V) before damage is shown. **b**, After the laser damage, the four receiver channels H, V, D, 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. Figure reprinted from [91]

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. 5.2b. 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 presented in section 4.3 without the pinhole, because of extra scattering at the edges of our laser-enlarged pinhole. The fact that a measurable amount of mismatch was found in the data showed that the system had become vulnerable again to the spatial-mode-efficiency-mismatch due to the application of laser damage. To quantify the insecurity, we used the attack model developed in section 4.4 and simulated the attack performance using the present data. The details are given next.

5.3.2 Predicted attack on free-space QKD system with damaged pinhole.

In this section, to quantify the insecurity, we model the same faked-state attack as described in section 4.4. The thresholds for H-polarized pulse is, $\eta_h(\tilde{H}) \geq 0.6$ and $\delta(\tilde{H}) = \min\left\{\frac{\eta_h(\tilde{H})}{\eta_d(\tilde{H})}, \frac{\eta_h(\tilde{H})}{\eta_a(\tilde{H})}\right\} \geq 100$ (see section 4.4 for details). Similarly, for V, D and A polar-

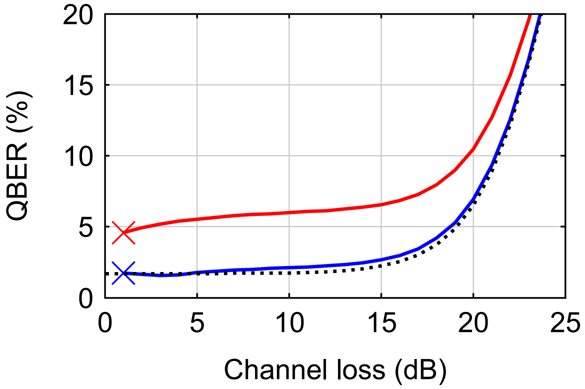


Figure 5.3: Modeled QBER observed by Bob in free-space QKD system. The dotted curve shows QBER without Eve. At lower channel loss, the QBER is due to imperfect fidelity, while at higher channel loss Bob's detector background counts become the dominant contribution. The lower solid curve (blue) shows 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. Figure reprinted from [91]

ized pulse, 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. 5.2b. We remark that the thresholds η and δ have been chosen manually so that they lead to a successfully attack and are not optimal values. As in section 4.4, we assumed that Alice–Bob and Alice–Eve fidelity $F = 0.9831$ [140, 143], while Eve–Bob experimentally measured $F = 0.9904$. All other assumptions were the same as in section 4.4 [143]. The simulation result is shown in fig. 5.3. It is clear that Eve can successfully perform the attack with a resultant QBER $\leq 6.6\%$ ($\leq 2.5\%$) in 1–15 dB channel loss range if we choose the constrain to maintain equal individual (total) detection rates (see section 4.4).

5.3.3 Risk evaluation

Our risk evaluation implies that it is feasible for an eavesdropper to create a deviation in a security critical component of free space QKD system on demand. In the present case, the target was damaging the pinhole to enlarge the opening. Once the pinhole diameter was enlarged, it was again possible to send light at higher mismatch angles as shown in fig. 5.2b. This enabled a faked-state attack under realistic conditions. Laser damage completely neutralizes the pinhole countermeasure, and the evaluated risk is very high.

5.4 Laser damage in fiber-optic quantum communication system

To evaluate the risk of laser damage on fiber based QKD systems, we choose the fiber based plug-and-play system Clavis2 [144] that implements both BB84 and SARG quantum key distribution protocols and also loss-tolerant quantum coin tossing (QCT) [19] protocol. As outlined in section 3.2.1, the secure operation of Clavis2 requires an upper bound on the mean photon number μ coming out of Alice (Otherwise, an eavesdropper can perform a Trojan-horse attack [81]). It is thus crucial that a portion of the incoming light is fed to the pulse-energy-monitoring detector (D_{pulse} such that that whenever extra energy is injected, an alarm is produced [142]. We tested the endurance of this countermeasure against laser damage.

5.4.1 Experimentation

First, we disconnected the channel Alice–Bob temporarily and connected Eve (fig. 5.4a). Then we injected 1550 nm laser light from an erbium-doped fiber amplifier for 20–30 s, delivering continuous-wave (c.w.) 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. The physical damage is shown in fig. 5.4b.

We tested a total of 6 photodiode samples and damaged each of them. 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 [142]. This quantified the reduction in sensitivity due to the damage. Three samples were exposed twice to a progressively higher power. Sample 1 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 sample 2 these numbers were 0.75 W with no change in sensitivity then 1.0 W with reduced photosensitivity of 1.6 dB (shown in 2nd microphotograph in fig. 5.4b). For sample 3 these numbers were 1.0 W, 5 dB then 1.5 W, 5.5 dB (shown in 3rd microphotograph in fig. 5.4b). 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. 5.4b) 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.

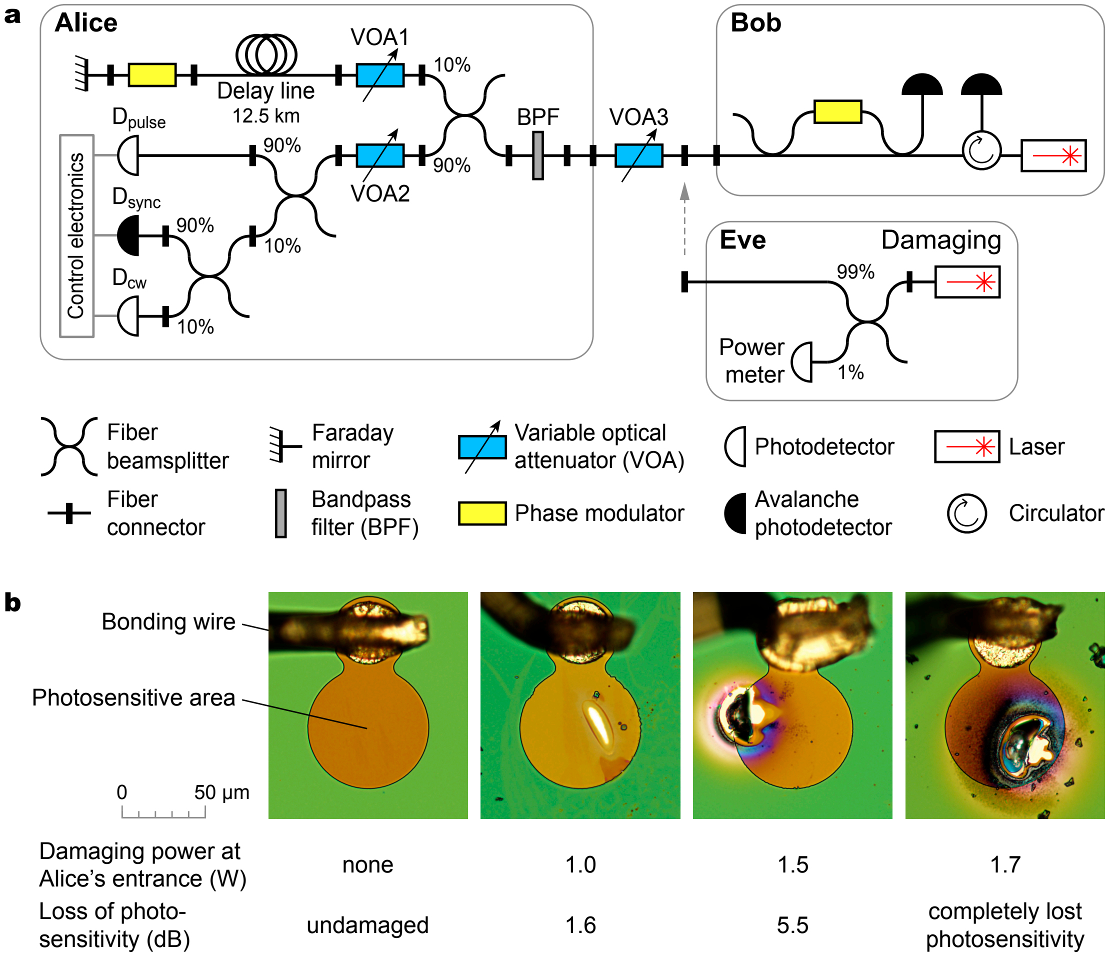


Figure 5.4: Attack on fiber-optic system Clavis2. **a**, Experimental setup. The system consists of Alice and Bob connected by a lossy fiber communication channel (simulated by variable optical attenuator VOA3). Bob sends to Alice pairs of bright coherent optical pulses, produced by his laser and two fiber arms of unequal length [144, 96]. 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 top-view of decapsulated photodiode chips. The last two samples have holes melted through their photosensitive area. Scattered dark specks are debris from decapsulation. Figure reprinted from [91]

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. 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.

For damaging and component tests, an erbium-doped fiber amplifier seeded from a 1550.7 nm laser source (EDFA; IPG Photonics ELR-70-1550-LP) was used. The injected 0–2 W c.w. power at Alice's entrance was monitored with a 1:99 fiber beamsplitter tap and a power meter (fig. 5.4a). 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 [96]) set the variable attenuator VOA2 at 2 dB. Thus, 44% of Alice's incoming light impinged $\mathrm{D_{pulse}}$, while smaller fractions impinged $\mathrm{D_{sync}}$ and $\mathrm{D_{cw}}$. The alarm threshold of $\mathrm{D_{pulse}}$ is calibrated when the system is assembled at the factory, and is not changed after that [142]. VOA3 introduced channel loss of 1.87 dB, to simulate the effect of ≈ 9 km long fiber line Alice–Bob. Figure 5.5 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.

5.4.2 Risk evaluation

In our risk evaluation, we use the simulation results presented in section 3.5. Considering the strong attack, i.e., assuming Eve is only limited by the laws of quantum mechanics, for the case of BB84 QKD protocol, she can extract partial or full key when the multiplication factor is increased in the range of $x \in \{1.1 - 1.2\}$, in fig. 3.12 in section 3.6.1. This corresponds to a sensitivity reduction of 0.4–0.8 dB range for $\mathrm{D_{pulse}}$ for Alice–Bob channel loss in the 1–7 dB range. Using the results from section 3.6.2 for SARG04 protocol, we can predict that Eve can extract partial or full key, for a sensitivity reduction in the range of $2.0 - 5.1$ dB for a channel loss of 3.4 dB. Finally, for QCT with a dishonest Bob, all

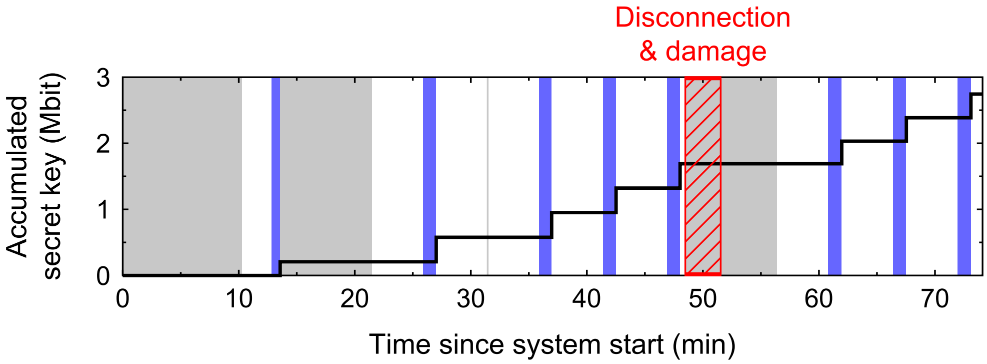


Figure 5.5: Fiber-optic QKD system operation during laser damage. The plot shows accumulated secret key amount versus time. Grey bands denote the system performing recalibration routines, white bands denote the quantum bit sending and receiving, and blue (darker) bands denote classical post-processing. 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. Figure reprinted from [91]

the quantum advantages of the protocol are eliminated if sensitivity reduces by 2.6 dB ($x = 1.805$) for a 15 km channel. Since we have demonstrated experimentally, that laser damage can achieve a sensitivity reductions of D_{pulse} higher than the threshold reduction values for BB84, SARG04 and QCT, we conclude that the security of all the three protocols have been compromised by laser damage. The evaluated risk is very high.

5.5 Conclusion

In this chapter, we have demonstrated a laser-damage attack which is capable of modifying device behavior on-demand. We have tested it on two different implementations of quantum communication system: a fiber based system running QKD and a coin-tossing protocols; and a free space based system running polarization encoded QKD protocol. In both cases, we demonstrate that laser damage can create deviations in one or more components without being detected by the implemented detection mechanisms. We further show that the newly created deviations are enough to break the security of the system completely. This reveals that laser damage is a potential security risk to existing QC systems, and necessitates further testing to guarantee security against it.

Chapter 6

Security evaluation of detector-device-independent-QKD

This chapter is based on a published paper [145].

6.1 Motivation

Since the first theoretical proposal in 1984 [11] to the recently published result of key exchange over 1200 km [36], QKD has marched forward a long way. However, the march has been consistently questioned by a number of attacks [30, 79, 70, 73, 78, 207, 143] that exploited the deviations between theory and practice. Ironically, the attacks eventually aided the march by equipping QKD with improved protocols and countermeasures; measurement-device-independent QKD (mdiQKD) protocol being one of the most effectives of them all.

Although mdiQKD is secure from all detector side-channels [79, 70, 212, 73, 78, 151, 77, 94, 143] and its practicality has also been confirmed numerous times [119, 118, 120, 121, 126, 123, 35], it has a major drawback. It requires high-visibility two-photon interference between independent sources, which makes its implementation more demanding than that of conventional QKD schemes. In addition, although recent proposals [213] significantly improve the performance, current finite-key security bounds against general attacks [214] require larger post-processing data block sizes than those of standard QKD. As a result, an alternative protocol, having the ease of implementation of conventional QKD along with the superior security of mdiQKD was the holy grail for the community. This is where detector-device-independent QKD (ddiQKD) came into the picture.

Within a span of one year, four different groups came up with the idea of Detector-device-independent QKD (ddiQKD) [114, 113, 115, 116] that followed the same spirit of mdiQKD. Later, an implementation was also carried out [215]. The key idea of ddiQKD is to replace the two-photon Bell state measurement (BSM) with a two-qubit single-photon BSM [117]. This requires that Alice and Bob use two different degrees of freedom of the single-photons to encode their bit information. In so doing, one avoids the need for interfering photons from independent light sources.

Although ddiQKD promises of being easily implementable along with providing device independent security at the detector side, its security proofs were based on assumptions that restricted the ability of an eavesdropper [114, 215]. It is not clear how secure the protocol would have been, had the restrictions been absent. This is the motivation for performing a security evaluation of the ddiQKD protocol.

6.2 Theory

6.2.1 Measurement-device-independent QKD

Measurement-device-independent QKD (mdiQKD) was first proposed in [110] with the aim to make practical QKD immune to detector side-channel attacks [73, 151, 156, 76, 75, 157, 158, 159, 79, 70, 212, 78, 77, 94, 143].An example of a possible implementation of mdiQKD is illustrated in fig. 6.1(a) [110]. To simplify the discussion, we shall assume that Alice and Bob have perfect single-photon sources, although other type of sources (for example, phase-randomized weak coherent pulses in combination with decoy states [97, 99, 98]) can also be used. The steps of the protocol are as follow:

1. Alice and Bob generate BB84 states [11] (i.e., horizontal (H), vertical (V), +45° (D), and $-45°$ (A)) randomly and independently of each other and send them to an untrusted relay Charles.

2. An honest Charles is supposed to perform a two-photon BSM that projects the incoming signals into a Bell state. Then he has to broadcast which of his measurements were successful together with the results (i.e., the Bell states obtained).

3. Alice and Bob broadcast the bases used.

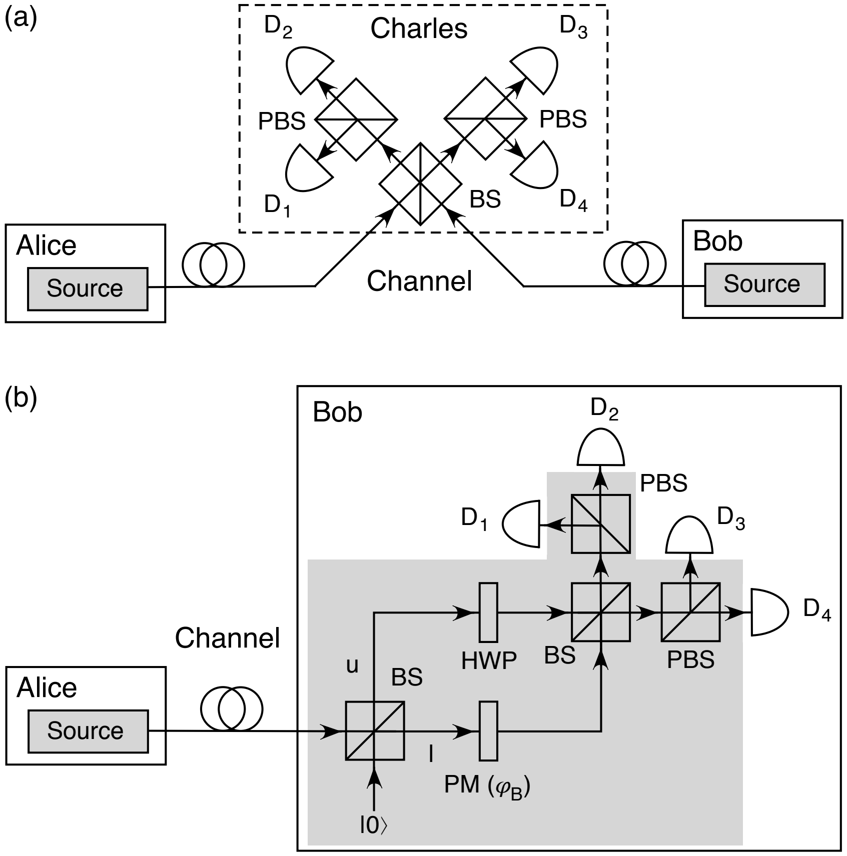


Figure 6.1: Possible implementations of partially-device-independent QKD with linear optics. (a) mdiQKD [110]. PBS, polarizing beam splitter; BS, 50 : 50 beam splitter; and D$_i$, with $i \in \{1, 2, 3, 4\}$, Charles' single-photon detectors. (b) ddiQKD [113]. HWP, half-wave plate; and PM, phase modulator. One single click in the detector D$_1$, D$_2$, D$_3$, or D$_4$ corresponds to a projection into the Bell state $|\Psi^+\rangle$, $|\Phi^+\rangle$, $|\Psi^-\rangle$, or $|\Phi^-\rangle$ respectively (see main text for further details). In both schemes, the grey areas denote devices that need to be characterized and trusted. Also, Alice's and Bob's laboratories need to be protected from any information leakage to the outside. Figure reprinted from [145]

4. If Charles broadcast successful measurement, and Alice and Bob used the same basis, then they can extract a secret key from those successful events. Importantly, if Charles is honest, his BSM post-selects entanglement between Alice and Bob, and, therefore, he is not able to learn any information about their bit values.

5. To test whether or not Charles is honest, Alice and Bob simply compare a randomly chosen subset of their data to see if it satisfies the expected correlations associated to the Bell states announced. If it does not, they abort the protocol, otherwise, they start post processing for distilling secret keys.

Interestingly, the steps of mdiQKD protocol can be seen as a time-reversed version of the Einstein-Podolsky-Rosen QKD (EPR-QKD) protocol [111]. Therefore, similar to (EPR-QKD), its security can be proven without any assumption on the behavior of Charles' measurement unit.

6.2.2 Detector-device-independent QKD

An example of a possible implementation is illustrated in fig. 6.1(b) [113] (see also [114, 115, 116] for similar proposals). The steps are:

1. Alice encodes her bit value in polarization degree of freedom by preparing the states: $|\psi\rangle_A = (|\mathrm{H}\rangle + e^{i\theta_\mathrm{A}}|\mathrm{V}\rangle)/\sqrt{2}$, where $|\mathrm{H}\rangle$ ($|\mathrm{V}\rangle$) denotes the Fock state of a single-photon prepared in horizontal (vertical) polarization, and the phase $\theta_\mathrm{A} \in \{0, \pi/2, \pi, 3\pi/2\}$. The photon is then sent out through the quantum channel towards Bob.

2. Bob encodes his bit value using the spatial degree of freedom of the incoming photons. This is done with a $50:50$ beam splitter (BS) together with a phase modulator (PM) that applies a random phase $\varphi_\mathrm{B} \in \{0, \pi/2, \pi, 3\pi/2\}$. Thus Bob's prepared state is, $|\psi\rangle_B = (|\mathrm{u}\rangle + e^{i\theta_\mathrm{B}}|\mathrm{l}\rangle)/\sqrt{2}$ where $|\mathrm{u}\rangle$ ($|\mathrm{l}\rangle$) represents the state of a photon that goes through the upper (lower) arm of the interferometer (see fig. 6.1(b)). Note that, up to this point, both Alice's and bob's state preparation part are assumed to be characterized and trusted.

3. Then, similar to mdiQKD, a BSM is performed that projects the two qubits from Alice and Bob (contained in a single photon) into a Bell state. In fig. 6.1b, a detection event ("click") in only one detector D_i corresponds to a projection on one of the four Bell states, $|\Phi^\pm\rangle = (|\mathrm{H}\rangle|\mathrm{u}\rangle \pm |\mathrm{V}\rangle|\mathrm{l}\rangle)/\sqrt{2}$ and $|\Psi^\pm\rangle = (|\mathrm{H}\rangle|\mathrm{l}\rangle \pm |\mathrm{V}\rangle|\mathrm{u}\rangle)/\sqrt{2}$. Although in the figure, the BSM setup is placed in Bob's laboratory, but there is no restriction

on that. After the measurement, Bob has to broadcast which of his measurements were successful together with the results and the rest of the protocol follows that of mdiQKD.

6.3 Results

6.3.1 The security of ddiQKD is not based on post-selected entanglement

At a first sight, it may seem that the security of ddiQKD follows directly from that of mdiQKD, given, of course, that the assumptions on Alice's and Bob's state preparation processes are fulfilled [113, 114, 115, 116]. That is, Alice sends $|\psi\rangle_A$ and Bob sends $|\psi\rangle_B$, and a BSM is performed on them. Whenever the BSM is successful, it post-selects entanglement between Alice and Bob. A first indication that confronts this idea was given in [216]. There, it was shown that, in contrast to mdiQKD, ddiQKD is actually insecure if Eve is able to replace Bob's detectors with a measurement apparatus that leaks information to the channel [216]. Although this result is important from a conceptual point of view, it violates one of the security assumptions of ddiQKD: Bob's detectors have to be built by a trusted party (but do not need to be characterized) to avoid that they intentionally leak key information to the outside [114]. Here we show that even in this scenario, the security of ddiQKD cannot be based on post-selected entanglement alone, unlike mdiQKD.

For this, we will consider that Bob's receiver in the scheme of fig. 6.1(b) has only one active detector, say for instance, the detector D_1, while the other three detectors are disabled. Now a successful BSM projects the incoming photons only into the Bell state $|\Psi^+\rangle$. If the security of ddiQKD is based on post-selected entanglement, this modification should not affect its security other than reducing the secret key rate by a factor of four. Projection into a single Bell state should be sufficient to guarantee security [110]. Below we show that a blinding attack [73, 151] renders ddiQKD insecure in this situation.

LetÕs assume an intercept-resend attack scenario (see section 2.3.1). Eve measures each of Alice's signals in one of the two BB84 bases chosen randomly. For each measured signal, she sends Bob a coherent state prepared in the BB84 polarization state identified by her measurement. We also assume that Eve has blinded Bob's detector (see section 2.3.7). That is, Eve shines bright light onto Bob's detector D_1 to force it to enter linear-mode operation [73, 151]. In this mode the detector is no longer sensitive to single-photon pulses, but it can only detect strong light. We assume that when D_1 receives a bright pulse

Table 6.1: Mean photon number of the input light to Bob's detectors as a function of the phases ϕ_{E} and φ_{B}.

(a) $\phi_{\mathrm{E}} = 0$

φ_{B}	D_1	D_2	D_3	D_4
0	μ	μ	0	0
$\frac{\pi}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
π	0	0	μ	μ
$\frac{3\pi}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$

(b) $\phi_{\mathrm{E}} = \frac{\pi}{2}$

φ_{B}	D_1	D_2	D_3	D_4
0	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
$\frac{\pi}{2}$	μ	0	0	μ
π	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
$\frac{3\pi}{2}$	0	μ	μ	0

(c) $\phi_{\mathrm{E}} = \pi$

φ_{B}	D_1	D_2	D_3	D_4
0	0	0	μ	μ
$\frac{\pi}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
π	μ	μ	0	0
$\frac{3\pi}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$

(d) $\phi_{\mathrm{E}} = \frac{3\pi}{2}$

φ_{B}	D_1	D_2	D_3	D_4
0	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
$\frac{\pi}{2}$	0	μ	μ	0
π	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$	$\frac{\mu}{2}$
$\frac{3\pi}{2}$	μ	0	0	μ

of mean photon number μ it always produces a click, while if the pulse's mean photon number is $\mu/2$, it never produces a click. This behavior has been experimentally confirmed in many detectors types [73, 151, 156, 76, 75, 157, 158, 159]. Suppose, the signals from Eve are coherent states of the form $|\sqrt{2\mu}\rangle$ with creation operator $a^\dagger = (a^\dagger_{\mathrm{H}} + e^{i\phi_{\mathrm{E}}}a^\dagger_{\mathrm{V}})/\sqrt{2}$. Here, $a^\dagger_{\mathrm{H}}$ ($a^\dagger_{\mathrm{V}}$) denotes the creation operator for horizontally (vertically) polarized photons, and the phase $\phi_{\mathrm{E}} \in \{0, \pi/2, \pi, 3\pi/2\}$ depends on Eve's measurement result. Then, it can be shown that at the input of Bob's detector $\mathrm{D_i}$, the state is a coherent state of the form (see appendix A for the derivation)

$$\begin{aligned} |\psi\rangle = \quad & |\frac{\sqrt{\mu}}{2}\big(e^{i\phi_{\mathrm{E}}} + e^{i\varphi_{\mathrm{B}}}\big)\big\rangle_{\mathrm{D}_1} \otimes |\frac{\sqrt{\mu}}{2}\big(1 + e^{i(\phi_{\mathrm{E}}+\varphi_{\mathrm{B}})}\big)\big\rangle_{\mathrm{D}_2} \\ \otimes \quad & |\frac{\sqrt{\mu}}{2}\big(e^{i\phi_{\mathrm{E}}} - e^{i\varphi_{\mathrm{B}}}\big)\big\rangle_{\mathrm{D}_3} \otimes |\frac{\sqrt{\mu}}{2}\big(1 - e^{i(\phi_{\mathrm{E}}+\varphi_{\mathrm{B}})}\big)\big\rangle_{\mathrm{D}_4} . \end{aligned} \tag{6.1}$$

This situation is illustrated in table 6.1 where we show the mean photon number of the incoming light to Bob's detectors for all combinations of ϕ_{E} and φ_{B}. Most importantly, from this table we can see that if D_1 is the only active detector and Eve selects μ such that $\mu/2 < \mu_{\mathrm{th}} < \mu$, where μ_{th} is the threshold intensity (see section 2.3.7), then Bob only obtains a click when he uses the same measurement basis as Eve (*i.e.*, when $\varphi_{\mathrm{B}}, \phi_{\mathrm{E}} \in \{0, \pi\}$

or $\varphi_{\mathrm{B}}, \phi_{\mathrm{E}} \in \{\pi/2, 3\pi/2\}$), and $\varphi_{\mathrm{B}} = \phi_{\mathrm{E}}$. That is, this attack does not introduce any error. Note that, Intercept-resend attacks correspond to entanglement-breaking channels and, therefore, they cannot lead to a secure key [217]. Moreover, we have that Bob and Eve select the same basis with at least 1/2 probability. This means that the ddiQKD scheme illustrated in fig. 6.1(b) (with only one active detector) is actually insecure against the detector blinding attack for a total system loss beyond only 3 dB, just like standard QKD schemes. This proves that the security of ddiQKD cannot be based on post-selected entanglement. The same conclusion applies as well to the ddiQKD schemes introduced in Refs. [114], [115], and [116].

6.3.2 Security against detector side-channel attacks

In this section, we will evaluate ddiQKD security assuming all four detectors active as shown in fig. 6.1(b). In this scenario, one can see from table 6.1 that whenever Bob uses the same measurement basis as Eve there is always two detectors that click. For instance, when $\varphi_{\mathrm{B}} = \phi_{\mathrm{E}} = 0$ the detectors D_1 and D_2 always click, and similar for the other cases. This means that Alice and Bob could, in principle, try to monitor double-clicks to detect the presence of Eve. Here, we show that if the detectors are imperfect, then even though they are trusted and characterized, it is possible for an eavesdropper to break the security of ddiQKD.

Practical single-photon detectors respond differently to the same blinding power P_{B}. This has been recently analyzed in Ref. [159]. There, the authors compare the response of two single-photon detectors in a commercial QKD system Clavis2 [96] to varying blinding power. They first illuminate the detectors with continuous-wave bright light of power P_{B} to force them enter linear-mode operation. Then they record the maximum and minimum value of the trigger pulse energy E_{T} for which the click probabilities are 0 and 1 respectively. The results are shown in fig. 6.2(a). For a particular blinding power P_{B}, each point in the solid (dashed) curves shown in the figure represents the maximum (minimum) value of trigger pulse energy E_{T} for which the detection efficiency η_{det} is 0 (1). The blue and green colors identify the two detectors. (Note that if the energies E_{T} corresponding to the dashed curves are halved, the result is always below the solid curves, thus satisfying the assumption made in section 6.3.1 that pulses with mean photon number $\mu/2$ result in zero click probability.) Next, we show how these detector characteristics could be used to avoid double-clicks.

For this, we return to the blinding attack described in section 6.3.1. For simplicity, let us consider the case where $\varphi_{\mathrm{B}} = \phi_{\mathrm{E}} = 0$ and Eve wants to force a click only on detector say

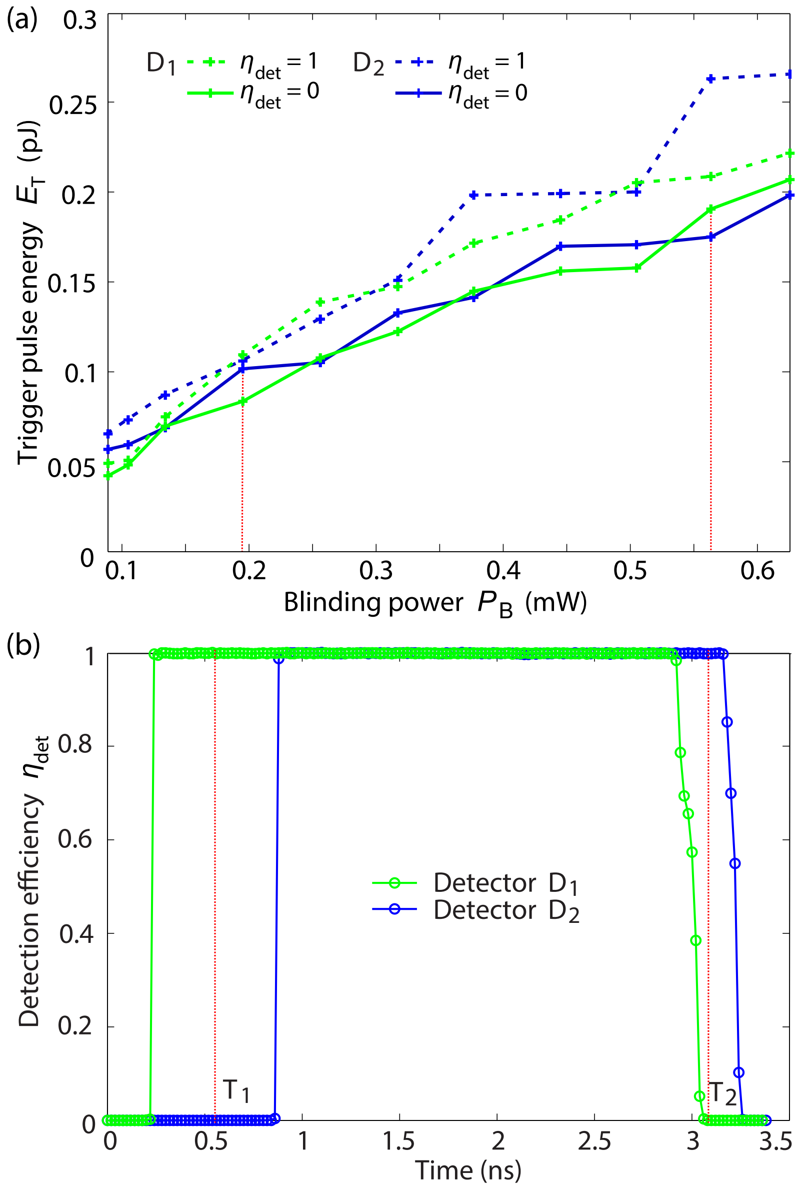


Figure 6.2: Detector click probability in bright-light blinded regime in commercial QKD system Clavis2. (a) Click trigger thresholds versus blinding power P_B for two different single-photon detectors D_1 and D_2. Here, for a particular blinding power P_B, each point in the solid (dashed) curves represents the maximum (minimum) value of trigger pulse energy E_T for which the detection efficiency η_{det} is 0 (1). The experimental data has been reprinted from Ref. [159]. (b) Measured detection efficiency mismatch in the time domain between two blinded single-photon detectors at $P_B = 0.32$ mW, $E_T = 0.24$ pJ, and 0.7 ns wide trigger pulse (see main text for further details). Figure reprinted from [145]

D_1, and no click on detector D_2. Then, she can simply choose a combination of P_B and E_T such that the detector D_1 (D_2) has a non-zero (zero) click probability. If the behavior of the detector D_1 (D_2) corresponds to the green (blue) curves shown in fig. 6.2(a), then the values $P_B \approx 0.2$ mW and $E_T \approx 0.1$ pJ constitute an example that satisfies this criterion. Similarly, if $P_B \approx 0.56$ mW and $E_T \approx 0.19$ pJ, then Eve could make the detector D_2 (D_1) to have a non-zero (zero) click probability. Importantly, note that when Bob's basis matches that of Eve, only two out of the four detectors D_i might produce a click (see table 6.1). Hence, in these instances Eve only needs to avoid double-clicks between two detectors in order to remain undetected. A similar argument can be applied to any other value of φ_B and ϕ_E. This particular attack demonstrates the fact that if Bob's detectors are uncharacterized, as assumed in ddiQKD, this type of schemes are indeed insecure against detector side-channel attacks. That is, Eve could learn the whole secret key without producing any error nor a double-click.

Other imperfections present in the detectors can also allow an eavesdropper to avoid double clicks. For example, if there are efficiency mismatch among the detectors, in either spatial or time domain [79, 70], it is possible for the eavesdropper to avoid the double clicks. Here we provide an example that exploits time-efficiency mismatch among the detectors. Figure 6.2b shows the temporal efficiency mismatch that exists between two detectors in the commercial QKD system Clavis2 [96]. In this case, Eve can perform a time-shift attack (see section 2.3.4). She can shift the arrival time of her signal such that only one detector can produce a click each given time. For instance, whenever Bob receives a trigger pulse at the time instance T_1 (T_2), only the detector D_1 (D_2) can produce a click because this instance is outside of the response region of the detector D_2 (D_1). That is, by combining the time-shift attack with the blinding attack, Eve could avoid double clicks and break the security of ddiQKD without introducing errors.

6.3.3 Side-channel attacks against Bob's linear optics network

One main assumption of ddiQKD is that Bob's linear optics network [*i.e.*, the grey area within Bob's receiver in fig. 6.1b] is fully characterized and trusted. However, this does not necessarily mean that they are perfect, as this would be very difficult to achieve in practice. So, what effect do imperfections have on the security? In this section we investigate this issue and show that Eve could also exploit various typical imperfections in Bob's linear optics to render the system insecure.

As an example of imperfection, we assume that Bob's phase modulator (PM) φ_B is not perfect. We assume that in practice, the PM applies a phase $\varphi_B = \bar{\varphi}_B + \Delta_{\varphi_B}$, where

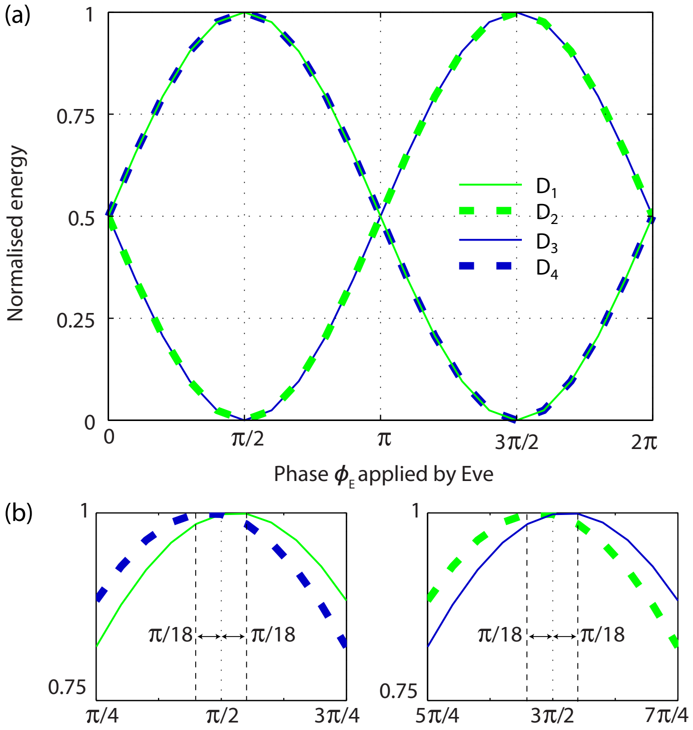


Figure 6.3: Normalized energy at the input ports of Bob's detectors D_i as a function of ϕ_{E}, when $\bar{\varphi}_{\mathrm{B}} = \pi/2$. (a) Ideal scenario with a perfect PM that has $\Delta_{\varphi_{\mathrm{B}}} = 0$. (b) Example of a practical case where $\Delta_{\varphi_{\mathrm{B}}} = \pi/36$ [206]. The normalized energy is defined as the energy divided by the energy of a coherent state with mean photon number μ. See text for further details. Figure reprinted from [145]

$\bar{\varphi}_{\mathrm{B}} \in \{0, \pi/2, \pi, 3\pi/2\}$ and the parameter $\Delta_{\varphi_{\mathrm{B}}}$ characterizes the imperfection. In this scenario, Eve can select her phase $\phi_{\mathrm{E}} = \bar{\phi}_{\mathrm{E}} + \Delta_{\phi_{\mathrm{E}}}$, where $\bar{\phi}_{\mathrm{E}} \in \{0, \pi/2, \pi, 3\pi/2\}$ and $\Delta_{\phi_{\mathrm{E}}} > 0$ is a deviation term that Eve can select to control the detectors. According to eq. (6.1), the energy at the input ports of Bob's detectors D_1, D_2, D_3 and D_4 is proportional to $\frac{\mu}{2}[1+\cos(\phi_{\mathrm{E}} - \varphi_{\mathrm{B}})]$, $\frac{\mu}{2}[1+\cos(\phi_{\mathrm{E}} + \varphi_{\mathrm{B}})]$, $\frac{\mu}{2}[1-\cos(\phi_{\mathrm{E}} - \varphi_{\mathrm{B}})]$, and $\frac{\mu}{2}[1-\cos(\phi_{\mathrm{E}} + \varphi_{\mathrm{B}})]$ respectively. For simplicity, below we focus on the case $\bar{\varphi}_{\mathrm{B}} = \pi/2$. The other cases can be analyzed similarly. We consider first the ideal scenario where $\Delta_{\varphi_{\mathrm{B}}} = 0$. The resulting normalized energies are illustrated in fig. 6.3a as a function of ϕ_{E}. That is, as already seen in section 6.3.1, when Eve's basis matches that of Bob, then two detectors receive maximum energy and, therefore, both click. If Bob and Eve use different bases then the total energy is equally distributed to all the four detectors and, given that E_{T} is chosen carefully, none of them click. Suppose now the practical scenario where Bob's state preparation is imperfect and $\Delta_{\varphi_{\mathrm{B}}}$ is equal to say, for instance, $\pi/36$ (or 5°, which is a typical accuracy in practical systems [206]). In this situation, the energy distributions shift with respect to each other as highlighted in fig. 6.3b. If $\bar{\phi}_{\mathrm{E}} = \pi/2$ and Eve selects say $\Delta_{\phi_{\mathrm{E}}} = \pi/18$ ($\Delta_{\phi_{\mathrm{E}}} = -\pi/18$) then the energy at the input ports of detectors D_1 and D_4 is, respectively, $E_+ \propto 0.998\mu$ and $E_- \propto 0.982\mu$ (E_- and E_+). Similarly, if $\bar{\phi}_{\mathrm{E}} = 3\pi/2$ the energy at the detectors D_2 and D_3 is, respectively, E_- and E_+ (E_+ and E_-). That is, if Eve chooses carefully a suitable value of $\Delta_{\phi_{\mathrm{E}}}$ and μ such that $0.998\mu \geq \mu_{\mathrm{th}}$ and $0.982\mu < \mu_{\mathrm{th}}$, she can guarantee that only one detector clicks each given time, and no double-click is produced.

6.4 Conclusion

In summary, we have analyzed the security of detector-device-independent QKD (ddiQKD) and shown that first, its security is not based on post-selected entanglement, as originally claimed. Secondly and more importantly, we have presented examples of two eavesdropping attacks that demonstrate that ddiQKD is vulnerable to detector side-channel attacks. Finally, we have analyzed the effect of imperfections in Bob's linear optics network on the security and shown that they can lead to a breach of security. We emphasize that these attacks are valid even when Alice's and Bob's state preparation processes are fully characterized and trusted, and Bob's detectors are built by a trusted party.

Whether or not ddiQKD could be made more robust against detector side-channel attacks by including appropriate countermeasures is a different issue. Our main focus in this analysis was to show that ddiQKD alone is not a solution to detector side-channel attacks. That is, in contrast to what has been claimed (*i.e.*, that ddiQKD itself is a countermeasure against any detector side-channel attack), we demonstrate that this is not and ddiQKD, by itself, does not provide the same level of security as mdiQKD. Of course, Alice and Bob might try to prevent these attacks by designing proper countermeasures at the detector side, just like standard QKD schemes. However, if such countermeasures would exist against any known and yet-to-be-known detector side-channel attack (besides mdiQKD), they could probably be used to protect standard QKD protocols as well. In such scenario, it is unclear what would be the real advantage (in terms of complexity and performance) of using ddiQKD instead of standard QKD systems.

Chapter 7

Other projects

In this chapter, I provide a brief summary of the other projects in which I participated. Please see reprints of research papers in appendices for details.

7.1 Invisible Trojan-horse attack

In year 2014, a Trojan-horse attack was attempted at the commercial QKD system Clavis2 running the SARG04 protocol [34] by N. Jain. et. al., [83]. Although the attack was able to extract information out of Bob's phase modulator by sending Trojan photons into it, the increase in QBER due to the resultant afterpulsing noise disclosed the presence of the attack. That attack used Trojan pulses at the operating wavelength $\lambda_s = 1550$ nm at which the employed detectors were highly sensitive. In contrast, we chose to perform the attack with a longer wavelength $\lambda_l = 1924$ nm at which detectors have less sensitivity compared to that at λ_s. Our results show experimentally with detailed numerical modeling that the current attack will succeed in breaking the security and stay inconspicuous unlike the previous attack. We conclude that the invisible nature of the attack poses a threat to the security of practical QKD if proper countermeasures are not adopted. The preprint version of the paper [218] is included in appendix B.

7.2 Finite-key-size effect in commercial plug-and-play QKD system

In order to guarantee unconditional security of practical quantum key distribution systems, every feature, imperfection, and loophole has to be considered and scrutinized. One such feature is the finite-key-size effect. In theory, it is possible to assume that the length of exchanged raw key is infinite and carry out the post processing steps based on that assumption. However, in practice, with limited resources and time, a QKD system can exchange only a finite length of raw key and the knowledge of an adversary about the key has to be estimated from the number of errors in that finite sample [219, 30]. Hence, the smaller the sample is, the less accurate the estimate becomes. Thus, the estimated bound of eavesdropper's knowledge might deviate from the actual value and, if it is underestimated, the key rate formula cannot guarantee security anymore. Thus, in practice, the assumption of infinite key length has to be abandoned and statistical deviations of the finite sample has to be taken into account. This branch of analysis is called 'Finite-key-size analysis' [220, 221, 69, 194] which essentially modifies the key rate equation and incorporate additional terms into the key rate equation due to the finite size of data.

The first rigorous analysis on finite-key-size was published in 2005 and the theory was further developed in the subsequent years [220, 221, 69, 194]. Many of the practical QKD systems used today were developed before that time and as a result the finite-key-size effect was not considered during post processing. Still, if the raw key size was large enough that the finite-key-size effects can be neglected, the generated secret key may still be secure. But this is no longer applicable when the raw key size becomes small enough that the finite-key-size effects can no longer be neglected. We have experimentally demonstrated this in this work [222]. We also demonstrate the ability of an eavesdropper to force a commercial QKD system to generate secret key from a smaller sample size. We further calculate the theoretical finite-key security bounds for the system under test and evaluate the risk based on our experimental data. We also test the manufacturer's patch and evaluate the risk afterwards. The preprint version of the paper [222] is included in appendix C.

7.3 Testing the random-detector-efficiency countermeasure against detector blinding attack

Whenever an implementation loophole is reported, the next step is to design a countermeasure. However, there is no guarantee that the newly designed countermeasures would work as expected. Hence, there is a need for testing the countermeasures. In this work, we tested the random-detector-efficiency countermeasure [223] that was designed by ID Quantique in order to prevent detector blinding attacks. As a third-party tester, we have found that although this countermeasure is effective in preventing the original blinding attack, it fails to guarantee security if the attack is modified slightly. The result is published [159] and included in appendix D.

7.4 Experimental quantum key distribution with source flaw

Most of the existing QKD systems assume perfect state preparation and do not consider the finite-key-size effects in obtaining the secure key rate. In this project, a loss tolerant long distance QKD implementation was demonstrated with the assumption of imperfect state preparations. We quantified the source flaws. We also achieved rigorous finite-key security bounds for decoy-state QKD against coherent attacks. Our results [206] constitute a large step towards secure practical QKD and is included in appendix E.

7.5 Experimental quantum fingerprinting

This project is the first proof-of-principle experimental demonstration of a quantum fingerprinting protocol that is capable of transmitting less information than the best-known classical protocol. The implementation is based on a modified version of a commercial quantum key distribution system using off-the-shelf optical components over telecom wavelengths, and is practical for messages as large as 100 M$bits$, even in the presence of experimental imperfections. Our results provide a first step in the development of experimental quantum communication complexity. The published article [23] is included in appendix F.

Chapter 8

Conclusion

When a new technology promises to supersede or supplement an existing technology, it requires at least one attractive feature that was not possessed by the existing technology. In case of quantum key distribution (QKD), this feature is the unconditional security against an all-powerful eavesdropper. This particular feature has driven researchers from all over the world to work on newer protocols to increase the key rate, make the system more secure, develop new characterization techniques and testing methodology etc. Funds have been poured into the field by industry and governments, and the participation is increasing day by day. All these necessitate rules and regulation for cooperation, coordination, compatibility and trust – in other words standardization and certification.

The standardization and certification process will involve identifying areas of vulnerability, characterizing installed system components for finding deviations in the behavior of practical devices from ideal ones, finding correct testing methodologies for finding the existence of side-channel, evaluating the risk of the side-channel, designing of countermeasure, testing countermeasure performance etc. This is a huge and time-consuming task and the number of areas to examine is very high due to the higher number of practical systems in different variants and forms being deployed around the world. Thus, contributions from independent research groups are highly necessary that motivated me to my research.

My research evolved around searching for loopholes in areas that were not scrutinized before and I was successful in finding them several times. The first project was evaluating the design of the pulse energy monitoring system (PEMS) implemented in a commercial QKD system and performing theoretical analyses to study the effect of mean photon number deviation on the security of three quantum communication protocols. My results showed that existing pulse-energy-monitoring-system design was based on flawed logic and needs redesigning to guarantee security. I further pointed out several new areas that must be tested when certifying a PEMS. This work also highlighted the limitations of closed security standards developed inside a manufacturing company – ID Quantique in this case. Although the company went above and beyond everyone else's prior research in this field to secure their commercial system by implementing the PEMS, it was not sufficient. In this case, as well as in numerous other instances [160, 79, 224, 71, 73, 76, 207, 84, 168], an independent research team – in this case, us – uncovered security problems that the original developers missed. This shows that a different point-of-view – other than that from the developers – are also important for standardization and certification.

I also performed security evaluation of a free space quantum communication system. More specifically, I studied a free-space QKD receiver prototype that was built for long-distance satellite communication and experimentally demonstrated the existence of spatial-mode-detector-efficiency mismatch. We confirmed that the discovered side-channel is exploitable by Eve to compromise security and checked the feasibility of such attacks. Our results identified a methodology for checking the spatial-mode-detector-efficiency mismatch in such QKD receivers, and also showed a simple, implementable countermeasure to block this side-channel. This is a step forward on the certification of these free-space receivers against spatial-mode-detector-efficiency mismatch.

Next, my project involved checking the feasibility of laser damage as a potential tool for eavesdropping. After testing on two different quantum communication systems, we confirmed that laser damage can be used to break the security for both. This result indicates that a characterized and side-channel free system does not necessarily guarantee security for ever, as with significant probability, an eavesdropper can create deviations and side-channels on-demand. Thus, any certification process must ensure that an eavesdropper cannot perform laser damage into the system and must consider other laser-damage related security critical issues for certification.

Finally, I scrutinized the assumptions of detector-device-independent QKD (ddiQKD) protocol which restricted the ability of an eavesdropper. It was not clear how secure the protocol was, had the ability of the eavesdropper been not restricted. We introduced several eavesdropping schemes that showed that the security of ddiQKD cannot be based on post selected entanglement. Our results point out that for the standardization and certification process, it is also important to test the assumptions of the protocols.

Along the way, I also participated in several other projects. I lead the project that performed a Trojan-horse attack on a commercial QKD system with a non-detectable-wavelength and was successful in compromising the security [218]. I participated in testing a detector-control-attack countermeasure to find it effective against the original attack;

but ineffective if the attack is slightly modified [159]. I also participated in projects that analyzed the effects of finite-key-size in a commercial QKD system [222] and imperfect state preparation on the security [206]. I also contributed to the project of experimental demonstration of quantum fingerprinting.

My results outline the importance of scrutinizing practical quantum communication systems for implementation loopholes and the need for standardization and certification. It is also fair to say that standards on security implementation and testing of systems should be done in a collective and cooperative way between the research community and industry. Although, this process has already taken place [137], it can be intensified in the security specifications aspect. Also, practice shows that more often than not, it is an independent research team that is focused on identifying side-channels, spots the security problems instead of the developers. Thus, it would not be a bad idea to hand over the task of testing and identifying unexpected security problems to independent security certification labs which will be their main focus. Unfortunately, no such lab exists at this moment, so initiatives can be taken to set up such labs. Until then, the task can be given to third party researchers that has shown outstanding abilities and great capability at that task.

In summary, during my Ph.D. research, I searched for gaps between theory and implementations. When found, I analyzed them to verify if they are exploitable by an eavesdropper to break the security. For a discovered side-channel, I looked for countermeasures to block it. All these are parts of an iterative process that a new and promising technology must go through before being able to supersede the old technology. Quantum cryptography is currently going through such a transitional stage and my Ph.D. research was an attempt to facilitate the transition. At present, it seems that the field is moving towards the right direction. Hopefully, after few more attempts like this, the state of the QKD will reach to a point where it can be trusted with high confidence and we will get a world with quantum cryptography where information will be truly secured. I want to finish my thesis on this note.

References

[1] A. Kerckhoffs. La cryptographie militaire. *J. des Sciences Militaires*, IX:5–38, January 1883.

[2] G. S. Vernam. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *J. Am. Inst. Electr. Eng.*, 45:109–115, 1926.

[3] S. Singh. *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography.* Fourth Estate, London, 1999.

[4] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26:1484–1509, 1997.

[5] Ralph Charles Merkle. *Secrecy, authentication, and public key systems.* Doctoral thesis, Stanford University, 1979.

[6] R. J. McEliece. A public-key cryptosystem based an algebraic coding theory. *Deep Space Network Progress Report*, 44:114–116, January 1978.

[7] Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. *NTRU: A ring-based public key cryptosystem*, pages 267–288. Springer Berlin Heidelberg, 1998.

[8] Jacques Patarin. *Hidden fields equations (HFE) and isomorphisms of polynomials (IP): two new families of asymmetric algorithms*, pages 33–48. Springer Berlin Heidelberg, 1996.

[9] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74(1):145–195, 2002.

[10] Stephen Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, January 1983.

[11] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)*, pages 175–179, New York, 1984. IEEE Press.

[12] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[13] D. Deutsch. Quantum theory, the church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 400(1818):97–117, 1985.

[14] A. K. Ekert. Quantum Cryptography Based on Bell's Theorem. *Phys. Rev. Lett.*, 67(6):661–663, 1991.

[15] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[16] David Bohm. *Quantum theory.* Englewood Cliffs, N.J. : Prentice-Hall, 1951.

[17] John. S. Bell. On the einstein podolsky rosen paradox. *Physics*, 1:195–200, 1964.

[18] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[19] Anna Pappa, Paul Jouguet, Thomas Lawson, André Chailloux, Matthieu Legré, Patrick Trinkler, Iordanis Kerenidis, and Eleni Diamanti. Experimental plug and play quantum coin flipping. *Nat. Commun.*, 5:3717, 2014.

[20] Guido Berlín, Gilles Brassard, Félix Bussières, Nicolas Godbout, Joshua A. Slater, and Wolfgang Tittel. Experimental loss-tolerant quantum coin flipping. *Nat. Commun.*, 2:561, 2011.

[21] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, Sep 2001.

[22] Juan Miguel Arrazola and Norbert Lütkenhaus. Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A*, 89:062305, Jun 2014.

[23] Feihu Xu, Juan Miguel Arrazola, Keijin Wei, Kenyan Wang, Pablo palacios Avila, Chen Feng, Shihan Sajeed, Norbert Lütkenhaus, and Hoi-Kwong Lo. Experimental quantum fingerprinting with weak coherent pulses. *Nat. Commun.*, 6:8735, 2015.

[24] Jian-Yu Guan, Feihu Xu, Hua-Lei Yin, Yuan Li, Wei-Jun Zhang, Si-Jing Chen, Xiao-Yan Yang, Li Li, Li-Xing You, Teng-Yun Chen, Zhen Wang, Qiang Zhang, and Jian-Wei Pan. Observation of quantum fingerprinting beating the classical limit. *Phys. Rev. Lett.*, 116:240502, Jun 2016.

[25] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993.

[26] Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, and Anton Zeilinger. Experimental quantum teleportation. *Nature*, 390:575–579, 1997.

[27] Xiao-Song Ma, Thomas Herbst, Thomas Scheidl, Daqing Wang, Sebastian Kropatschek, William Naylor, Bernhard Wittmann, Alexandra Mech, Johannes Kofler, Elena Anisimova, Vadim Makarov, Thomas Jennewein, Rupert Ursin, and Anton Zeilinger. Quantum teleportation over 143 kilometres using active feed-forward. *Nature*, 489:269, 2012.

[28] Ji-Gang Ren, Ping Xu, Hai-Lin Yong, Liang Zhang, Sheng-Kai Liao, Juan Yin, Wei-Yue Liu, Wen-Qi Cai, Meng Yang, Li Li, Kui-Xing Yang, Xuan Han, Yong-Qiang Yao, Ji Li, Hai-Yan Wu, Song Wan, Lei Liu, Ding-Quan Liu, Yao-Wu Kuang, Zhi-Ping He, Peng Shang, Cheng Guo, Ru-Hua Zheng, Kai Tian, Zhen-Cai Zhu, Nai-Le Liu, Chao-Yang Lu, Rong Shu, Yu-Ao Chen, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. Ground-to-satellite quantum teleportation. *arXiv:1707.00934*, 2017.

[29] Stefanie Barz, Elham Kashefi, Anne Broadbent, Joseph F. Fitzsimons, Anton Zeilinger, and Philip Walther. Demonstration of blind quantum computing. *Science*, 335:303–308, 2012.

[30] C. H. Bennett, F. Bessette, L. Salvail, G. Brassard, and J. Smolin. Experimental quantum cryptography. *J. Cryptology*, 5:3–28, 1992.

[31] C. H. Bennett. Quantum cryptography using any 2 nonorthogonal states. *Phys. Rev. Lett.*, 68(21):3121–3124, 1992.

[32] H. Bechmann-Pasquinucci and N. Gisin. Incoherent and coherent eavesdropping in the six-state protocol of quantum cryptography. *Phys. Rev. A*, 59:4238–4248, Jun 1999.

[33] D. Bruß. Optimal eavesdropping in quantum cryptography with six states. *Phys. Rev. Lett.*, 81:3018–3021, 1998.

[34] V. Scarani, A. Acín, G. Ribordy, and N. Gisin. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Phys. Rev. Lett.*, 92(5):057901, 2004.

[35] Hua-Lei Yin, Teng-Yun Chen, Zong-Wen Yu, Hui Liu, Li-Xing You, Yi-Heng Zhou, Si-Jing Chen, Yingqiu Mao, Ming-Qi Huang, Wei-Jun Zhang, Hao Chen, Ming Jun Li, Daniel Nolan, Fei Zhou, Xiao Jiang, Zhen Wang, Qiang Zhang, Xiang-Bin Wang, and Jian-Wei Pan. Measurement device independent quantum key distribution over 404 km optical fibre.

[36] Juan Yin, Yuan Cao, Yu-Huai Li, Sheng-Kai Liao, Liang Zhang, Ji-Gang Ren, Wen-Qi Cai, Wei-Yue Liu, Bo Li, Hui Dai, Guang-Bing Li, Qi-Ming Lu, Yun-Hong Gong, Yu Xu, Shuang-Lin Li, Feng-Zhi Li, Ya-Yun Yin, Zi-Qing Jiang, Ming Li, Jian-Jun Jia, Ge Ren, Dong He, Yi-Lin Zhou, Xiao-Xiang Zhang, Na Wang, Xiang Chang, Zhen-Cai Zhu, Nai-Le Liu, Yu-Ao Chen, Chao-Yang Lu, Rong Shu, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356(6343):1140–1144, 2017.

[37] T. C. Ralph. Continuous variable quantum cryptography. *Phys. Rev. A*, 61(1):010303, 1999.

[38] M. Hillery. Quantum cryptography with squeezed states. *Phys. Rev. A*, 61(2):022309, 2000.

[39] M. D. Reid. Quantum cryptography with a predetermined key, using continuous-variable Einstein-Podolsky-Rosen correlations. *Phys. Rev. A*, 62(6):062308, 2000.

[40] N. J. Cerf, M. Lévy, and G. Van Assche. Quantum distribution of gaussian keys using squeezed states. *Phys. Rev. A*, 63:052311, Apr 2001.

[41] Frédéric Grosshans and Philippe Grangier. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.*, 88:057902, Jan 2002.

[42] Frédéric Grosshans, Gilles Van Assche, Jerome Wenger, Rosa Brouri, Nicolas J. Cerf, and Philippe Grangier. Quantum key distribution using gaussian-modulated coherent states. *Nature*, 421:238–241, Jan 2003.

[43] Jérôme Lodewyck, Matthieu Bloch, Raúl García-Patrón, Simon Fossier, Evgueni Karpov, Eleni Diamanti, Thierry Debuisschert, Nicolas J. Cerf, Rosa Tualle-Brouri, Steven W. McLaughlin, and Philippe Grangier. Quantum key distribution over 25 km with an all-fiber continuous-variable system. *Phys. Rev. A*, 76:042305, Oct 2007.

[44] Quyen Dinh Xuan, Zheshen Zhang, and Paul L. Voss. A 24 km fiber-based discretely signaled continuous variable quantum key distribution system. *Opt. Express*, 17(26):24244–24249, 2009.

[45] S. Fossier, E. Diamanti, T. Debuisschert, A. Villing, R. Tualle-Brouri, and P. Grangier. Field test of a continuous-variable quantum key distribution prototype. *New J. Phys.*, 11(4):045023, 2009.

[46] Paul Jouguet, Sébastien Kunz-Jacques, Thierry Debuisschert, Simon Fossier, Eleni Diamanti, Romain Alléaume, Rosa Tualle-Brouri, Philippe Grangier, Anthony Leverrier, Philippe Pache, and Philippe Painchault. Field test of classical symmetric encryption with continuous variables quantum key distribution. *Opt. Express*, 20(13):14030–14041, 2012.

[47] Paul Jouguet, Sébastien Kunz-Jacques, Anthony Leverrier, Philippe Grangier, and Eleni Diamanti. Experimental demonstration of long-distance continuous-variable quantum key distribution. *Nat. Photonics*, 7:378, 2013.

[48] Chao Wang, Duan Huang, Peng Huang, Dakai Lin, Jinye Peng, and Guihua Zeng. 25mhz clock continuous-variable quantum key distribution system over 50km fiber channel. *Scientific Reports*, 5:14607, September 2015.

[49] Duan Huang, Dakai Lin, Chao Wang, Weiqi Liu, Shuanghong Fang, Jinye Peng, Peng Huang, and Guihua Zeng. Continuous-variable quantum key distribution with 1 mbps secure key rate. *Opt. Express*, 23(13):17511–17519, 2015.

[50] Duan Huang, Peng Huang, Dakai Lin, and Guihua Zeng. Long-distance continuous-variable quantum key distribution by controlling excess noise. *Scientific Reports*, 6:19201, January 2016.

[51] K. Inoue, E. Waks, and Y. Yamamoto. Differential phase shift quantum key distribution. *Phys. Rev. Lett.*, 89(3):037902, 2002.

[52] K. Inoue, E. Waks, and Y. Yamamoto. Differential-phase-shift quantum key distribution using coherent light. *Phys. Rev. A*, 68(2):022317, 2003.

[53] H. Takesue, E. Diamanti, T. Honjo, C. Langrock, M. M. Fejer, K. Inoue, and Y. Yamamoto. Differential phase shift quantum key distribution experiment over 105 km fibre. *New J. Phys.*, 7(1):232, 2005.

[54] E. Diamanti, H. Takesue, C. Langrock, M. M. Fejer, and Y. Yamamoto. 100 km differential phase shift quantum key distribution experiment with low jitter up-conversion detectors. *Opt. Express*, 14(26):13073–13082, 2006.

[55] H. Takesue, S. W. Nam, Q. Zhang, R. H. Hadfield, T. Honjo, K. Tamaki, and Y. Yamamoto. Quantum key distribution over a 40-dB channel loss using superconducting single-photon detectors. *Nat. Photonics*, 1(6):343–348, 2007.

[56] N. Gisin, G. Ribordy, H. Zbinden, D. Stucki, N. Brunner, and V. Scarani. Towards practical and fast quantum cryptography. *arXiv:0411022v1*, 2004.

[57] D. Stucki, C. Barreiro, S. Fasel, J.-D. Gautier, O. Gay, N. Gisin, R. Thew, Y. Thoma, P. Trinkler, F. Vannel, and H. Zbinden. High speed coherent one-way quantum key distribution prototype. *arXiv:0809.5264*, 2008.

[58] D. Stucki, C. Barreiro, S. Fasel, J.-D. Gautier, O. Gay, N. Gisin, R. Thew, Y. Thoma, P. Trinkler, F. Vannel, and H. Zbinden. Continuous high speed coherent one-way quantum key distribution. *Opt. Express*, 17(16):13326–13334, 2009.

[59] N. Walenta, A. Burg, D. Caselunghe, J. Constantin, N. Gisin, O. Guinnard, R. Houlmann, P. Junod, B. Korzh, N. Kulesza, M. Legré, C. W. Lim, T. Lunghi, L. Monat, C. Portmann, M. Soucarros, R. T. Thew, P. Trinkler, G. Trolliet, F. Vannel, and H. Zbinden. A fast and versatile quantum key distribution system with hardware key distillation and wavelength multiplexing. *New J. Phys.*, 16:013047, 2014.

[60] Clavis3 specification sheet, `http://marketing.idquantique.com/acton/attachment/11868/f-00f9/1/-/-/-/-/Clavis_3_Datasheet_260116.pdf`, visited 11 May 2017.

[61] D. Mayers. Advances in cryptology. In N. Koblitz, editor, *Proceedings of Crypto'96*, volume 1109, pages 343–357. Springer, New York, 1996.

[62] Eli Biham and Tal Mor. Security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 78:2256–2259, Mar 1997.

[63] H.-K. Lo and H. F. Chau. Unconditional security of quantum key distribution over arbitrarily long distances. *Science*, 283(5410):2050–2056, 1999.

[64] P. W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85(2):441–444, 2000.

[65] N. Lütkenhaus. Security against individual attacks for realistic quantum key distribution. *Phys. Rev. A*, 61(5):052304, 2000.

[66] B. Kraus, N. Gisin, and R. Renner. Lower and upper bounds on the secret-key rate for quantum key distribution protocols using one-way classical communication. *Phys. Rev. Lett.*, 95:080501, Aug 2005.

[67] R. Renner, N. Gisin, and B. Kraus. Information-theoretic security proof for quantum-key-distribution protocols. *Phys. Rev. A*, 72(1):012332, 2005.

[68] R. Renner and R. Koenig. Universally composable privacy amplification against quantum adversaries. In J. Kilian, editor, *Second Theory of Cryptography Conference, TCC 2005*, volume 3378 of *LNCS*, pages 407–425. Springer Verlag, Berlin, February 2005.

[69] R. Y. Q. Cai and V. Scarani. Finite-key analysis for practical implementations of quantum key distribution. *New J. Phys.*, 11:045024, 2009.

[70] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma. Time-shift attack in practical quantum cryptosystems. *Quantum Inf. Comput.*, 7(1-2):73–82, 2007.

[71] Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K. Lo. Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems. *Phys. Rev. A*, 78(4):042333, 2008.

[72] V. Makarov. Controlling passively quenched single photon detectors by bright light. *New J. Phys.*, 11(6):065003, 2009.

[73] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nat. Photonics*, 4:686–689, 2010.

[74] L. Lydersen, J. Skaar, and V. Makarov. Tailored bright illumination attack on distributed-phase-reference protocols. *J. Mod. Opt.*, 58(8):680–685, 2011.

[75] L. Lydersen, M. K. Akhlaghi, A. H. Majedi, J. Skaar, and V. Makarov. Controlling a superconducting nanowire single-photon detector using tailored bright illumination. *New J. Phys.*, 13:113042, 2011.

[76] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs. After-gate attack on a quantum cryptosystem. *New J. Phys.*, 13:013043, 2011.

[77] H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauerth, and H. Weinfurter. Quantum eavesdropping without interception: an attack exploiting the dead time of single-photon detectors. *New J. Phys.*, 13:073024, 2011.

[78] F. Xu, B. Qi, and H.-K. Lo. Experimental demonstration of phase-remapping attack in a practical quantum key distribution system. *New J. Phys.*, 12:113026, 2010.

[79] V. Makarov, A. Anisimov, and J. Skaar. Effects of detector efficiency mismatch on security of quantum cryptosystems. *Phys. Rev. A*, 74(2):022313, 2006. erratum ibid. **78**, 019905 (2008).

[80] V. Makarov and J. Skaar. Faked states attack using detector efficiency mismatch on SARG04, phase-time, DPSK, and Ekert protocols. *Quantum Inf. Comput.*, 8:622–635, 2008.

[81] A. Vakhitov, V. Makarov, and D. R. Hjelme. Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography. *J. Mod. Opt.*, 48(13):2023–2038, 2001.

[82] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy. Trojan-horse attacks on quantum-key-distribution systems. *Phys. Rev. A*, 73(2):022320, 2006.

[83] Nitin Jain, Elena Anisimova, Imran Khan, Vadim Makarov, Christoph Marquardt, and Gerd Leuchs. Trojan-horse attacks threaten the security of practical quantum cryptography. *New J. Phys.*, 16:123030, 2014.

[84] N. Jain, C. Wittmann, L. Lydersen, C. Wiechers, D. Elser, C. Marquardt, V. Makarov, and G. Leuchs. Device calibration impacts security of quantum key distribution. *Phys. Rev. Lett.*, 107:110501, 2011.

[85] B. Huttner, N. Imoto, N. Gisin, and T. Mor. Quantum cryptography with coherent states. *Phys. Rev. A*, 51:1863–1869, Mar 1995.

[86] N. Lütkenhaus. Quantum key distribution with realistic states: photon-number statistics in the photon-number splitting attack. *New J. Phys.*, 4(5):44, 2002.

[87] M. Dušek, M. Jahma, and N. Lütkenhaus. Unambiguous state discrimination in quantum cryptography with weak coherent states. *Phys. Rev. A*, 62(2):022306, Jul 2000.

[88] S. Félix, N. Gisin, A. Stefanov, and H. Zbinden. Faint laser quantum key distribution: eavesdropping exploiting multiphoton pulses. *J. Mod. Opt.*, 48:2009–2021, 2001.

[89] Marcos Curty and Norbert Lütkenhaus. Intercept-resend attacks in the bennett-brassard 1984 quantum-key-distribution protocol with weak coherent pulses. *Phys. Rev. A*, 71:062301, Jun 2005.

[90] M. Curty, L.-L. Zhang, H.-K. Lo, and N. Lütkenhaus. Sequential attacks against differential-phase-shift quantum key distribution with weak coherent states. *Quantum Inf. Comput.*, 7(7):665–688, 2007.

[91] Vadim Makarov, Jean-Philippe Bourgoin, Poompong Chaiwongkhot, Mathieu Gagné, Thomas Jennewein, Sarah Kaiser, Raman Kashyap, Matthieu Legré, Carter Minshull, and Shihan Sajeed. Creation of backdoors in quantum communications via laser damage. *Phys. Rev. A*, 94:030302, 2016.

[92] Agnes Ferenczi, Philippe Grangier, and Frederic Grosshans. Calibration attack and defense in continuous variable quantum key distribution. In *CLEO/Europe and IQEC 2007 Conference Digest*, page IC13. Optical Society of America, 2007.

[93] Hauke Häseler, Tobias Moroder, and Norbert Lütkenhaus. Testing quantum devices: Practical entanglement verification in bipartite optical systems. *Phys. Rev. A*, 77:032303, Mar 2008.

[94] Paul Jouguet, Sébastien Kunz-Jacques, and Eleni Diamanti. Preventing calibration attacks on the local oscillator in continuous-variable quantum key distribution. *Phys. Rev. A*, 87:062313, 2013.

[95] Xiang-Chun Ma, Shi-Hai Sun, Mu-Sheng Jiang, and Lin-Mei Liang. Local oscillator fluctuation opens a loophole for eve in practical continuous-variable quantum-key-distribution systems. *Phys. Rev. A*, 88:022339, Aug 2013.

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

[97] W.-Y. Hwang. Quantum key distribution with high loss: Toward global secure communication. *Phys. Rev. Lett.*, 91(5):057901, 2003.

[98] X.-B. Wang. Beating the photon-number-splitting attack in practical quantum cryptography. *Phys. Rev. Lett.*, 94(23):230503, 2005.

[99] H.-K. Lo, X. Ma, and K. Chen. Decoy state quantum key distribution. *Phys. Rev. Lett.*, 94(23):230504, 2005.

[100] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill. Security of quantum key distribution with imperfect devices. *Quantum Inf. Comput.*, 4(5):325–360, 2004.

[101] C.-H. F. Fung, K. Tamaki, B. Qi, H.-K. Lo, and X. Ma. Security proof of quantum key distribution with detection efficiency mismatch. *Quantum Inf. Comput.*, 9(1 & 2):131–165, 2009.

[102] Ø. Marøy, L. Lydersen, and J. Skaar. Security of quantum key distribution with arbitrary individual imperfections. *Phys. Rev. A*, 82:032337, 2010.

[103] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proc. 39th Annual Symposium on Foundations of Computer Science*, pages 503–509. IEEE, 1998.

[104] A. Acín, N. Gisin, and L. Masanes. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.*, 97:120405, 2006.

[105] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, 2007.

[106] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Phys. Rev. Lett.*, 113:140501, 2014.

[107] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575):682–686, 10 2015.

[108] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel,

Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla Lambrocco, Edward Tortorici, Alan L. Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam. Strong loophole-free test of local realism. *Phys. Rev. Lett.*, 115:250402, 2015.

[109] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger. Significant-loophole-free test of bell's theorem with entangled photons. *Phys. Rev. Lett.*, 115:250401, 2015.

[110] H.-K. Lo, M. Curty, and B. Qi. Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.*, 108:130503, 2012.

[111] Eli Biham, Bruno Huttner, and Tal Mor. Quantum cryptographic network based on quantum memories. *Phys. Rev. A*, 54:2651–2658, 1996.

[112] H. Inamori. Security of Practical Time-Reversed EPR Quantum Key Distribution. *Algorithmica*, 34:340–365, 2002.

[113] Charles Ci Wen Lim, Boris Korzh, Anthony Martin, Félix Bussières, Rob Thew, and Hugo Zbinden. Detector-device-independent quantum key distribution. *Appl. Phys. Lett.*, 105:221112, 2014.

[114] P. González, L. Rebón, T. Ferreira da Silva, M. Figueroa, C. Saavedra, M. Curty, G. Lima, G. B. Xavier, and W. A. T. Nogueira. Quantum key distribution with untrusted detectors. *Phys. Rev. A*, 92:022337, 2015.

[115] Wen-Fei Cao, Yi-Zheng Zhen, Yu-Lin Zheng, Zeng-Bing Chen, Nai-Le Liu, Kai Chen, and Jian-Wei Pan. Highly efficient quantum key distribution immune to all detector attacks. *arXiv:1410.2928v1*, 2014. (manuscript withdrawn by authors on 23 Aug 2016 owing to the insecurity of the proposed scheme).

[116] Wen-Ye Liang, Mo Li, Zhen-Qiang Yin, Wei Chen, Shuang Wang, Xue-Bi An, Guang-Can Guo, and Zheng-Fu Han. Simple implementation of quantum key distribution based on single-photon bell-state measurement. *Phys. Rev. A*, 92:012319, 2015.

[117] Yoon-Ho Kim. Single-photon two-qubit entangled states: Preparation and measurement. *Phys. Rev. A*, 67:040301, 2003.

[118] T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. C. do Amaral, G. P. Temporão, and J. P. von der Weid. Proof-of-principle demonstration of measurement-device-independent quantum key distribution using polarization qubits. *Phys. Rev. A*, 88:052303, 2013.

[119] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel. Real-world two-photon interference and proof-of-principle quantum key distribution immune to detector attacks. *Phys. Rev. Lett.*, 111:130501, 2013.

[120] Yang Liu, Teng-Yun Chen, Liu-Jun Wang, Hao Liang, Guo-Liang Shentu, Jian Wang, Ke Cui, Hua-Lei Yin, Nai-Le Liu, Li Li, Xiongfeng Ma, Jason S. Pelc, M. M. Fejer, Cheng-Zhi Peng, Qiang Zhang, and Jian-Wei Pan. Experimental measurement-device-independent quantum key distribution. *Phys. Rev. Lett.*, 111:130502, 2013.

[121] Yan-Lin Tang, Hua-Lei Yin, Si-Jing Chen, Yang Liu, Wei-Jun Zhang, Xiao Jiang, Lu Zhang, Jian Wang, Li-Xing You, Jian-Yu Guan, Dong-Xu Yang, Zhen Wang, Hao Liang, Zhen Zhang, Nan Zhou, Xiongfeng Ma, Teng-Yun Chen, Qiang Zhang, and Jian-Wei Pan. Measurement-device-independent quantum key distribution over 200 km. *Phys. Rev. Lett.*, 113:190501, 2014.

[122] Raju Valivarthi, Itzel Lucio-Martinez, Philip Chan, Allison Rubenok, Caleb John, Daniel Korchinski, Cooper Duffin, Francesco Marsili, Varun Verma, Mathew D. Shaw, Jeffrey A. Stern, Sae Woo Nam, Daniel Oblak, Qiang Zhou, Joshua A. Slater, and Wolfgang Tittel. Measurement-device-independent quantum key distribution: from idea towards application. *Journal of Modern Optics*, 62(14):1141–1150, 2015.

[123] Yan-Lin Tang, Hua-Lei Yin, Qi Zhao, Hui Liu, Xiang-Xiang Sun, Ming-Qi Huang, Wei-Jun Zhang, Si-Jing Chen, Lu Zhang, Li-Xing You, Zhen Wang, Yang Liu, Chao-Yang Lu, Xiao Jiang, Xiongfeng Ma, Qiang Zhang, Teng-Yun Chen, and Jian-Wei Pan. Measurement-device-independent quantum key distribution over untrustful metropolitan network. *Phys. Rev. X*, 6:011024, 2016.

[124] Guang-Zhao Tang, Shi-Hai Sun, Feihu Xu, Huan Chen, Chun-Yan Li, and Lin-Mei Liang. Experimental asymmetric plug-and-play measurement-device-independent quantum key distribution. *Phys. Rev. A*, 94:032326, Sep 2016.

[125] Stefano Pirandola, Carlo Ottaviani, Gaetana Spedalieri, Christian Weedbrook, Samuel L Braunstein, Seth Lloyd, Tobias Gehring, Christian S Jacobsen, and Ulrik L Andersen. High-rate measurement-device-independent quantum cryptography. *Nat. Photonics*, 9:397–402, 2015.

[126] L C Comandar, M Lucamarini, B Fröhlich, J F Dynes, A W Sharpe, S W-B Tam, Z L Yuan, R V Penty, and A J Shields. Quantum cryptography without detector vulnerabilities using optically-seeded lasers. *Nat. Photonics*, 10:312–315, 2016.

[127] C. Erven, N. Ng, N. Gigov, R. Laflamme, S. Wehner, and G. Weihs. An experimental implementation of oblivious transfer in the noisy storage model. *Nat. Commun.*, 5:3418, 2014.

[128] http://www.idquantique.com/about-idq/company-profile/, visited 12 July 2017.

[129] Quantis specification sheet, http://marketing.idquantique.com/acton/attachment/11868/f-0174/1/-/-/-/-/2016%2008%2024%20-%20QRNG%20paper.pdf, visited 12 July 2017.

[130] ID Quantique Cerberis datasheet: http://www.idquantique.com/images/stories/PDF/cerberis-encryptor/cerberis-specs.pdf, visited 19 April 2011.

[131] Centauris specification sheet, http://marketing.idquantique.com/acton/attachment/11868/f-0100/1/-/-/-/-/Centauris-Datasheet.pdf, visited 12 July 2017.

[132] M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. F. Dynes, S. Fasel, S. Fossier, M. Fürst, J.-D. Gautier, O. Gay, N. Gisin, P. Grangier, A. Happe, Y. Hasani, M. Hentschel, H. Hübel, G. Humer, T. Länger, M. Legré, R. Lieger, J. Lodewyck, T. Lorünser, N Lütkenhaus, A. Marhold, T. Matyus, O. Maurhart, L. Monat, S. Nauerth, J.-B. Page, A. Poppe, E. Querasser, G. Ribordy, S. Robyr, L. Salvail, A. W. Sharpe, A. J. Shields, D. Stucki, M. Suda, C. Tamas, T. Themel, R. T. Thew, Y. Thoma, A. Treiber, P. Trinkler, R. Tualle-Brouri, F. Vannel, N. Walenta, H. Weier, H. Weinfurter, I. Wimberger, Z. L. Yuan, H. Zbinden, and A. Zeilinger. The SECOQC quantum key distribution network in Vienna. *New J. Phys.*, 11(7):075001, 2009.

[133] Sheng-Kai Liao, Wen-Qi Cai, Wei-Yue Liu, Liang Zhang, Yang Li, Ji-Gang Ren, Juan Yin, Qi Shen, Yuan Cao, Zheng-Ping Li, Feng-Zhi Li, Xia-Wei Chen, Li-Hua

Sun, Jian-Jun Jia, Jin-Cai Wu, Xiao-Jun Jiang, Jian-Feng Wang, Yong-Mei Huang, Qiang Wang, Yi-Lin Zhou, Lei Deng, Tao Xi, Lu Ma, Tai Hu, Qiang Zhang, Yu-Ao Chen, Nai-Le Liu, Xiang-Bin Wang, Zhen-Cai Zhu, Chao-Yang Lu, Rong Shu, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. Satellite-to-ground quantum key distribution. *arXiv: 1707.00542*, 2017.

[134] Eleni Diamanti, Hoi-Kwong Lo, Bing Qi, and Zhiliang Yuan. Practical challenges in quantum key distribution. *Npj Quantum Inf.*, 2:16025EP–, 2016.

[135] Elizabeth Gibney. Europe's billion-euro quantum project takes shape. *Nature*, 545, 2017.

[136] http://pulsenews.co.kr/view.php?year=2017&no=412607,visited 12 July 2017.

[137] Thomas Länger and Gaby Lenhart. Standardization of quantum key distribution and the ETSI standardization initiative ISG-QKD. *New J. Phys.*, 11:055051, 2009.

[138] ETSI white paper, http://www.etsi.org/images/files/ETSIWhitePapers/QuantumSafeWhitepaper.pdf, visited 12 July 2017.

[139] J.-P. Bourgoin, E. Meyer-Scott, B. L. Higgins, B. Helou, C. Erven, H. Hübel, B. Kumar, D. Hudson, I. D'Souza, R. Girard, R. Laflamme, and T. Jennewein. A comprehensive design and performance analysis of low earth orbit satellite quantum communication. *New J. Phys.*, 15:023006, 2013.

[140] Jean-Philippe Bourgoin, Nikolay Gigov, Brendon L. Higgins, Zhizhong Yan, Evan Meyer-Scott, Amir K. Khandani, Norbert Lütkenhaus, and Thomas Jennewein. Experimental quantum key distribution with simulated ground-to-satellite photon losses and processing limitations. *Phys. Rev. A*, 92:052339, 2015.

[141] Audun Nystad Bugge, Sebastien Sauge, Aina Mardhiyah M. Ghazali, Johannes Skaar, Lars Lydersen, and Vadim Makarov. Laser damage helps the eavesdropper in quantum cryptography. *Phys. Rev. Lett.*, 112:070503, Feb 2014.

[142] Shihan Sajeed, Igor Radchenko, Sarah Kaiser, Jean-Philippe Bourgoin, Anna Pappa, Laurent Monat, Matthieu Legré, and Vadim Makarov. Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing. *Phys. Rev. A*, 91:032326, 2015.

[143] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov. Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch. *Phys. Rev. A*, 91:062301, 2015.

[144] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden. Quantum key distribution over 67 km with a plug&play system. *New J. Phys.*, 4:41, 2002.

[145] S. Sajeed, A. Huang, S. Sun, F. Xu, V. makarov, and M. Curty. Insecurity of detector-device-independent quantum key distribution. *Phys. Rev. Lett.*, 117:250505, 2016.

[146] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev. The security of practical quantum key distribution. *Rev. Mod. Phys.*, 81(3):1301–1350, 2009.

[147] C. Branciard, N. Gisin, B. Kraus, and V. Scarani. Security of two quantum cryptography protocols using the same four qubit states. *Phys. Rev. A*, 72(3):032301, 2005.

[148] H. Inamori, N. Lütkenhaus, and D. Mayers. Unconditional security of practical quantum key distribution. *Eur. Phys. J. D*, 41:599–627, 2007.

[149] C. A. Fuchs, N. Gisin, R. B. Griffiths, C. S. Niu, and A. Peres. Optimal eavesdropping in quantum cryptography. 1. Information bound and optimal strategy. *Phys. Rev. A*, 56(2):1163–1172, 1997.

[150] V. Makarov and D. R. Hjelme. Faked states attack on quantum cryptosystems. *J. Mod. Opt.*, 52:691–705, 2005.

[151] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov. Full-field implementation of a perfect eavesdropper on a quantum cryptography system. *Nat. Commun.*, 2:349, 2011.

[152] M. Lucamarini, I. Choi, M. B. Ward, J. F. Dynes, Z. L. Yuan, and A. J. Shields. Practical security bounds against the trojan-horse attack in quantum key distribution. *Phys. Rev. X*, 5:031030, Sep 2015.

[153] Nitin Jain, Birgit Stiller, Imran Khan, Vadim Makarov, Christoph Marquardt, and Gerd Leuchs. Risk analysis of Trojan-horse attacks on practical quantum key distribution systems. *arXiv:1408.0492*, 2014.

[154] B. Stiller, I. Khan, N. Jain, P. Jouguet, S. Kunz-Jacques, E. Diamanti, C. Marquardt, and G. Leuchs. Quantum hacking of continuous-variable quantum key distribution systems: Realtime trojan-horse attacks. In *2015 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2, May 2015.

[155] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa. Evolution and prospects for single-photon avalanche diodes and quenching circuits. *J. Mod. Opt.*, 51(9):1267–1288, 2004.

[156] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov. Thermal blinding of gated detectors in quantum cryptography. *Opt. Express*, 18:27938–27954, 2010.

[157] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov. Controlling an actively-quenched single photon detector with bright light. *Opt. Express*, 19:23590–23600, 2011.

[158] Jonathan Jogenfors, Ashraf Mohamed Elhassan, Johan Ahrens, Mohamed Bourennane, and Jan-Åke Larsson. Hacking the Bell test using classical light in energy-time entanglement-based quantum key distribution. *Sci. Adv.*, 1:e1500793, 2015.

[159] Anqi Huang, Shihan Sajeed, Poompong Chaiwongkhot, Mathilde Soucarros, Matthieu Legré, and Vadim Makarov. Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption. *IEEE Journal of Quantum Electronics*, 52(11), 2016.

[160] N. Lütkenhaus. Estimates for practical quantum cryptography. *Phys. Rev. A*, 59(5):3301–3319, 1999.

[161] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo. Practical decoy state for quantum key distribution. *Phys. Rev. A*, 72:012326, 2005.

[162] Duan Huang, Peng Huang, Tao Wang, Huasheng Li, Yingming Zhou, and Guihua Zeng. Continuous-variable quantum key distribution based on a plug-and-play dual-phase-modulated coherent-states protocol. *Phys. Rev. A*, 94:032305, Sep 2016.

[163] Yujun Choi, Osung Kwon, Minki Woo, Kyunghwan Oh, Sang-Wook Han, Yong-Su Kim, and Sung Moon. Plug-and-play measurement-device-independent quantum key distribution. *Phys. Rev. A*, 93:032319, Mar 2016.

[164] I. V. Radchenko, K. S. Kravtsov, S. P. Kulik, and S. N. Molotkov. Relativistic quantum cryptography. *Laser. Phys. Lett.*, 11:065203, 2014.

[165] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin. “Plug and play” systems for quantum cryptography. *Appl. Phys. Lett.*, 70(7):793–795, 1997.

[166] Clavis2 specification sheet, http://www.idquantique.com.

[167] Texas Instruments OPA380 precision, high-speed transimpedance amplifier, http://www.ti.com/lit/ds/symlink/opa380.pdf, visited 5 May 2014.

[168] Agnes Ferenczi, Varun Narasimhachar, and Norbert Lütkenhaus. Security proof of the unbalanced phase-encoded bennett-brassard 1984 protocol. *Phys. Rev. A*, 86:042327, 2012.

[169] Marcos Curty and Norbert Lütkenhaus. Effect of finite detector efficiencies on the security evaluation of quantum key distribution. *Phys. Rev. A*, 69:042321, 2004.

[170] Armand Niederberger, Valerio Scarani, and Nicolas Gisin. Photon-number-splitting versus cloning attacks in practical implementations of the bennett-brassard 1984 protocol for quantum cryptography. *Phys. Rev. A*, 71:042316, Apr 2005.

[171] O. Gittsovich, N. J. Beaudry, V. Narasimhachar, R. Romero Alvarez, T. Moroder, and N. Lütkenhaus. Squashing model for detectors and applications to quantum-key-distribution protocols. *Phys. Rev. A*, 89:012325, 2014.

[172] Agiltron, Inc. FC-1-5-1-1-9-1-1-2 fiberoptic coupler/splitter, http://www.agiltron.com/PDFs/fiberopticwidebandcoupler1310-1550nm.pdf, visited 5 May 2014.

[173] Boston Applied Technologies, Inc. Nanona high speed & low loss optical switch, http://www.bostonati.com/Products_Nanona.html, visited 5 May 2014.

[174] Agiltron, Inc. NanoSpeed 1x2 fiberoptic switch NSSW-11-5-1-1-1-1-1-2, http://www.agiltron.com/PDFs/NS1x2switch-B.pdf, visited 5 May 2014.

[175] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam. Detecting single infrared photons with 93% system efficiency. *Nat. Photonics*, 7:210, 2013.

[176] Photon Spot, http://www.photonspot.com, visited 25 October 2014.

[177] Guido Berlín, Gilles Brassard, Félix Bussières, and Nicolas Godbout. Fair loss-tolerant quantum coin flipping. *Phys. Rev. A*, 80:062321, 2009.

[178] A. Pappa, A. Chailloux, E. Diamanti, and I. Kerenidis. Practical quantum coin flipping. *Phys. Rev. A*, 84:052305, 2011.

[179] J. D. Franson and H. Ilves. Quantum cryptography using optical fibers. *Appl. Opt.*, 33(14):2949–2954, May 1994.

[180] Thomas Jennewein, Christoph Simon, Gregor Weihs, Harald Weinfurter, and Anton Zeilinger. Quantum cryptography with entangled photons. *Phys. Rev. Lett.*, 84:4729–4732, May 2000.

[181] Y. Liu, T.-Y. Chen, J. Wang, W.-Q. Cai, X. Wan, L.-K. Chen, J.-H. Wang, S.-B. Liu, H. Liang, L. Yang, C.-Z. Peng, K. Chen, Z.-B. Chen, and J.-W. Pan. Decoy-state quantum key distribution with polarized photons over 200 km. *Opt. Express*, 18(8):8587–8594, 2010.

[182] Y. Zhao, B. Qi, and H.-K. Lo. Experimental quantum key distribution with active phase randomization. *Appl. Phys. Lett.*, 90(4):044106, 2007.

[183] D. Stucki, N. Walenta, F. Vannel, R. T. Thew, N. Gisin, H. Zbinden, S. Gray, C. R. Towery, and S. Ten. High rate, long-distance quantum key distribution over 250 km of ultra low loss fibres. *New J. Phys.*, 11(7):075003, 2009.

[184] MagiQ QPN 5505 security gateway quantum key distribution system, specification sheet, http://web.archive.org/web/20050209013643/http://www.magiqtech.com/press/qpn.pdf, visited 31 March 2010.

[185] N. J. Beaudry, T. Moroder, and N. Lütkenhaus. Squashing models for optical measurements in quantum communication. *Phys. Rev. Lett.*, 101(9):093601, 2008.

[186] T. Tsurumaru and K. Tamaki. Security proof for quantum-key-distribution systems with threshold detectors. *Phys. Rev. A*, 78(3):032302, 2008.

[187] W. T. Buttler, R. J. Hughes, P. G. Kwiat, S. K. Lamoreaux, G. G. Luther, G. L. Morgan, J. E. Nordholt, C. G. Peterson, and C. M. Simmons. Practical free-space quantum key distribution over 1 km. *Phys. Rev. Lett.*, 81:3283, 1998.

[188] C. Kurtsiefer, P. Zarda, M. Halder, H. Weinfurter, P. M. Gorman, P. R. Tapster, and J. G. Rarity. Quantum cryptography: A step towards global key distribution. *Nature*, 419(6906):450–450, 2002.

[189] C. Kurtsiefer, P. Zarda, M. Halder, P. M. Gorman, P. R. Tapster, J. G. Rarity, and H. Weinfurter. Long distance free space quantum cryptography. *Proc. SPIE*, 4917:25–31, 2002.

[190] R. J. Hughes, J. E. Nordholt, D. Derkacs, and C. G. Peterson. Practical free-space quantum key distribution over 10 km in daylight and at night. *New J. Phys.*, 4:43, 2002.

[191] R. Ursin, F. Tiefenbacher, T. Schmitt-Manderbach, H. Weier, T. Scheidl, M. Lindenthal, B. Blauensteiner, T. Jennewein, J. Perdigues, P. Trojek, B. Ömer, M. Fürst, M. Meyenburg, J. Rarity, Z. Sodnik, C. Barbieri, H. Weinfurter, and A. Zeilinger. Entanglement-based quantum communication over 144 km. *Nat. Phys.*, 3(7):481–486, 2007.

[192] C. Erven, C. Couteau, R. Laflamme, and G.Weihs. Entangled quantum key distribution over two free-space optical links. *Opt. Express*, 16:16840–16853, 2008.

[193] Sebastian Nauerth, Florian Moll, Markus Rau, Christian Fuchs, Joachim Horwath, Stefan Frick, and Harald Weinfurter. Air-to-ground quantum communication. *Nat. Photonics*, 7:382, 2013.

[194] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner. Tight finite-key analysis for quantum cryptography. *Nat. Commun.*, 3:634, 2012.

[195] Henning Weier, Tobias Schmitt-Manderbach, Nadja Regner, Christian Kurtsiefer, and Harald Weinfurter. Free space quantum key distribution: Towards a real life application. *Fortschr. Phys.*, 54:840, 2006.

[196] M. P. Peloso, I. Gerhardt, C. Ho, A. Lamas-Linares, and C. Kurtsiefer. Daylight operation of a free space, entanglement-based quantum key distribution system. *New J. Phys.*, 11(4):045007, 2009.

[197] Robert Tyson. *Principles of Adaptive Optics*. CRC Press, 3rd edition, 2010.

[198] J. C. Bienfang, A. J. Gross, A. Mink, B. J. Hershman, A. Nakassis, X. Tang, R. Lu, D. H. Su, Charles W. Clark, and Carl J. Williams. Quantum key distribution with 1.25 Gbps clock synchronization. *Opt. Express*, 12:2011, 2004.

[199] Hideki Takenaka, Morio Toyoshima, and Yoshihisa Takayama. Experimental verification of fiber-coupling efficiency for satellite-to-ground atmospheric laser downlinks. *Opt. Express*, 20:15301, 2012.

[200] Thiago Ferreira da Silva, Gustavo C. do Amaral, Guilherme B. Xavier, Guilherme P. Temporão, and Jean Pierre von der Weid. Safeguarding quantum key distribution through detection randomization. *IEEE J. Sel. Top. Quantum Electron.*, 21:1–9, 2015.

[201] A. N. Bugge. Controlled laser damage of single-photon avalanche photodiodes. Master's thesis, Norwegian University of Science and Technology, 2012.

[202] Robert J. Collins, Ross J. Donaldson, Vedran Dunjko, Petros Wallden, Patrick J. Clarke, Erika Andersson, John Jeffers, and Gerald S. Buller. Realization of quantum digital signatures without the requirement of quantum memory. *Phys. Rev. Lett.*, 113:040502, 2014.

[203] T. Lunghi, J. Kaniewski, F. Bussières, R. Houlmann, M. Tomamichel, A. Kent, N. Gisin, S. Wehner, and H. Zbinden. Experimental bit commitment based on quantum communication and special relativity. *Phys. Rev. Lett.*, 111:180504, 2013.

[204] W. P. Grice, P. G. Evans, B. Lawrie, M. Legré, P. Lougovski, W. Ray, B. P. Williams, B. Qi, and A. M. Smith. Two-party secret key distribution via a modified quantum secret sharing protocol. *Opt. Express*, 23:7300, 2015.

[205] Hong-Wei Li, Shuang Wang, Jing-Zheng Huang, Wei Chen, Zhen-Qiang Yin, Fang-Yi Li, Zheng Zhou, Dong Liu, Yang Zhang, Guang-Can Guo, Wan-Su Bao, and Zheng-Fu Han. Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources. *Phys. Rev. A*, 84:062308, 2011.

[206] Feihu Xu, Kejin Wei, Shihan Sajeed, Sarah Kaiser, Shihai Sun, Zhiyuan Tang, Li Qian, Vadim Makarov, and Hoi-Kwong Lo. Experimental quantum key distribution with source flaws. *Phys. Rev. A*, 92:032305, 2015.

[207] S.-H. Sun, M.-S. Jiang, and L.-M. Liang. Passive Faraday-mirror attack in a practical two-way quantum-key-distribution system. *Phys. Rev. A*, 83(6):062331, 2011.

[208] Jonathan Barrett, Roger Colbeck, and Adrian Kent. Memory attacks on device-independent quantum cryptography. *Phys. Rev. Lett.*, 110:010503, 2013.

[209] C. Kurtsiefer, P. Zarda, S. Mayer, and H. Weinfurter. The breakdown flash of silicon avalanche photodiodes-back door for eavesdropper attacks? *J. Mod. Opt.*, 48:2039–2047, 2001.

[210] P. V. P. Pinheiro *et al.,* poster presented at *QCrypt 2015, Tokyo, Japan, September 28 – October 2, 2015;* P. V. P. Pinheiro *et al.,* manuscript in preparation.

[211] A. Meda, I. P. Degiovanni, A. Tosi, Z. L. Yuan, G. Brida, and M. Genovese. Backflash light characterization to prevent QKD zero-error hacking. *Nat. Commun.*

[212] A. Lamas-Linares and C. Kurtsiefer. Breaking a quantum key distribution system through a timing side channel. *Opt. Express*, 15(15):9388–9393, 2007.

[213] Yi-Heng Zhou, Zong-Wen Yu, and Xiang-Bin Wang. Making the decoy-state measurement-device-independent quantum key distribution practically useful. *Phys. Rev. A*, 93:042324, 2016.

[214] Marcos Curty, Feihu Xu, Wei Cui, Charles Ci Wen Lim, Kiyoshi Tamaki, and Hoi-Kwong Lo. Finite-key analysis for measurement-device-independent quantum key distribution. *Nat. Commun.*, 5:3732, 2014.

[215] Alberto Boaron, Boris Korzh, Raphael Houlmann, Gianluca Boso, Charles Ci Wen Lim, Anthony Martin, and Hugo Zbinden. Detector-device-independent quantum key distribution: Security analysis and fast implementation. *Journal of Applied Physics*, 120(6):063101, 2016.

[216] Bing Qi. Trustworthiness of detectors in quantum key distribution with untrusted detectors. *Phys. Rev. A*, 91:020303, 2015.

[217] Marcos Curty, Maciej Lewenstein, and Norbert Lütkenhaus. Entanglement as a precondition for secure quantum key distribution. *Phys. Rev. Lett.*, 92:217903, 2004.

[218] S. Sajeed, C. Minshull, N. Jain, and V. Makarov. Invisible trojan-horse attack. *Sci. Rep. (in press). arXiv:1704.07749*, 2017.

[219] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. *Lect. Notes Comp. Sci.*, 765:410, 1994.

[220] Renato Renner. *Security of quantum key distribution.* 2005.

[221] V. Scarani and R. Renner. Quantum cryptography with finite resources: Unconditional security bound for discrete-variable protocols with one-way postprocessing. *Phys. Rev. Lett.*, 100(20):200501, 2008.

[222] P. Chaiwongkhot, S. Sajeed, L. Lydersen, and V. Makarov. Finite-key-size effect in commercial plug-and-play qkd system. *Quantum Sci. Technol. (in press), arXiv:1610.06876*, 2017.

[223] Charles Ci Wen Lim, Nino Walenta, Matthieu Legré, Nicolas Gisin, and Hugo Zbinden. Random variation of detector efficiency: A countermeasure against detector blinding attacks for quantum key distribution. *IEEE J. Sel. Top. Quantum Electron.*, 21:6601305, 2015.

[224] H.-K. Lo and J. Preskill. Security of quantum key distribution using weak coherent states with nonrandom phases. *Quantum Inf. Comput.*, 7:431–458, 2007.

APPENDICES

Appendix A

Quantum state at the input ports of Bob's detectors

In this section, we present the derivation of eq. (6.1) from section 6.3.1. To simplify the discussion, we have labeled different modes involved in the calculations in fig. A.1. Suppose that the input state in mode a is a coherent state $|\sqrt{2\mu}\rangle_a$ with creation operator $a^\dagger = (a_{\mathrm{H}}^\dagger + e^{i\phi_{\mathrm{E}}} a_{\mathrm{V}}^\dagger)/\sqrt{2}$. Also, suppose that the input signal in mode b is the vacuum state $|0\rangle_b$. Then, the output signal in modes c and d, after the action of the 50 : 50 beamsplitter (BS), is given by $|\sqrt{\mu}\rangle_c \otimes |\sqrt{\mu}\rangle_d$, where $c^\dagger = (c_{\mathrm{H}}^\dagger + e^{i\phi_{\mathrm{E}}} c_{\mathrm{V}}^\dagger)/\sqrt{2}$ and $d^\dagger = (d_{\mathrm{H}}^\dagger + e^{i\phi_{\mathrm{E}}} d_{\mathrm{V}}^\dagger)/\sqrt{2}$ denote the corresponding creation operators for modes c and d.

Next, we consider the phase modulator (PM) and the half-wave plate (HWP) that act on modes c and d. The former performs the unitary transformation $c^\dagger = e^{i\varphi_{\mathrm{B}}} e^\dagger$, where $e^\dagger$ is the creation operator at the output port of the PM. The HWP applies the unitary transformation $d_{\mathrm{H}}^\dagger = f_{\mathrm{V}}^\dagger$ and $d_{\mathrm{V}}^\dagger = f_{\mathrm{H}}^\dagger$, where $f_{\mathrm{H}}^\dagger$ and $f_{\mathrm{V}}^\dagger$ denote the creation operators at the output port of the HWP. This means, in particular, that the quantum state in modes e and f has the form

$$|\sqrt{\mu} e^{i\varphi_{\mathrm{B}}}\rangle_e \otimes |\sqrt{\mu}\rangle_f \,, \tag{A.1}$$

with the creation operators $e^\dagger$ and $f^\dagger$ given by $e^\dagger = (e_{\mathrm{H}}^\dagger + e^{i\phi_{\mathrm{E}}} e_{\mathrm{V}}^\dagger)/\sqrt{2}$ and $f^\dagger = (e^{i\phi_{\mathrm{E}}} f_{\mathrm{H}}^\dagger + f_{\mathrm{V}}^\dagger)/\sqrt{2}$, respectively.

Then, after applying the 50 : 50 BS on modes e and f, we have that the output state

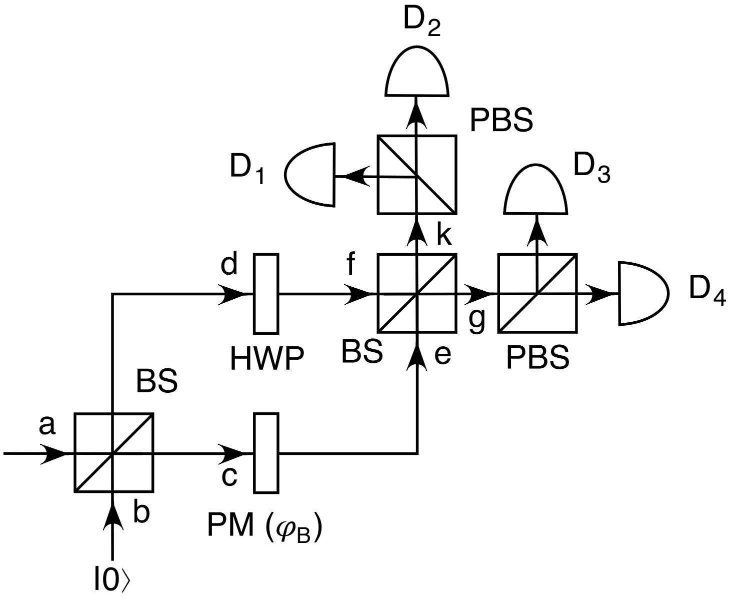


Figure A.1: Schematic representation of Bob's ddiQKD receiver [113]. For simplicity, the modes have been labeled differently than fig. 6.1 from section 6.2; the receiver scheme is otherwise identical. Figure reprinted from [145]

in modes g and k can be expressed as

$$\begin{aligned}\exp\Bigg\{\frac{\sqrt{\mu}}{2}\Big[&\big(e^{i\phi_{\rm E}}-e^{i\varphi_{\rm B}}\big)g_{\rm H}^{\dagger}+\big(1-e^{i(\phi_{\rm E}+\varphi_{\rm B})}\big)g_{\rm V}^{\dagger}\\&+\big(e^{i\phi_{\rm E}}+e^{i\varphi_{\rm B}}\big)k_{\rm H}^{\dagger}+\big(1+e^{i(\phi_{\rm E}+\varphi_{\rm B})}\big)k_{\rm V}^{\dagger}\Big]\Bigg\}\left|0\right\rangle.\end{aligned}\tag{A.2}$$

Finally, if we apply the polarizing beam splitters (PBS) (which we assume reflect horizontally polarized light and let vertically polarized light pass) on modes g and k, we find that the state $|\psi\rangle$ at the input ports of Bob's detectors $\mathrm{D_i}$, with $i \in \{1, 2, 3, 4\}$, is a tensor product of coherent states given by eq. (6.1) in section 6.3.1.

Appendix B

Invisible Trojan-horse attack

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Invisible Trojan-horse attack

Shihan Sajeed[1,2], Carter Minshull[1,3], Nitin Jain[4] & Vadim Makarov[3,1,2]

Received: 28 April 2017
Accepted: 10 July 2017
Published online: 21 August 2017

We demonstrate the experimental feasibility of a Trojan-horse attack that remains nearly invisible to the single-photon detectors employed in practical quantum key distribution (QKD) systems, such as Clavis2 from ID Quantique. We perform a detailed numerical comparison of the attack performance against Scarani-Ac´ın-Ribordy-Gisin (SARG04) QKD protocol at 1924 nm versus that at 1536 nm. The attack strategy was proposed earlier but found to be unsuccessful at the latter wavelength, as reported in N. Jain *et al.*, New J. Phys. 16, 123030 (2014). However at 1924 nm, we show experimentally that the noise response of the detectors to bright pulses is greatly reduced, and show by modeling that the same attack will succeed. The invisible nature of the attack poses a threat to the security of practical QKD if proper countermeasures are not adopted.

Quantum cryptography allows two parties, Alice and Bob, to obtain random but correlated sequences of bits by exchanging quantum states[1–3]. The bit sequences can then be classically processed to get shorter but secret keys. The security of the key relies on the fact that an adversary Eve cannot eavesdrop on the exchange without introducing errors noticeable to Alice and Bob. This constitutes a solution to the problem of key distribution in cryptography, and is better known as quantum key distribution (QKD).

The security of keys distributed over the 'quantum channel' connecting Alice and Bob can be validated by a theoretical security proof. If the amount of errors observed by the two parties exceed a certain threshold, they abort the QKD protocol. Conversely, if the incurred quantum bit error rate (QBER) is below the abort threshold Q_{abort}, the protocol guarantees that Eve cannot know the secret key, except with a vanishingly small probability[3].

However, due to discrepancies between theory and practice, the operation of the QKD protocol may be manipulated by Eve in order to gain information about the key without introducing too many errors. Such discrepancies can arise due to imperfections in the physical devices used in the implementation and/or incorrect assumptions in the theoretical security proofs[3–5]. The field of 'quantum hacking' investigates practical QKD implementations to find such theory-practice deviations, demonstrate the resultant vulnerability via proof-of-principle attacks, and propose countermeasures to protect Alice and Bob from Eve. Over the years, many vulnerabilities have been discovered and attacks have been proposed and demonstrated on both commercial and laboratory QKD systems; see refs 6–8 for reviews. In most cases, it was shown that under attack conditions, the QBER $Q \leq Q_{\text{abort}}$ but Eve's knowledge of the secret key was substantially larger than the predictions of the security proof.

In the so-called Trojan-horse attack[9] (introduced as a 'large pulse attack' a few years before[10]), Eve probes the properties of a component inside Alice or Bob by sending in a bright pulse and analyzing a suitable back-reflected pulse. This attack was recently demonstrated[11] with the intention to breach the security of the Scarani-Acín-Ribordy-Gisin QKD protocol (SARG04)[12] running on the commercial QKD system Clavis2 from ID Quantique[13]. SARG04 is a four-state protocol that is equivalent to the Bennett-Brassard QKD protocol (BB84)[1] in the quantum stage. Their difference comes in the classical processing stage: in SARG04, the bases selections of Bob are used for coding the secret bits, unlike in BB84 where they are publicly revealed. Therefore, if Eve surreptitiously gets information about Bob's bases selections at any time, she can compromise the security of the QKD system running SARG04. (In contrast, a Trojan-horse attack on Bob running the BB84 protocol is normally useless[10], unless it is combined with other attacks[14–16]).

In the attack demonstration[11], it was shown that getting the bases' information in a remote manner was indeed possible via homodyne measurement of the back-reflected photons. The path taken by these photons at 1550 nm, as depicted by the green dotted line in Fig. 1, traverses Bob's phase modulator (PM) twice. The homodyne measurement thus allowed discerning the phase applied by Bob, which is equivalent to knowing his basis selection. This 'phase readout' was accurate in >90% cases even when the mean photon number of the back-reflected pulses was ≈3.

[1]Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. [2]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. [3]Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. [4]Department of Physics, Technical University of Denmark, Fysikvej, Kongens Lyngby, 2800, Denmark. Correspondence and requests for materials should be addressed to S.S. (email: shihan.sajeed@gmail.com) or N.J. (email: nitinj@iitbombay.org)

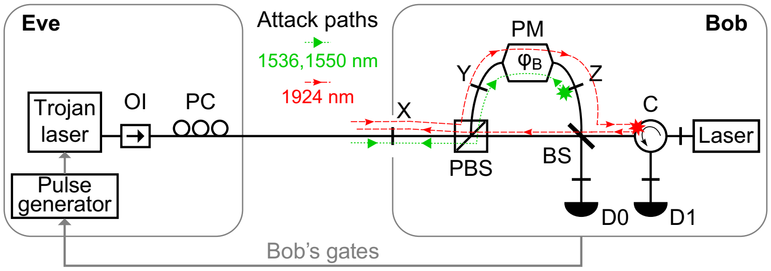


Figure 1. Basic experimental schematic and attack paths at $\lambda_s = 1536\,\text{nm}$ and $\lambda_l = 1924\,\text{nm}$. The scheme and operation of Bob's setup is described in detail in refs 13 and 17. The stars indicate the back-reflection sources exploited in ref. 11 and in this work. Trojan laser models: Eblana Photonics EP1925-DM-B06-FA at λ_l and Alcatel 1905 LMI at λ_s. OI, optical isolator; PC, polarization controller; PBS, polarizing beamsplitter; BS, 50:50 beamsplitter; C, circulator; D, single-photon detectors; X, Y, Z, bulkhead fiber-optic connectors.

Despite that, an overall attack on the QKD system did not have a chance to succeed owing to a side effect produced when the bright pulses went on to hit the detectors D0 and D1, as may be visualized in Fig. 1. To elaborate, the bright pulses result in a severe afterpulsing in these InGaAs/InP single-photon detectors (SPDs), which are operated in a gated mode. For a single bright pulse that hits D1, even if well outside a gate, the cumulative probability of a spurious detection event due to afterpulsing crosses 40% (which is ~4 times the detection probability of a single photon) in just 5 gate periods[18]. The resulting detection events (clicks) are accidental, i.e., erroneous in half of the cases. Hence, only a handful of Trojan-horse pulses (THPs) suffice to rapidly elevate the number of erroneous clicks and make the QBER surpass Q_{abort}, even though Eve's actual knowledge I_E^{act} of the key is still quite small. An elaborate attack strategy to improve I_E^{act} was proposed and numerically simulated, however, it could also not simultaneously satisfy $Q \le Q_{\text{abort}}$ together with $I_E^{\text{act}} > I_E^{\text{est}}$, where I_E^{est} is the estimated (theoretical) security bound on Eve's knowledge that Clavis2 uses to produce the final secret key[11]. While ref. 11 did not prove that a better attack could not be constructed, the attack proposed failed in practice by a large margin.

In this Article, we provide experimental evidence that this Trojan-horse attack could however succeed if Eve were to craft bright pulses at a wavelength where the afterpulsing experienced by the SPDs is considerably lower. The underlying physics is that photons with energy lower than the bandgap of the SPD absorption layer material (InGaAs) mostly pass the material unabsorbed, thereby causing negligible afterpulsing. Indeed, we confirm experimentally that at a relatively longer wavelength $\lambda_l = 1924\,\text{nm}$, the SPD has much less afterpulsing than at $\lambda_s = 1536\,\text{nm}$ (similar to the wavelength used in ref. 11). We then perform a numerical comparison of the attack conditions and performance at λ_l with these at λ_s. By means of an optimized simulation that assumes fairly realistic conditions, we show that the actual attack at λ_l can break the security of Clavis2. The attack in itself is general enough to be potentially applicable to most discrete-variable QKD systems, and can be categorized with those that exploit vulnerabilities arising from the wavelength-dependence of optical components[19,20].

Experiment

While using $\lambda_l = 1924\,\text{nm}$ for the attack offers the benefit of reduced afterpulsing, the transmittance and reflectance properties of different optical components inside Bob vary greatly in comparison with those measured at $\lambda_s = 1536\,\text{nm}$. Most relevant to the attack, the attenuation is generally higher; for instance, the optical loss through the PM at λ_l is $\gtrsim 20\,\text{dB}$ higher than that at λ_s. Furthermore, the modulation itself varies with λ since the modulator's half-wave voltage is a function of wavelength. If Eve uses light at λ_l to estimate Bob's randomly modulated phase ($\varphi_B = 0$ or $\pi/2$ at λ_s) through the homodyne measurement of a pulse that made a single pass through the PM, the measurement outcomes will not be on orthogonal quadratures.

Altogether, it is thus likely that compared to ref. 11, Eve would not only need to inject a larger mean photon number $\mu_{E\to B}$ into Bob, but may also require a higher mean photon number $\mu_{B\to E}$ in the back-reflection for successful homodyne measurements. To calculate the efficacy of the attack, we experimentally quantify at λ_l (relative to λ_s) the following three aspects: increased attenuation, altered phase modulation, and decreased afterpulsing. Figure 1 shows a schematic of the experimental setup used for various measurements.

Increased attenuation.

To gauge the increase in attenuation, we measured the optical loss of various components of Bob at both λ_s and λ_l. In Fig. 1, the dotted line (path X–Y–Z*–Y–X, where $\star$ indicates the source of reflection) shows the attack path used in ref. 11. Relevant loss values are given in the left column of Table 1. With a round trip loss of $L_{\text{X}-\text{Y}-\text{Z}^\star-\text{Y}-\text{X}}(\lambda_s) = 2L_{\text{X}-\text{Y}}(\lambda_s) + \Gamma_{\text{Z}^\star} + 2L_{\text{Y}-\text{Z}}(\lambda_s) = 58.7\,\text{dB}$, Trojan-horse pulses injected with $\mu_{E\to B} \approx 2 \times 10^6$ photons yielded $\mu_{B\to E} \approx 4$ photons in the back-reflection from Bob. Here, $\Gamma_{\text{Z}^\star} = 51.7\,\text{dB}$ is the loss during reflection at Z, the fiber connector after Bob's PM.

For an attack at λ_l with Trojan-horse pulses traversing the same path, the round trip loss would be $L_{\text{X}-\text{Y}-\text{Z}^\star-\text{Y}-\text{X}}(\lambda_l) = 104.9\,\text{dB}$ (with the further assumption that $\Gamma_{\text{Z}^\star}$ is independent of wavelength). The attack pulses at λ_l would therefore face 46.2 dB more attenuation than at λ_s. A major contribution to this large attenuation is from the PM, which even gets doubled since the THPs travel through the PM twice.

Paths & points	Loss at λs (dB)	Loss at λ_l (dB)
X–Y	0.9	3.6
Y–Z	2.6	23.0
Z$^{☆}$	51.7	
Z–C$^{☆}$–X		58.4 to 65.8 (polarization-dependent)
X–D0	8.8 (via long arm)	15.5 (via short arm)
X–C–D1	9.2 (via long arm)	25.8 (via short arm)

Table 1. Comparison of optical losses in Bob at λs versus λ_l. See Fig. 1 for location of the paths and points. The loss during reflection $\Gamma_{Z^☆}$ was measured at 1550 nm[11], which we consider to be close enough to our $\lambda_s = 1536$ nm.

However, since a single pass can also yield information about φ_B, Eve can opt for a different route where only either the input forward-traveling THP or the back-reflected pulse passes through Bob's PM. All Eve requires is a reasonably large source of reflection from any component after the 50:50 beamsplitter (BS). Indeed, during our loss measurements at λ_l we observed a large attenuation through the optical circulator (C), a part of which stems from a rather generous back-reflection. We estimated the loss $L_{Z-C^☆-X}(\lambda_l)$ for the path Z–C$^☆$–X (via BS twice and polarizing beamsplitter once) using a photon-counting method, described below.

We temporarily connected the polarization-controlled output of the 1924 nm laser at Z to send light towards the BS. The average power of the pulsed laser, operated at 5 MHz repetition rate, was $P_{\text{avg}} = 21.55\,\mu$W, corresponding to a mean photon number per pulse $\mu_Z = 4.14 \times 10^7$. An SPD was connected at X to detect the back-reflections from C. To prevent other back-reflections from contributing to the photon counts, Bob's laser and detectors D0 and D1 were disconnected, and the patchcords (with open connectors) were coiled on a pencil to strongly attenuate the propagating light.

Two counters (Stanford Research Systems SR620) were used to measure the number of optical pulses sent by the laser $N = 4.98 \times 10^6$ and the number of pulses received by the detector $n = 323$ maximized over input polarization at Z. The mean photon number per pulse at X was estimated as $\mu_X \approx 59.7$ from the relation,

$$\frac{n-d}{N} = 1 - e^{-\mu_X \eta_d} \approx \mu_X\ \eta_D, \tag{1}$$

where $d = 60$ is the number of dark counts and $\eta_D = 8.85 \times 10^{-7}$ is the single-photon detection efficiency at λ_l, which was estimated in a separate experiment similar to the one in ref. 20. The ratio of the mean photon numbers μ_Z/μ_X provides the overall loss $L_{Z-C^☆-X}(\lambda_l) \approx 58.4$ dB. The dashed line in Fig. 1 shows the complete attack path. Eve's THPs from the quantum channel enter the long arm of Bob, pass through the modulator, and after a reflection from the BS, propagate to the circulator. Here, they get back-reflected and then take the short arm to exit Bob, passing through the BS again. Using Table 1, this path can be characterized by a total loss $L_{X-Y-Z-C^☆-X}(\lambda_l) = L_{X-Y}(\lambda_l) + L_{Y-Z}(\lambda_l) + L_{Z-C^☆-X}(\lambda_l) = 85.0$ dB.

As noted above, the value of μ_X was polarization-sensitive. For the worst input polarization, μ_X decreased by 7.4 dB, changing the overall loss to $L_{X-Y-Z-C^☆-X}(\lambda_l) = 92.4$ dB. For the rest of the paper, we shall assume the attack pulses to be in a polarization midway between the best and the worst, leading to a loss figure of $L_{X-Y-Z-C^☆-X}(\lambda_l) = 87.3$ dB used to decide Eve's photon budget. In terms of photon numbers, this implies that in order to get the same number of photons out from Bob (i.e., $\mu_{B\to E} \approx 4$), Eve needs to inject $\rho = 10^{(-58.7+87.3)/10} = 7.24 \times 10^2$ times more photons at λ_l than at λ_s.

Altered phase modulator response. We now explain an impact of the altered phase modulation experienced by Eve's THPs at λ_l as they travel through Bob's PM. As mentioned before, Bob randomly chooses between voltages $V_0(=0)$ or $V_{\pi/2}$ to apply a phase $\varphi_B = 0$ or $\pi/2$ on Alice's incoming quantum signal at (or in the vicinity of) $\lambda_s = 1536$ nm. Eve's objective is to learn φ_B. The double pass through the PM in ref. 11 implied that Eve had to discriminate between a pair of coherent states with angle $\theta(\lambda_s) \equiv \theta_s = 2 \times \pi/2 = \pi$ between them, as illustrated in Fig. 2(a). At $\lambda_l = 1924$ nm, the phase modulator is expected to lose efficiency and provide less phase shift at the same voltage. Furthermore, Eve's THP only traverses it once. Assuming a linear response of the PM, one can calculate the angle $\theta_l = [V_{\pi/2}(\lambda_s)/V_{\pi/2}(\lambda_l)] \times \pi/2$ between the coherent states available to Eve.

Since the half-wave voltage of the PM at 1924 nm was not specified by the manufacturer, we experimentally measured it. We constructed a balanced fiber-optic Mach-Zehnder interferometer, incorporating the path X–Z (Fig. 1) into one of its arms. We applied a square modulation voltage to the PM, and observed interference fringes at the output port of the interferometer. We adjusted the voltage amplitude until it was causing no light modulation at the output port, indicating an exact 2π phase shift. From this, we found that $V_{\pi/2}(\lambda_l) = 5.7$ V. By the same method with the 1536 nm laser, we found $V_{\pi/2}(\lambda_s) = 3.35$ V.

From this measurement, we calculated $\theta_l \approx 0.294\pi < \theta_s$. The increased overlap between the two states $|\alpha\rangle$ and $|\beta\rangle$ with $|\alpha| = |\beta|$, as depicted in Fig. 2(b), would make discrimination between Bob's choices of φ_B more difficult. Eve can however increase the brightness of the injected Trojan-horse pulse: this would elicit a higher mean photon number in the back-reflection, effectively translating the states farther from the origin to diminish the overlap. The increment factor that makes the distance between the states at λ_l equal to that at λ_s is given by

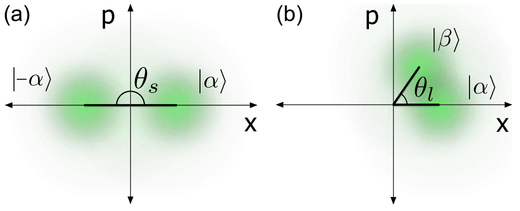


Figure 2. Illustrative phase space representation of the back-reflected states. Eve attempts to discern $\varphi_B = 0$ or $\pi/2$ by performing optimal detection on the back-reflected weak coherent states $|\alpha\rangle$ and $|\beta\rangle$ that have a non-zero overlap. (**a**) The complex amplitude $\beta = \alpha e^{i\theta_s} = -\alpha$, as a result of the double pass at the attack wavelength of λ_s. (**b**) $\beta = \alpha e^{i\theta_l}$, as a result of the single pass at λ_l through Bob's modulator.

$$\nu = \frac{|\alpha - \beta|^2 \text{at } \lambda_s}{|\alpha - \beta|^2 \text{at } \lambda_l} = \frac{1 - \cos\theta_s}{1 - \cos\theta_l} = 5.04, \tag{2}$$

implying that a mean photon number $\mu_{B\to E} \approx 20$ at λ_l would ensure a close-to-unity probability in the phase readout[11].

Decreased afterpulse probability. To quantify the decrease in the afterpulse probabilities in Bob's detectors, we used the setup shown in Fig. 1. A single THP was synchronized to the first in a series of detection gates[11,18] of Bob, and the times at which clicks occurred in the onward gates were then recorded. The delay of the THP relative to the first gate was adjusted such that the pulses going through Bob's long arm hit the detectors just a few nanoseconds after the gate was applied by Bob. Although we did utilize a polarization controller, only a maximum of ~45% of the incoming optical power at λ_l could be routed through the long arm. The remaining light, after having suffered propagation losses through the short arm, hit D0 and D1 around 50 ns *before* the first gate (propagation time through the short arm is ≈50 ns faster than the long arm in Clavis2[17]). These light pulses before the gate were found to be the dominant cause for increased noise in the detectors.

Figure 3 shows the time distribution of counts recorded in detector D0 at the wavelengths λ_s and λ_l. Each of the histograms was prepared by recording 10^6 counts. To make the most of the limited number of histogram bins in the counter (SR620), each bin was 0.4 µs wide and included counts from two consecutive gates. This allowed us to cover a time range of >80 µs. THPs with mean photon numbers $\mu_s = 2.68 \times 10^4$ and $\mu_l = 8.32 \times 10^7$ were used for wavelengths λ_s and λ_l respectively. Despite $\mu_s \ll \mu_l$, the data acquisition for the latter took much longer, indicating that most of the clicks were actually (thermal) dark counts. The number of counts per bin settled down at a constant value, representing dark counts, after ~40 µs (right half of the histogram). The total number of thermal dark counts collected could then be calculated by multiplying this value by the total number of bins in the entire histogram. All remaining counts could then be attributed to afterpulsing. Table 2 lists these counts at the two wavelengths. The afterpulse counts (ApC) make the bulk of the counts at λ_s, while dark counts (DC) are in the majority at λ_l.

It can also be observed in Fig. 3 that afterpulsing decay profile at both wavelengths is roughly similar, however the ratio of longer to shorter lifetime components is slightly larger at λ_l. Although this would help our modeled attack[11], for simplicity we have conservatively assumed that the decay parameters at λ_l are the same as at λ_s[18], aside from different overall afterpulse probability. The decay parameters and Z★ were measured at 1550 nm[11,18], which we consider to be close enough at our wavelength $\lambda_s = 1536$ nm.

To compute a numerical factor γ that compares the afterpulsing noise induced at the two wavelengths, we first take the ratio (ApC/DC) at each wavelength. Then, assuming the dark count probability per detector gate stayed constant between the two measurements, we take a ratio of these ratios. We assume a linear scaling of the afterpulse probability with the energy of the THP, and further normalise for the dissimilar mean photon numbers μ_s and μ_l of the THPs. The numerical factor is then

$$\gamma = \frac{\mu_s}{\mu_l}\frac{(ApC_l/DC_l)}{(ApC_s/DC_s)} = 2.83 \times 10^{-6}. \tag{3}$$

In other words, a photon at λ_l is only 2.83×10^{-6} times as likely to cause an afterpulse as a photon at λ_s.

Attack modeling and discussion. Relative to λ_s, an attack at λ_l can thus effectively decrease the afterpulsing probability in D0 by

$$\delta_0 = \rho\nu\gamma = 1.03 \times 10^{-2}. \tag{4}$$

The factor $\rho\nu = 3.65 \times 10^3$ combines the results discussed previously on the aspects of increased attenuation and altered phase modulation, which required THPs injected into Bob at λ_l to be $\rho\nu$ times brighter than at λ_s to ensure optimal attack performance.

λ (nm)	μ	*ApC*	*DC*
1536	2.68×10^4	867760	162854
1924	8.32×10^7	44981	962140

Table 2. Counts due to thermal dark noise (*DC*) and afterpulsing (*ApC*), extracted from Fig. 3 and corrected for the saturation effect. (*ApC* + *DC*) is greater than 10^6 owing to this correction.

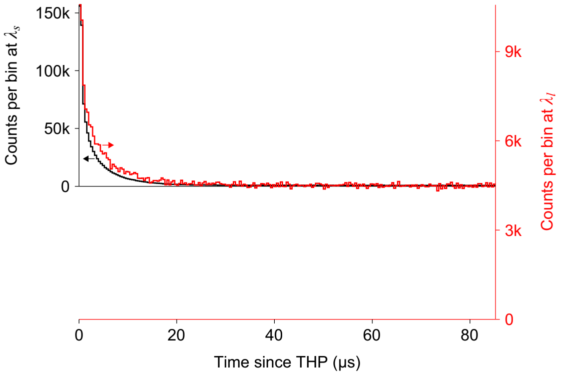


Figure 3. Afterpulse profiles at $\lambda_s = 1536$ nm and $\lambda_l = 1924$ nm. Note that the histograms are rescaled such that their peak counts and dark count rates match in the plot, making visual comparison of decay curves easy. The decay curves are similar but not identical. A total of 10^6 counts were histogrammed at each wavelength. The originally collected histogram data exhibited a saturation effect, in which count rate in later bins was slightly suppressed (by 6.4% for λ_s, 1.0% for λ_l) because of significant click probability in early bins. This has been corrected in the plotted histograms, increasing their total count number above 10^6.

To calculate the afterpulsing probability for D1, one must also consider different losses from Bob's entrance to detectors D0 and D1 for the two attack paths (via the long arm at λ_s and short arm at λ_l, as shown in Fig. 1). We minimised $L_{\mathrm{X-Y}}(\lambda_l)$ by adjusting input polarisation at X, then measured losses between X and the detectors through the short arm. $L_{\mathrm{X-C-D1}}(\lambda_l)$ varied by a factor of 11 over the input polarization, while $L_{\mathrm{X-D0}}(\lambda_l)$ unexpectedly was independent of the input polarization. Using the measured loss values (listed in the last two rows in Table 1), we calculate the effective decrease in the afterpulsing probability in D1

$$\begin{aligned}\delta_1 &= \delta_0 \times 10^{[L_{\mathrm{X-C-D1}}(\lambda_s) - L_{\mathrm{X-D0}}(\lambda_s) - L_{\mathrm{X-C-D1}}(\lambda_l) + L_{\mathrm{X-D0}}(\lambda_l)]/10} \\ &= 1.05 \times 10^{-3}. \end{aligned} \tag{5}$$

With afterpulsing amplitudes reduced by δ_0 and δ_1, we have repeated the simulation of the attack strategy proposed in ref. 11. Let us first recap this strategy, in which Eve manipulates packets or 'frames'[13] of quantum signals traveling from Alice to Bob in the quantum channel. For instance, she may simply block the quantum signals for several contiguous time slots in a frame, thereby preventing any detection clicks (except those arising from dark counts) in Bob over a certain period of time. Conversely, she could substitute the quantum channel with a low-loss version to increase the detection probability in another group of slots. Such actions increase the efficacy of Eve's attack; they provide her some control over when inside a frame Bob's SPDs enter 'deadtime' – a period in which both D0 and D1 are insensitive to single photons and cannot register detection clicks. (In Clavis2, a 10 μs long deadtime is automatically triggered by a click in either of the detectors[18]). This is essentially done by attacking in bursts, i.e., probing the phase modulator by sending bright THPs in a group of slots, thus making the SPDs enter deadtime as quickly as possible to let the afterpulses decay harmlessly and contribute as little as possible to the QBER. By balancing the usage of the low-loss line and the number of slots blocked per frame, Eve can also ensure that Bob does not notice any significant deviation of the observed detection rate (typically averaged over a large number of frames).

A numerical simulation modeling the above attack strategy during the operation of the QKD protocol is used to calculate Bob's incurred QBER Q and Eve's actual knowledge of the raw key I_E^{act}. This is performed for different attack combinations, i.e., by varying the number of slots that are blocked or simply passed via the low-loss line (with or without accompanying THPs). If for at least one combination, I_E^{act} exceeds the estimation I_E^{est} from the security proof but $Q < Q_{\mathrm{abort}}$, the attack strategy is successful in breaching the security.

For an attack at λ_l, we have been able to find several such combinations for the given frame size of $N_f = 1075$ slots and a quantum channel transmittance $T = 0.25$. For instance, in one such combination, a total of 433 slots out of N_f are blocked by Eve. The remaining 642 slots pass from Alice to Bob via a low-loss line with transmittance $T_{LL} = 0.5$, and out of them only 334 slots–periodically distributed in 12 bursts of 28 slots each inside the frame–are accompanied by THPs to read the modulation. With this attack combination, we were able to obtain $I_E^{\mathrm{act}} = 0.515 > I_E^{\mathrm{est}} = 0.506$ (calculation based on Clavis2 parameters and the attack conditions[11]) and $Q = 7.8\% < Q_{\mathrm{abort}} \approx 8\%$ (empirically determined in ref. 21). We remark here that for a similar value of Q, the best optimized attacks at λ_s could not even yield $I_E^{\mathrm{act}} \sim 0.080$. Furthermore, in contrast to the $T_{LL} = 0.9$ used in ref. 11, implementing the attack strategy with $T_{LL} = 0.5$ here makes the attack closer to be feasible in practice.

Note that in the simulation, we have mixed measurement results from two samples of Clavis2 system. The optical loss measurements at λ_l and the relative decrease in afterpulsing come from the system installed in Waterloo (Bob module serial number 08020F130), while the decay parameters of trap levels in avalanche photodiodes measured at λ_s come from the system in Erlangen (Bob module serial number 08008F130). The decay parameters and $Z^\star$ were measured at 1550 nm[11, 18], which we consider to be close enough at our wavelength $\lambda_s = 1536$ nm.We further note that the latter figures vary significantly between D0 and D1, although the two avalanche photodiodes were of the same type and at the same temperature[18]. Therefore our simulation only gives a rough indication of attack performance. Results of the actual attack, if it is performed, will vary from sample to sample. However, also note that we have tested a single long wavelength of 1924; a different wavelength may well yield better attack performance. Finally, more recent commercial systems deploy SPDs with much better efficiencies and afterpulsing characteristics and, as noted in ref. 11, this benefits the eavesdropping strategy.

We expect homodyne detection at 1924 nm to be easy to implement by using p-i-n diodes with extended infrared response[22, 23]. Based on the published specs, the latter should provide detection performance in our setting similar to that demonstrated at 1550 nm[11]. Separating Eve from Bob by some distance of fiber does not degrade the attack very fast; we have measured 7.5 dB/km loss at 1924 nm in a 16.5 cm diameter spool of Corning SMF-28e[24] fiber.

The easiest countermeasure to protect the QKD system from this attack is to properly filter the light entering the system[20, 25]. E.g., adding a narrow-pass filter at Bob's entrance will force Eve to use the signal wavelength λ_s and reduce her attack performance to the original failure, provided poor detector afterpulsing properties are maintained in production[11]. Another countermeasure would be to use a QKD protocol that does not require the receiver's PM settings to be secret, such as BB84 with decoy states[3, 10, 26]. However, protecting the source's PM settings will still be required in most QKD protocols[25, 27].

Conclusion

In conclusion, we have shown that despite the increased attenuation and sub-optimal phase modulation experienced around 1924 nm, the Trojan-horse attack performed at this wavelength has a very good chance of being invisible, because the afterpulsing experienced by Bob's detectors is extremely low. This attack is mostly implementable with commercial off-the-shelf components. Therefore, an urgent need exists to incorporate effective countermeasures into practical QKD systems to thwart such threats.

References

1. Bennett, C. H. & Brassard, G. Quantum cryptography: Public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)*, 175–179 (IEEE Press, New York, 1984).
2. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
3. Scarani, V. *et al.* The security of practical quantum key distribution. *Rev. Mod. Phys.* **81**, 1301–1350 (2009).
4. Makarov, V. Cracking quantum cryptography. In *CLEO/Europe and EQEC 2011 Conference Digest*, ED3_1 (Optical Society of America, 2011).
5. Scarani, V. & Kurtsiefer, C. The black paper of quantum cryptography: real implementation problems. *Theor. Comput. Sci.* **560**, 27–32 (2014).
6. Lo, H.-K., Curty, M. & Tamaki, K. Secure quantum key distribution. *Nat. Photonics* **8**, 595–604 (2014).
7. Jain, N. *et al.* Attacks on practical quantum key distribution systems (and how to prevent them). *Contemp. Phys.* **57**, 366–387 (2016).
8. Liang, L.-M., Sun, S.-H., Jiang, M.-S. & Li, C.-Y. Security analysis on some experimental quantum key distribution systems with imperfect optical and electrical devices. *Front. Phys.* **9**, 613–628 (2014).
9. Gisin, N., Fasel, S., Kraus, B., Zbinden, H. & Ribordy, G. Trojan-horse attacks on quantum-key-distribution systems. *Phys. Rev. A* **73**, 022320 (2006).
10. Vakhitov, A., Makarov, V. & Hjelme, D. R. Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography. *J. Mod. Opt.* **48**, 2023–2038 (2001).
11. Jain, N. *et al.* Trojan-horse attacks threaten the security of practical quantum cryptography. *New J. Phys.* **16**, 123030 (2014).
12. Scarani, V., Acín, A., Ribordy, G. & Gisin, N. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Phys. Rev. Lett.* **92**, 057901 (2004).
13. Clavis2 specification sheet, http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf, visited (16 Apr 2017).
14. Makarov, V., Anisimov, A. & Skaar, J. Effects of detector efficiency mismatch on security of quantum cryptosystems. *Phys. Rev. A* **74**, 022313 (2006). Erratum ibid. **78**, 019905 (2008).
15. Qi, B., Fung, C.-H. F., Lo, H.-K. & Ma, X. Time-shift attack in practical quantum cryptosystems. *Quant. Inf. Comp.* **7**, 73–82 (2007).
16. Lydersen, L. & Skaar, J. Security of quantum key distribution with bit and basis dependent detector flaws. *Quant. Inf. Comp.* **10**, 60–76 (2010).
17. Stucki, D., Gisin, N., Guinnard, O., Ribordy, G. & Zbinden, H. Quantum key distribution over 67 km with a plug&play system. *New J. Phys.* **4**, 41 (2002).
18. Wiechers, C. *et al.* After-gate attack on a quantum cryptosystem. *New J. Phys.* **13**, 013043 (2011).
19. Li, H.-W. *et al.* Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources. *Phys. Rev. A* **84**, 062308 (2011).
20. Jain, N. *et al.* Risk analysis of Trojan-horse attacks on practical quantum key distribution systems. *IEEE J. Sel. Top. Quantum Electron.* **21**, 6600710 (2015).
21. Jain, N. *et al.* Device calibration impacts security of quantum key distribution. *Phys. Rev. Lett.* **107**, 110501 (2011).
22. Extended InGaAs PIN photodiodes IG22-series, http://www.lasercomponents.com/us/product/ingaas-500-2600-nm-1/, visited (16 Apr 2017).

23. InGaAs PIN photodiodes G12182 series, http://www.hamamatsu.com/resources/pdf/ssd/g12182_series_kird1118e.pdf, visited (16 Apr 2017).
24. Corning SMF-28e optical fiber, http://www.princetel.com/datasheets/SMF28e.pdf, visited (16 Apr 2017).
25. Lucamarini, M. *et al.* Practical security bounds against the Trojan-horse attack in quantum key distribution. *Phys. Rev. X* **5**, 031030 (2015).
26. Hwang, W.-Y. Quantum key distribution with high loss: Toward global secure communication. *Phys. Rev. Lett.* **91**, 057901 (2003).
27. Sajeed, S. *et al.* Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing. *Phys. Rev. A* **91**, 032326 (2015).

Acknowledgements

We thank ID Quantique for cooperation, technical assistance, and providing us the QKD hardware. This work was funded by Industry Canada, CFI, NSERC (programs Discovery and CryptoWorks21), Ontario MRI, and the US Office of Naval Research. N.J. acknowledges the warm hospitality of the Institute for Quantum Computing, where this work was carried out.

Author Contributions

S.S. and C.M. performed the experiments. N.J. performed attack modeling and contributed to the experiments. V.M. supervised the study. All authors performed data analysis and contributed to writing the article.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix C

Finite-key-size effect in commercial plug-and-play QKD system

IOP Publishing Quantum Sci. Technol. 2 (2017) 044003 https://doi.org/10.1088/2058-9565/aa804b

Quantum Science and Technology

CrossMark

RECEIVED
12 January 2017

REVISED
12 July 2017

ACCEPTED FOR PUBLICATION
18 July 2017

PUBLISHED
14 August 2017

PAPER

Finite-key-size effect in a commercial plug-and-play QKD system

Poompong Chaiwongkhot[1,2], **Shihan Sajeed**[1,3], **Lars Lydersen**[4] **and Vadim Makarov**[2,3]

[1] Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1 Canada
[2] Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1 Canada
[3] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1 Canada
[4] Department of Electronics Systems, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

E-mail: poompong.ch@gmail.com

Keywords: quantum key distribution, finite-key-size effect, plug-and-play system

Abstract

A security evaluation against the finite-key-size effect was performed for a commercial plug-and-play quantum key distribution (QKD) system. We demonstrate the ability of an eavesdropper to force the system to distill key from a smaller length of sifted-key. We also derive a key-rate equation that is specific for this system. This equation provides bounds above the upper bound of secure key under finite-key-size analysis. From this equation and our experimental data, we show that the keys that have been distilled from the smaller sifted-key size fall above our bound. Thus, their security is not covered by finite-key-size analysis. Experimentally, we could consistently force the system to generate the key outside of the bound. We also test manufacturer's software update. Although all the keys after the patch fall under our bound, their security cannot be guaranteed under this analysis. Our methodology can be used for security certification and standardization of QKD systems.

1. Introduction

Quantum key distribution (QKD) systems are expected to provide unconditionally secure keys between two parties [1–6]. To fulfill that expectation, every feature, imperfection, and loophole both in theory and practice has to be taken into account. One of these features is that, with limited resources and time, a QKD system can exchange only a finite length of raw key. The knowledge of an adversary about the key is estimated by the number of errors in it [7, 8]. Since the bound on the adversary's knowledge is estimated from a finite sample, the smaller the sample is, the less accurate the estimate becomes. Thus, the estimated knowledge might deviate from the actual value and, if it is underestimated, the security of the secret key might be compromised. Finite-key-size analysis [9–14] takes these statistical deviations into account and modifies the key-rate equation accordingly.

Many of the practical QKD systems used today were developed before the finite-key-size analysis in QKD protocols became available. Although some form of finite-key-size effect has been considered in the literature since the year 2000 [4], a rigorous proof was first published in 2005 and developed in the subsequent years [9–14]. While the finite-size analysis was not considered in the security assumptions of the early systems, the generated secret key may still be secure if the raw-key sample size is large enough to neglect the finite-size effects. However, if the sample size is smaller, the effects can no longer be neglected and an absence of the finite-key analysis may render the generated key insecure. This is the main focus of this work. We emphasize the significance of the finite-key-size effects in a practical QKD system. We also demonstrate the ability of an eavesdropper to amplify these effects by actively interfering with the transmission and forcing the system to generate secret key from a smaller sample size. In section 2, we experimentally demonstrate a simple attack that forces a commercial QKD system to use a smaller sample size. The key-rate equation for this specific system is derived in section 3. In section 4, we compare the finite-key security bounds with our experimental data. We test the system again after manufacturer's security update in section 5, and conclude in section 6.

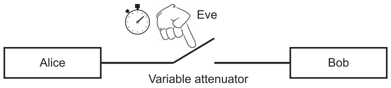


Figure 1. Scheme of experiment.

2. Experiment

The subject of this study is the security of a plug-and-play QKD system Clavis2 produced by ID Quantique [15, 16]. Although updated configurations for plug-and-play systems exist [17], we have not modified the system under test and all tests were performed in the same configuration, as provided by the manufacturer. The QKD protocol under study is Bennett–Brassard 1984 (BB84) protocol without decoy states [1, 7], as implemented in Clavis2. The security of this system implemented in the manufacturer's software is based on the security analysis in [18], which is an analysis against photon-number-splitting attack and cloning attack. The analysis in [18] neither considers the finite-key-size effects nor takes into account the lack of phase randomization in the system. It also assume that Eve cannot change detectors efficiency.

Under normal operation, the system exchanges the quantum signals until the memory buffer for the states sent by Alice is filled. This leads to the raw-key size being limited. This limit varies depending on the line loss (at higher loss fewer photons are received by Bob, and the key is smaller as our experimental data show below). Then, Alice and Bob perform post-processing: sifting, error correction, and privacy amplification [7, 15, 16]. One of the features of Clavis2 is that the system terminates the raw-key exchange process if Bob's photon detection efficiency drops below a certain threshold, and performs a recalibration procedure for the timing alignment of detector gates [16, 19]. This timing alignment greatly affects the photon detection efficiency, is sensitive to environmental fluctuations, and needs to be restored from time to time by performing this recalibration. However the system does not discard the raw key already accumulated in the buffer (as long as it has accumulated at least 80 kbit), and performs the post-processing from the available amount at the time of termination. Eve may take an advantage of this feature. Since the security proof of the system did not take into account the statistical deviation due to non-infinite key length, the deviation can be further amplified if the interruption for recalibration occurs early in the raw-key exchange session.

To demonstrate Eve's ability to force the system to distill from a short raw-key length, we first ran the system in a normal operation mode. The quantum channel between Alice and Bob consisted of a 2 m long optical fiber, and a variable attenuator (OZ Optics DD-100-11-1550) was added to simulate transmission line loss of 2, 3 and 4 dB (see figure 1). We ran multiple sessions of key distribution. In each session, during the raw-key exchange phase, we let the system exchange quantum signals for time τ, then abruptly increased the attenuation to $\approx$40 dB. This reduced the detection rate in Bob below the threshold and forced the system to terminate the key exchange. After that, the system performed post-processing of the already exchanged raw key and reported the secret-key length for that session. At the same time, we reset the variable attenuator to the original loss value. The system then recalibrated the timing alignment, and proceeded to the next raw-key exchange session. We varied τ between 10 and 280 s, so that the raw-key size after termination was between the system's minimum threshold of 80 kbit and the memory buffer limit of 1.6–4 Mbit in Bob (depending on the line loss), corresponding to the leftmost and rightmost experimental points in each plot in figure 2. We also allowed the system to complete some of the sessions naturally without Eve's intervention, which mostly resulted in the maximum key length but occasionally a shorter one. The plots show the variation of secret-key size as a function of the sifted-key size, for different transmission loss values. Note that the sifted-key size plotted is half the raw-key size. The amount of the raw key exchanged did not depend solely on τ. Some sessions experienced fluctuations in transmission loss and detection rate, which caused a lower key exchange rate but not below the termination threshold. Some sessions terminated before we induced the loss, when the detection efficiency dropped below the threshold as the result of naturally occurring timing drift, without Eve's help.

In our analysis, we consider the length of secret key as a function of the sifted-key length, rather than the session time duration. For each session that produced non-zero secret key, we recorded the length of the sifted key, the number of bits disclosed in the error correction, the error rate, and the length of the secret key reported by the system. The system under test did not include finite-key-size analysis in its post-processing. Rather, the post-processing step was programmed to subtract an arbitrarily chosen amount of the key in addition to the value given by the asymptotic security analysis [20]. This subtraction was done to account for any unknown effects that were not included in the system's security analysis. Prior to this study, the security of this arbitrary key subtraction has not been verified. We check this hypothesis below. Note that we consider only the case where

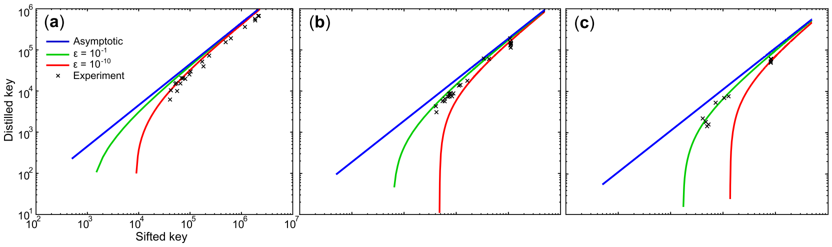


Figure 2. Secret-key size versus sifted-key size. Black × are experimental results with (a) 2 dB line loss and 2.5% error rate averaged over all sessions, (b) 3 dB line loss with 5.2% error rate, and (c) 4 dB line loss with 6.2% error rate. The experiment error rate in individual sessions deviated less than ±0.5% from the average. Blue (dark gray) curve is the infinite-key bound obtained from equation (6). Red (gray) curve is a finite-key-size bound with $\varepsilon = 10^{-10}$ using equation (5). Green (light gray) curve is the finite-key-size bound with $\varepsilon = 10^{-1}$. Secure-key bounds in each sub-figure were calculated separately according to the error rate and line loss of each experiment. Experimental points that lie above each line are considered insecure under those conditions used in each calculation. The detector's efficiency during the experiment was 0.08. The dark count rate was 2×10^{-5} counts per second for each detector.

Eve attempts to control the sifted-key-size before the post-processing. No other attack or flaw is considered in this study.

3. Derivation of key-rate equation

For finite-key-size effects, we need to formulate the key-rate equation for this specific system. To our knowledge, there is no finite-key-size analysis that covers all assumption in this system without hardware modification [21]. In this section, we use available derivation technique to find a secure key bound of the key generated by the system. We assume here that Eve does not interfere with the bright pulses sent from Bob to Alice, and assume that the phase of signal in different time slots is random. As a result, the key bound in this analysis would lie above the upper bound of secure key, which takes into account the lack of phase randomization [22]. Although we cannot conclude that the keys below our bound are secure, it can be used to justify that the secret keys that fall above this bound are not covered by the finite-key-size analysis. Thus, we need to assume the worst case that such keys are insecure.

Our analysis covers the process starting with the raw-key exchange step of plug-and-play system, where Alice attenuates the laser pulses from Bob and encodes each pulse in one of the four possible phase values: 0, $\pi/2$, π, and $3\pi/2$. Alice then sends the encoded signal back to Bob where he measures the signal in one of the two bases, and gets his raw key. They perform sifting and error correction afterwards. The system then performs privacy amplification process where the key is shortened with a universal-2 hash function to exclude Eve's information about the key. The key after this step is the secret key. Eve's information is estimated from quantum bit error rate found during the error correction and probability of having multi-photon pulses during raw key exchange. This process allows us to use a common procedure of secret-key analysis based on [9–11], which stated that, by using the universal-2 hash function as privacy amplification, a secret key K of secret key probability per bit l_K is ε-secure if the protocol is not aborted, and l_K satisfies the relation

$$\varepsilon_{\mathrm{PA}} < 2^{-\frac{1}{2}(H_{\min}(K|E')-l_K)}. \tag{1}$$

Here, $\varepsilon_{\mathrm{PA}}$ is the collision probability of hash function, which is the probability of two different input strings being projected into the same string of output. $H_{\min}(K|E')$ is smooth min-entropy of the system, which represents the probability of Eve guessing the key K correctly using an optimal strategy, given her information about the key before privacy amplification E'.

The goal of this derivation is to replace the smooth min-entropy with a function of measurable parameters from the system. Since the information leakage during error correction is independent of other processes prior to that, E' can be decomposed into Eve's knowledge before error correction E and information leakage during error correction process L. By inequality of smooth entropy [11], we have

$$H_{\min}(K|E') \geqslant H_{\min}(K|E'') - L - 7\sqrt{\frac{1}{n}\log_2\frac{2}{\tilde{\varepsilon}}}, \tag{2}$$

where $H_{\min}(K|E'')$ is the smooth min-entropy of the system before the error correction step. The last term is a statistical correction under finite-key-size regime, where $\tilde{\varepsilon}$ is the probability that Eve's information is underestimated when using smooth min-entropy [12]. The analysis in [12, 23] gave us the bound of $H_{\min}(K|E'')$

as a function of measurable parameters

$$H_{\min}(K|E'') \geqslant A\left(1 - h\left(\frac{\tilde{E}}{A}\right)\right), \tag{3}$$

where $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary Shannon entropy, $\tilde{E} = E + \frac{1}{2}\sqrt{\{2\ln(1/\varepsilon_{\mathrm{PE}}) + 2\ln(n+1)\}(1/n)}$ takes into account a chance that the error rate estimated from a sifted key of size n in the protocol might deviate from the actual value [12], $\varepsilon_{\mathrm{PE}}$ is the probability that such deviation occurs, and E is the observed error rate. The single photon detection probability $A = (p_{\mathrm{det}} - p_{\mathrm{multi}})/p_{\mathrm{det}}$ is a correction term for weak coherent laser used to exchange the raw key in the system [5], where p_{det} is the probability of detection and p_{multi} is the probability of a multi-photon pulse generated by Alice [16].

Now we consider information leakage during the error correction. In theory, the minimum portion of the key with error probability E that needs to be disclosed to correct all the errors is $h(E)$. Using this limit along with the finite-key-size analysis from [13], we have the upper bound of information leakage during error correction

$$L \leqslant \mathrm{leak}_{\mathrm{EC}}A + \log_2\frac{8}{\varepsilon_{\mathrm{EC}}}, \tag{4}$$

where $\mathrm{leak}_{\mathrm{EC}} = f_{\mathrm{EC}}\,h(E)$ is an estimated portion of the key disclosed during error correction. The factor $f_{\mathrm{EC}} = 1.2$ is a practical efficiency of the error correction protocol [8, 12]. In the system log of system under test, this value varied between 1.1 and 1.3. The last term takes account of a failure probability $\varepsilon_{\mathrm{EC}}$ that the error correction leaves non-zero number of errors [13]. This can occur, for example, owing to a non-zero probability of at least one parity check block containing an even number of error bits in every iteration of CASCADE error-correction code and the following parity check rounds in Clavis2 [16].

Since the experimental results are the secret key size as a function of the sifted key length n, we need a secure key bound $l = nl_K$. Substituting equations (2)–(4) into equation (1), taking the logarithm, then multiplying by n on both sides, we obtain

$$l \leqslant nA\left(1 - h\left(\frac{\tilde{E}}{A}\right)\right) - n\ \mathrm{leak}_{\mathrm{EC}} - 7n\sqrt{\frac{1}{n}\log_2\frac{2}{\tilde{\varepsilon}}} - 2\log_2\frac{1}{\varepsilon_{\mathrm{PA}}} - \log_2\frac{2}{\varepsilon_{\mathrm{EC}}}, \tag{5}$$

with security parameter $\varepsilon = \varepsilon_{\mathrm{PE}} + \tilde{\varepsilon} + \varepsilon_{\mathrm{PA}} + \varepsilon_{\mathrm{EC}}$ [11–14]. Since secret-key-rate analyses under collective and coherent attack on non-decoy state BB84 are equivalent [6, 24], the present analysis also covers coherent attack, which is the most general form of attacks on QKD system.

The asymptotic key-rate equation for this specific system can be derived in the same way, but without considering statistical deviation due to finite-size effects. The asymptotic key-rate is

$$l_{\infty} \leqslant nA\left(1 - h\left(\frac{E}{A}\right)\right) - n\ \mathrm{leak}_{\mathrm{EC}}. \tag{6}$$

4. Security verification

To verify the security of the secret key, we compare the experimental result with the bound of the secret key under the asymptotic assumption and finite-key-size analysis. For the asymptotic case, we use equation (6) as the secure key bound. For finite-key-size effect, we use a numerical optimization to find a combination of security parameters ($\varepsilon_{\mathrm{PE}}$, $\tilde{\varepsilon}$, $\varepsilon_{\mathrm{PA}}$, and $\varepsilon_{\mathrm{EC}}$) that maximizes the key length in equation (5). The observed error rate E is an average of the error rates reported by the system after each key distillation at a given transmission loss. The term A is calculated assuming the Poisson distribution with a mean photon number per pulse $\mu = 0.2$ sent by Alice. The value of μ varied between 0.2 and 0.4 in the experiment, however the lowest value gives the highest bound for the secret key rate. We thus obtain bounds of secure key length, plotted in figure 2. Above each curve lies the zone where the security of the key is not covered by finite-key-size analysis.

The experimental secret key sizes, denoted by black ×, always satisfy the security criteria for the asymptotic assumption. When the size of the input sifted key is large, the key-rate bounds with and without finite-size assumption lie very close to each other (see figure 2). This might put the experimentally distilled key size below the finite-key-size bound, i.e., on the safe side. However, when the sifted key size is reduced, the key-rate bounds with and without finite-key assumption diverge significantly. Higher loss results in higher divergence. A fraction of the experimental results falls outside the secure zone for the finite-key-size analysis with values of ε up to 10^{-1}. The latter value means there is a 10% chance that the information of the key generated under this condition might be leaked to Eve. In practice, the security parameter ε can be picked to be of the same order as the probability of major natural disasters such as a serious earthquake, nearby volcanic eruption or nuclear power plant meltdown [25]. If such disaster happened, it is most likely that the security of the key would not matter

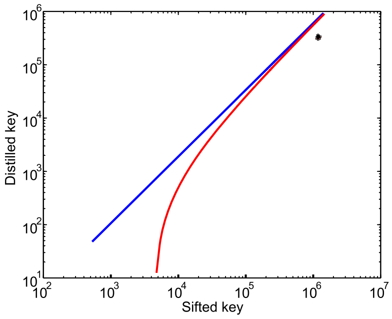


Figure 3. Experimental result with new software. The line loss was 3 dB and error rate was 1%. Blue (dark gray) curve is the asymptotic key bound given by equation (6). Red (gray) curve is the finite-key-size bound for $\varepsilon = 10^{-10}$ given by equation (5). A group of × presents results of 8 key distillations.

anymore. For example, the probability of a nuclear power plant meltdown is 10^{-4} per year, according to the Nuclear Regulatory Commission [26]. If our QKD machine generates two keys every minute or approximately 10^6 keys a year, one might pick $\varepsilon = 10^{-10}$ so that the probability that at least one key leaks to Eve is of the same order as such disasters [25]. However, our experiment shows that Eve can consistently induce a much higher risk probability of key leakage. She can do this by applying our channel interruption technique for BB84 protocol at channel loss values >2 dB (or line distances longer than about 12 km, given typical fiber loss value of 0.17 dB km^{-1}).

5. Testing manufacturer's patch

In the middle of our study in 2014, ID Quantique released a software update for Clavis2. After the update, the system accumulates the raw key over multiple key exchange sessions, and performs post-processing only when the sifted-key size reaches a threshold of about 2 Mbit.

We have repeated our experiment and recalculated our plot using the new parameters acquired from the updated system. The result shows that the secret key is within the secure bound of $\varepsilon = 10^{-10}$ (see figure 3). Regardless of our channel interruptions, we observed that the system has retained the raw key exchanged before termination of each raw-key exchange session, and accumulated it until the size reached about 2 Mbit before proceeding to the distillation. This behavior is clearly visible in the system log and confirmed by the manufacturer [20].

6. Conclusion

In this work, we have done a security evaluation of the finite-key-size effect for Clavis2 system that included derivation of the specific key-rate equation, developing a testing methodology, using it to test the system's security against finite-key-size effects, and testing the manufacturer's patch. Although rigorous security proofs with finite-key-size assumptions were abundant in the literature during the start of this work, they were not assembled together into a key-rate equation suitable for the system under test. Our work has assembled the components of the key-rate equation, verified the assumptions, and put them together into the form of equation (5). However, under our assumptions, the equation does not give the upper bound to evaluate the security of the secret key. Using our result, we can only verify that the keys that fall above the bound are not secure under finite-key-size analysis.

We have shown that by dynamically controlling the channel loss, Eve can force the system to distill key from a shorter sifted-key length to bring the finite-key-effects into play. Using our derived key-rate equation, equation (5), we have shown that key distilled from a sufficiently small length of sifted-key is not guaranteed to be secure, even with the manufacturer's added post-processing step of secret-key subtraction. We have also investigated the security update from ID Quantique, and found that all experimental results fall under the bound in this study. Unfortunately the security of the key against this attack cannot be concluded from this result. Our study only covers statistical evidence from the system against a theoretical bound. An explicit attack that exploits this effect is still open for future study.

Our investigation highlights the significance and importance of finite-key-size analysis in the implementations of QKD, especially in commercial systems. Our method of attack can be used as basis of a testing methodology for security certification. It should be incorporated in the standardization of QKD, which is the next step this technology field faces [27].

We responsibly disclosed to ID Quantique partial results of this investigation before the 2014 patch. Publication has been delayed in order to give the company enough time for patch deployment.

Acknowledgments

We thank the anonymous referee for the insightful and thorough criticism that led to the improvement of the theory in this study. We thank M Curty, N Lütkenhaus, R Renner, and J Skaar for discussions. We thank ID Quantique for cooperation, technical assistance, and providing us the QKD hardware. This work was supported by Industry Canada, NSERC (programs Discovery and CryptoWorks21), CFI and Ontario MRI. PC acknowledges support by Thai DPST scholarship.

Competing interests. The authors have a research collaboration on security analysis of QKD products with ID Quantique. No other competing interests are declared.

ORCID iDs

Poompong Chaiwongkhot https://orcid.org/0000-0002-2825-8287

References

[1] Bennett C H and Brassard G 1984 *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)* (New York: IEEE Press) pp 175–9
[2] Ekert A K 1991 *Phys. Rev. Lett.* **67** 661
[3] Lo H-K and Chau H F 1999 *Science* **283** 2050
[4] Shor P W and Preskill J 2000 *Phys. Rev. Lett.* **85** 441
[5] Lütkenhaus N 2000 *Phys. Rev.* A **61** 052304
[6] Renner R, Gisin N and Kraus B 2005 *Phys. Rev.* A **72** 012332
[7] Bennett C H, Bessette F, Salvail L, Brassard G and Smolin J 1992 *J. Cryptol.* **5** 3
[8] Brassard G and Salvail L 1994 *Lect. Notes Comp. Sci.* **765** 410
[9] Ben-Or M, Horodecki M, Leung D, Mayers D and Oppenheim J 2005 *Lect. Notes Comp. Sci.* **3378** 386
[10] Renner R and Koenig R 2005 *Lect. Notes Comp. Sci.* **3378** 407
[11] Renner R 2005 Security of quantum key distribution *PhD Thesis* ETH Zürich (https://doi.org/10.3929/ethz-a-005115027)
[12] Scarani V and Renner R 2008 *Phys. Rev. Lett.* **100** 200501
[13] Cai R Y Q and Scarani V 2009 *New J. Phys.* **11** 045024
[14] Tomamichel M, Lim C C W, Gisin N and Renner R 2012 *Nat. Commun.* **3** 634
[15] Stucki D, Gisin N, Guinnard O, Ribordy G and Zbinden H 2002 *New J. Phys.* **4** 41
[16] Clavis2 specification sheet, http://idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf (visited: 4 July 2014)
[17] Zhao Y, Qi B, Lo H-K and Qian L 2010 *New J. Phys.* **12** 023024
[18] Niederberger A, Scarani V and Gisin N 2005 *Phys. Rev.* A **71** 042316
[19] Jain N, Wittmann C, Lydersen L, Wiechers C, Elser D, Marquardt C, Makarov V and Leuchs G 2011 *Phys. Rev. Lett.* **107** 110501
[20] I D Quantique 2014 private communication
[21] Zhao Y, Qi B and Lo H-K 2007 *Appl. Phys. Lett.* **90** 044106
[22] Lo H-K, Ma X and Chen K 2005 *Phys. Rev. Lett.* **94** 230504
[23] Tomamichel M and Renner R 2011 *Phys. Rev. Lett.* **106** 110506
[24] Kraus B, Gisin N and Renner R 2005 *Phys. Rev. Lett.* **95** 080501
[25] Renner R 2014 Private communication and lectures
[26] Office of Nuclear Regulatory Research, Regulatory analysis guidelines of the U.S. Nuclear Regulatory Commission NUREG/BR-0058, Rev. 4 (2004), http://nrc.gov/reading-rm/doc-collections/nuregs/brochures/br0058/br0058r4.pdf
[27] Länger T and Lenhart G 2009 *New J. Phys.* **11** 055051

Appendix D

Testing Random-Detector-Efficiency Countermeasure in a Commercial System Reveals a Breakable Unrealistic Assumption

Testing Random-Detector-Efficiency Countermeasure in a Commercial System Reveals a Breakable Unrealistic Assumption

Anqi Huang, Shihan Sajeed, Poompong Chaiwongkhot, Mathilde Soucarros, Matthieu Legré, and Vadim Makarov

***Abstract*— In the last decade, efforts have been made to reconcile theoretical security with realistic imperfect implementations of quantum key distribution. Implementable countermeasures are proposed to patch the discovered loopholes. However, certain countermeasures are not as robust as would be expected. In this paper, we present a concrete example of ID Quantique's random-detector-efficiency countermeasure against detector blinding attacks. As a third-party tester, we have found that the first industrial implementation of this countermeasure is effective against the original blinding attack, but not immune to a modified blinding attack. Then, we implement and test a later full version of this countermeasure containing a security proof. We find that it is still vulnerable against the modified blinding attack, because an assumption about hardware characteristics on which the proof relies fails in practice.**

***Index Terms*— Quantum key distribution, detector blinding attack, countermeasure testing.**

I. Introduction

Currently, applied cryptography systems rely on the hardness of certain mathematical assumptions, which only provides computational security [1], [2]. Once an eavesdropper has enough computing power, such as a quantum computer, the security of these classical encryption algorithms will be broken [3], [4]. However, quantum key distribution (QKD) allows two parties, Alice and Bob, to share a secret key based on the laws of quantum mechanics [5]–[8]. Because of no-cloning theorem [9], an eavesdropper with arbitrary computing power cannot copy the information sent by Alice without leaving any trace, which guarantees the unconditional security of communication [10]–[15].

For this gradually maturing technology, practical QKD systems have been realised in laboratories [16]–[19] and several companies have provided commercial QKD systems to general customers [20]. However, imperfect components used in the implementations lead to security issues that have attracted an increasing attention in the last decade [21]–[30]. Since increasing number of quantum attacks have been demonstrated, academic community is already aware of the security threat from practical loopholes. Therefore, the next step is to come up with loophole-free countermeasures. Importantly, the security of these countermeasures should be verified.

In this paper, an example of testing the security of an implemented countermeasure is given. We examine ID Quantique's attempted countermeasure to earlier discovered bright-light detector control attacks [26], [31], [32] that were demonstrated 6 years ago on ID Quantique's and MagiQ Technologies' QKD products. The countermeasure is to randomly remove some detector gates to force the effective detection efficiency to zero during those slots [33]. The idea is that when an eavesdropper is performing the blinding attack, she will produce click during these removed gates and thus get caught. This countermeasure has been implemented in a commercial system Clavis2 by two authors of this paper working at ID Quantique (M.S. and M.L.), then provided as-is in a form of firmware update to the remaining four authors from the University of Waterloo who played the role of a third-party testing team. The authors from ID Quantique did not participate in the test, however results of the test produced by the testing team were discussed by all authors and agreed upon.

The experimental results produced by the testing team show that although this countermeasure is effective against the original detector blinding attack [26], it is no longer effective if the eavesdropper modifies her attack slightly. We note here that this countermeasure implemented by ID Quantique is the simplest possible version of the original countermeasure proposal [33], and has already been criticised as unreliable

Manuscript received June 6, 2016; revised September 7, 2016; accepted September 8, 2016. Date of publication September 19, 2016; date of current version October 12, 2016. This work was supported in part by Industry Canada, in part by NSERC under the Discovery and CryptoWorks21 Programs, in part by CFI, in part by Ontario MRI, in part by the U.S. Office of Naval Research, in part by ID Quantique, in part by European Commission's FET QICT SIQS, and in part by EMPIR 14IND05 MIQC2 Projects. The work of P. Chaiwongkhot was supported by Thai DPST Scholarship.

A. Huang and S. Sajeed are with the Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: a42huang@uwaterloo.ca).

P. Chaiwongkhot is with the Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 3G1 Canada, and also with the Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

M. Soucarros and M. Legré are with ID Quantique SA, 1227 Carouge, Geneva, Switzerland.

V. Makarov is with the Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 3G1, Canada, also with the Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JQE.2016.2611443

in a later theoretical work [34]. Hence, the testing team has gone further ahead and manually implemented a full version of the countermeasure using two non-zero detection efficiency levels [33], [34], and tested it. Our testing shows that even the full countermeasure is vulnerable to the modified blinding attack. Specifically, we experimentally disprove an assumption that Bob's detection probability under blinding attack cannot be proportional to his single-photon detection efficiency, on which the theoretical analysis in Ref. [34] relies.

The paper is organized as follows. Section II reviews a hacking-and-patching timeline of ID Quantique's Clavis2 QKD system and introduces the countermeasure. In Section III, testing results of ID Quantique's first countermeasure implementation are reported and our modified blinding attack is introduced. Section IV theoretically analyses conditions of a successful attack and shows that the modified blinding attack satisfies them. Moreover, in Section V, based on certain assumptions about a future implementation of the full countermeasure [34], we demonstrate two possible methods to hack this full version implementation. We discuss the practicality of our attacks against installed commercial QKD lines in Section VI and conclude in Section VII.

II. From Loophole Discovery to Countermeasure Implementation

In 2009, the vulnerability of the commercial QKD system Clavis2 [35] to detector blinding attacks was identified and a confidential report was submitted to ID Quantique (the work was published shortly afterwards [26]). After this, ID Quantique has been trying to figure out an experimental countermeasure against these attacks. The timeline of this security problem is shown in Fig. 1. In 2010, ID Quantique proposed a countermeasure that randomizes the efficiency of a gated avalanche photodiode (APD) by randomly choosing one out of two different gate voltages, and filed this idea for a patent [33]. In this way, an eavesdropper Eve does not know the exact efficiency of Bob in every gated slot and thus cannot maintain his detection statistics. At the sifting phase, if the observed detection rates differ from the expected values, Alice and Bob would be aware of Eve's presence and discard their raw keys.

In 2014, Lim *et al.* [34] proposed a specific protocol to realize this countermeasure, which analyses the security mathematically for blinding attacks that obey a certain assumption on their behavior. In the protocol, Bob randomly applies two non-zero detection efficiencies $\eta_1 > \eta_2 > 0$, and measures detection rates R_1 and R_2 conditioned on these efficiencies. The effect of detector blinding attack is accounted via the factor $(\eta_1 R_2 - \eta_2 R_1)\,/\,(\eta_1 - \eta_2)$. Without the blinding attack, the detection rate is proportional to the efficiency, making this factor zero. The analysis makes a crucial assumption that the detection rate under blinding attack $R_1 = R_2$, i.e., it will be independent of Bob's choice of $\eta_{1,2}$. Then, under attack the factor will be greater than zero, and reduces the secure key rate. This solution intends to introduce an information gap between Eve and Bob, for Eve has no information about Bob's random efficiency choice.

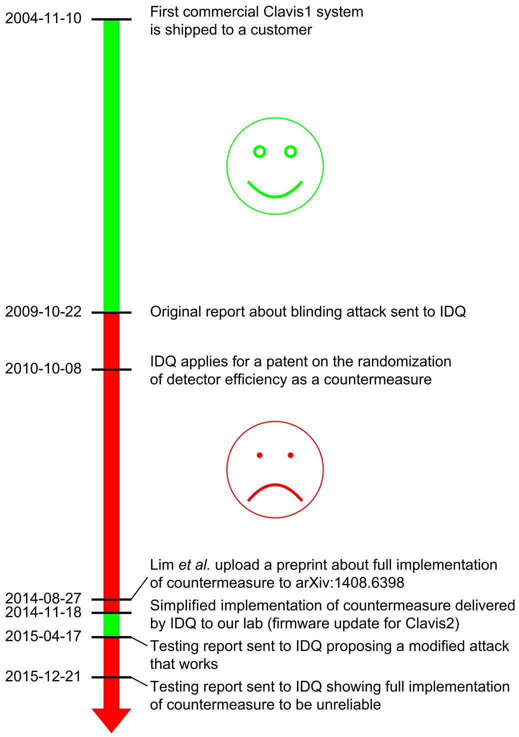

2004-11-10
First commercial Clavis1 system is shipped to a customer
2009-10-22
Original report about blinding attack sent to IDQ
2010-10-08
IDQ applies for a patent on the randomization of detector efficiency as a countermeasure
Lim *et al.* upload a preprint about full implementation of countermeasure to arXiv:1408.6398
2014-08-27
2014-11-18
Simplified implementation of countermeasure delivered by IDQ to our lab (firmware update for Clavis2)
2015-04-17
Testing report sent to IDQ proposing a modified attack that works
2015-12-21
Testing report sent to IDQ showing full implementation of countermeasure to be unreliable

Fig. 1. Timeline of hacking-countermeasure-hacking for the bright-light detector control class of attacks.

Later in 2014, ID Quantique implemented the countermeasure as a firmware patch. The hardware in Clavis2 is not capable of generating two nonzero efficiency levels that switch randomly between adjacent detector gates. As a result, implementation is in a simple form by suppressing gates randomly with 2% probability. The suppressed gates represent zero efficiency $\eta_2 = 0$, while the rest of the gates represent calibrated efficiency $\eta_1 = \eta$. Ideally, in the updated system, there should be no click in the absence of the gate. In practice, transient electromagnetic interference may extremely infrequently lead to a click without a gate. Therefore, an alarm counter is used with the system lifetime limit of 15 clicks in the absence of the gate. If this limit is reached, it triggers the firmware to brick the system and require factory maintenance. This implementation assumes that under blinding attack [26], click probability should not depend on the gate voltage and the attack should therefore cause clicks at the slots of gate absence.

III. Testing the Countermeasure

In this section, we demonstrate that the countermeasure presently implemented by ID Quantique is effective against the original blinding attack [26], but not sufficient against the general class of attacks attempting to take control of Bob's single-photon detectors.

Let us briefly remind the reader how Clavis2 and the original blinding attack against it work. Clavis2 is a bidirectional phase-encoding QKD system [35], [36]. After Bob sends multi-photon bright pulses to Alice, Alice randomly modulates

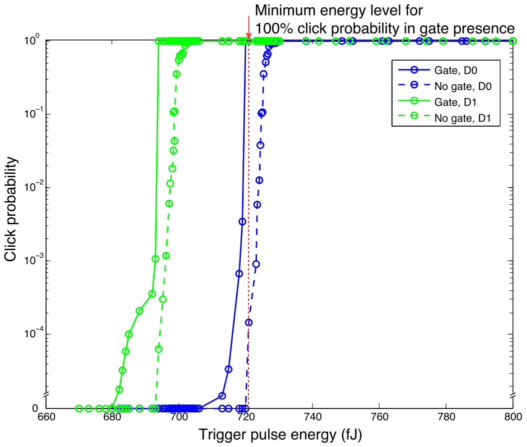


Fig. 2. Click probability under original blinding attack [26] versus energy of trigger pulse. The blinding power is 1.08 mW, as the same as the power used in the published original attack [26]. The timing of trigger pulse is 0.7 ns long, 3 ns after the centre of the gate signal, which should roughly reproduce the original attack [26].

one of the four BB84 phase states [5], attenuates the pulses and sends them back to Bob. Bob randomly chooses one out of two measurement bases. Interference happens between pulses from longer and shorter paths of an interferometer at Bob's side, and the outcomes of interference depend on the phase difference between Alice's and Bob's modulation [37]. However, Eve is able to control the outcomes by the following strategy. She shines bright light to blind the detectors, and then intercepts Alice's states [26]. According to Eve's interception results, she re-sends faked states by multi-photon pulses to Bob's blinded detectors. If Bob chooses the same measurement basis as Eve's, the pulses interfere at Bob's interferometer, so that all power of the pulse goes to one detector to trigger a click. If the measurement bases chosen by Bob and Eve are mismatched, there is no interference, and the power of the pulse is split equally between Bob's two detectors. In this case, neither detector clicks. In this attack, Eve can fully control Bob's detectors and obtain the whole key tracelessly [26].

For the original blinding attack, Eve sends bright-light continuous-wave (c.w.) laser light to blind Bob's detectors. Then a trigger pulse is sent slightly after the gate to make a click. We repeat this attack for improved Clavis2 system and test the amount of energy to trigger a click which is shown in Fig. 2. From Fig. 2, we can see the trigger pulse energy for gate presence (solid curves) is lower than that for gate absence (dashed curves), because minute electrical fluctuations of APD voltage following the gate signal lower the click threshold slightly.

However, if Eve tries to trigger a click with 100% probability when the gate is applied, this amount of trigger pulse energy (marked by a dotted vertical line in Fig. 2) also might trigger a click with non-zero probability when the gate is suppressed, which is monitored and results in an alarm. Therefore, Eve cannot hack the system with full controllability. To avoid clicks in slots of gate suppression, Eve could in theory decrease the level of trigger pulse energy to trigger a click sometimes with gate presence, but never with gate absence. This also satisfies a necessary condition of a successful attack which we will discuss in Section IV later. Unfortunately, in practice, our testing result shows the amount of trigger pulse energy required to trigger D0 without the gate is about 710 fJ, which is only 1.5% less than the amount of energy for 100% click (720 fJ) when the gate is present. The 1.5% difference of these two energy levels is likely not big enough to achieve a reliable attack operation that avoids triggering the countermeasure. Also, D1 will always trigger at these energy levels, revealing the attack. Eve could target D1 using a slightly lower energy level, but the relative precision required is similar there. Routine fluctuations of temperature and other equipment parameters may lead to some instability of these trigger pulse energy levels, causing a risk for Eve to trigger a few clicks in the gate absence and brick the system being attacked. From this point of view, we think this first implementation of countermeasure is effective against the original blinding attack.

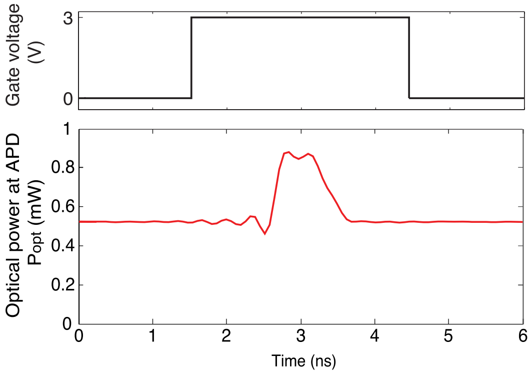


Fig. 3. Idealized APD gate signal and real oscillogram of optical trigger pulse. Relative time between the gate voltage transitions and the optical pulse is approximate. The c.w. signal is generated by a 1536 nm laser diode; the trigger pulse signal is obtained by modulating pump current of a separate 1551 nm laser diode, using an electrical pulse generator [26].

We can slightly modify our blinding attack to break the security of this countermeasure. Similarly to the original blinding attack, Bob's detectors are blinded by a bright-light laser first. Then, instead of sending a trigger pulse slightly after the gate as in the original attacks [26], we send a 0.7 ns long trigger pulse on top of the c.w. illumination *during the detector gate*, as shown in Fig. 3. This trigger pulse produces a click in one of Bob's two detectors only if Bob applies the gate and his basis choice matches that of Eve; otherwise there is no click.

To explain why this modified attack succeeds, let us remind the reader the normal operation of an avalanche photodiode (APD). The detectors in Clavis2 are gated APDs. When the gate signal is applied, the voltage across the APD V_{APD} is greater than its breakdown voltage V_{br}. If a single photon comes during the gated time, an avalanche happens and causes a large current. This current is converted into a voltage by the detector electronic circuit. If the peak voltage is larger than

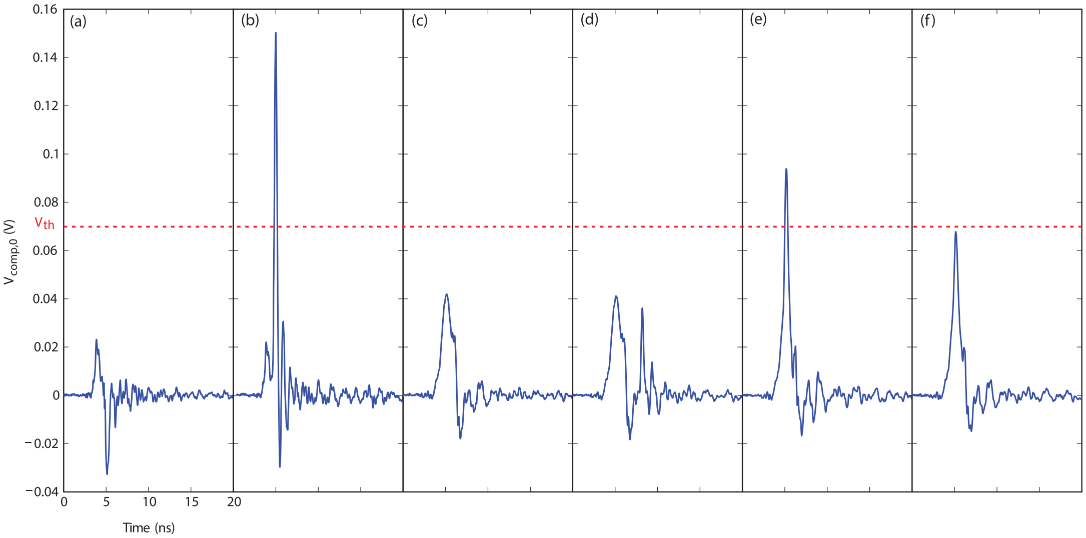


Fig. 4. Oscillograms at comparator input in the detector circuit, proportional to APD current. (a) Geiger mode. The small positive and negative pulses are due to gate signal leakage through the APD capacitance of ∼ 1 pF. (b) Geiger mode, single-photon avalanche. (c–f) The detector is blinded with 0.56 mW c.w. illumination, with (c) no trigger pulse applied, (d) 0.32 pJ trigger pulse applied 5 ns after the gate, (e) 0.32 pJ trigger pulse applied in the gate, and (f) 0.16 pJ trigger pulse applied in the gate.

a threshold $V_{\mathrm{th}} = 70$ mV, the detector registers a photon detection (a ‘click’). Fig. 4(a) and (b) show the cases of no photon coming and a photon introducing an avalanche. Section A explains more details of the detector operation principle and the blinding attack.

A bright laser is able to blind the APDs. Under c.w. illumination, the APD produces constant photocurrent that overloads the high-voltage supply and lowers V_{APD}. Then, even when the gate signal is applied, V_{APD} does not exceed V_{br} and the APD remains in the linear mode as a classical photodetector that is no longer sensitive to single photons. This means the detectors become blinded.

Under the blinding attack, Fig. 4(c–e) shows the detector voltages in different cases: when no trigger pulse is applied and when the trigger pulse is applied either after or in the gate. Since in the linear mode the gain factor of secondary electron-hole pairs generation in the APD depends on the voltage across it, the 3 V gate applied to the APD increases the gain factor. This larger gain during the gated time assists the APD in generating a larger photocurrent than the photocurrent outside the gate. Therefore the gate signal causes a positive pulse as shown in Fig. 4(c). The trigger pulse applied after the gate produces a second pulse, but the peak voltages of neither pulses exceed V_{th} [Fig. 4(d)]. However, when the trigger pulse is shifted inside the gate, the two pulse amplitudes add up, reach V_{th} and produce a detector click [Fig. 4(e)]. If Bob chooses a different measurement basis than Eve, only half of the trigger pulse energy arrives at each detector [26]. In this case, the peak voltage does not reach V_{th} [Fig. 4(f)]. Overall, only when the trigger pulse is applied during the gate time and Bob chooses the same basis as Eve, the detector under the blinding attack clicks. As a result, Eve can control Bob's detectors to make Bob obtain the same measurement result as her, and does not introduce extra errors [26].

Contrary to most of previously demonstrated attacks attempting to take control of single-photon detectors [26], [28], [31], in the present demonstration the timing of the trigger pulse has to be aligned with the gate. Besides timing alignment, another important factor of the attack is the trigger pulse energy E. To test the effect of different trigger pulse energy, we gradually increase it and observe the detection outcomes. Figure 5 shows schematically in which order clicks appear in Clavis2 as E is increased. We observe three thresholds.

- If $E \leq E^{\mathrm{gate}}_{\mathrm{never},i}$ (where $i \in \{0, 1\}$ is detector number), the detector never clicks when the gate is applied.
- If $E \geq E^{\mathrm{gate}}_{\mathrm{always},i}$, the detector always clicks when the gate is applied.
- If $E \leq E^{\mathrm{no\,gate}}_{\mathrm{never},i}$, the detector never clicks when the gate is suppressed.

Figure 6 shows these detection thresholds measured for a range of c.w. blinding powers. All the thresholds rise with the blinding power, because higher blinding power leads to a larger photocurrent and lower V_{APD}. The decreased V_{APD} leads to smaller gain and thus lower sensitivity to the trigger pulse. (Section B contains a more detailed investigation of the processes inside the detector.) As can be seen, for any given blinding power, $E^{\mathrm{no\,gate}}_{\mathrm{never},i}$ is much higher than the other click thresholds. This easily allows the original detector control attack [26] to proceed undetected by the countermeasure. A more formal analysis will be stated in the next section.

IV. Conditions of a Successful Attack

Experimental result of the previous section shows that the attack of Ref. 26 is possible in Clavis2. However, general

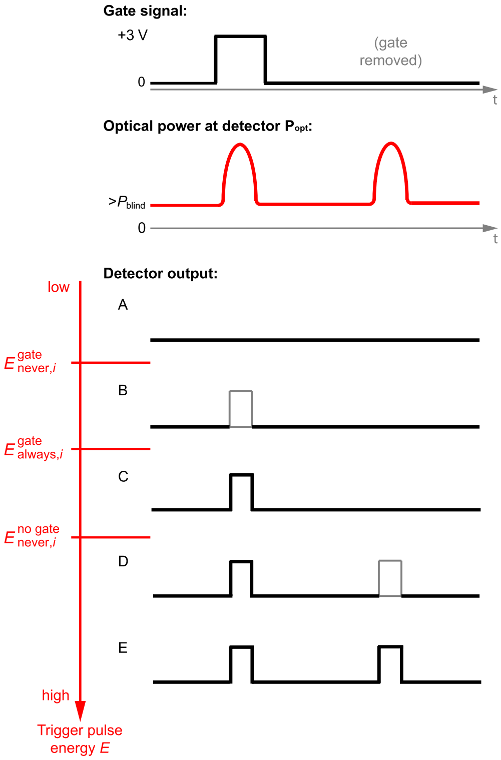


Fig. 5. Output of a blinded detector in Clavis2 under control of trigger pulses of different energy. The top graph shows a gate applied at the first slot, but suppressed at the second slot. However, an optical trigger pulse is sent to the detector in both slots. Graphs A–E show detector output versus trigger pulse energy E. In graph A, the energy is insufficient to produce a click. As the energy is increased above $E^{\text{gate}}_{\text{never},i}$, clicks intermittently appear in the presence of the gate, as shown in graph B. At the energy level above $E^{\text{gate}}_{\text{always},i}$, the gate always has a click, as shown in graph C. However, there is never a click when there is no gate. At a higher energy level above $E^{\text{no gate}}_{\text{never},i}$, clicks in the gate absence appear intermittently (graph D) or always (graph E).

conditions for a successful attack should be analysed theoretically. In this section, we first consider *strong conditions* for a perfect attack, in which Eve induces a click in Bob with 100% probability if their bases match and the gate is applied, and 0% probability otherwise. These conditions are definitely sufficient for a successful attack [26]. However, as we remark later in this section, even if these strong conditions are not satisfied, an attack may still be possible.

Strong conditions. If the detection outcome varies as Fig. 5 with the increase of trigger pulse energy, the order of the three thresholds is:

$$E^{\text{no gate}}_{\text{never},i} > E^{\text{gate}}_{\text{always},i} > E^{\text{gate}}_{\text{never},i}. \tag{1}$$

If Eve and Bob select opposite bases, half of the energy of trigger pulse goes to each Bob's detector. In this case, none of the detectors should click despite the gate presence.

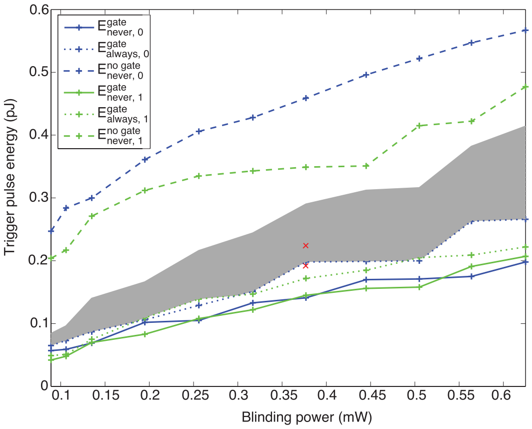


Fig. 6. Energy thresholds of trigger pulse versus c.w. blinding power. Shaded area shows the range of trigger pulse energies of the perfect attack.

This is achieved if [26]

$$\frac{1}{2}\max_i\left\{E^{\text{gate}}_{\text{always},i}\right\} < \left(\min_i\left\{E^{\text{gate}}_{\text{never},i}\right\}\right). \tag{2}$$

The random gate suppression imposes additional conditions. In case of basis mismatch, half of the trigger pulse energy is arriving at each detector. It should induce a click in neither detector when the gate signal is absent. For the target detector i, there is no click once Eq. (1) is satisfied. For the other detector $i\oplus 1$, no click is achieved when half of the trigger pulse energy is still lower than the detection threshold in the no-gate case. That is,

$$\frac{1}{2}E^{\text{gate}}_{\text{always},i} < E^{\text{no gate}}_{\text{never},i\oplus 1}. \tag{3}$$

If the bases match, we need to make sure there is no click when the gate is suppressed, but always a click in the expected detector in the gate presence. This is achieved if $E^{\text{gate}}_{\text{always},i} < E^{\text{no gate}}_{\text{never},i}$, which is already included in inequality (1). Although inequality (3) has a physical meaning, it mathematically follows from inequalities (1) and (2). Thus satisfying inequalities (1) and (2) represents the strong attack conditions and guarantees the same performance as in Ref. 26. The shaded area in Fig. 6 indicates a range of the trigger pulse energies Eve can apply for the perfect attack. The range is sufficiently wide to allow for a robust implementation, only requiring Eve to set correct energy with about ±15% precision.

Necessary condition. An attack may still be possible even if Eve's trigger pulse does not always cause a click in Bob when their bases match, and/or sometimes causes a click when their bases do not match [38]. The latter introduces some additional QBER but as long as it's below the protocol abort threshold, Alice and Bob may still produce key. The random gate removal countermeasure imposes the condition

$$E^{\text{no gate}}_{\text{never},i} > E^{\text{gate}}_{\text{never},i}, \tag{4}$$

which means Eve should be able to at least sometimes cause a click in the gate while never causing a click without the gate (lest the alarm counter is increased). This is a necessary condition for an attack. As the present paper details, there are strong engineering reasons why this condition is likely to be satisfied in a detector. Additional conditions will depend on exact system characteristics [38].

V. Will a Full Implementation of the Countermeasure Be Robust?

We have proved so far that the current countermeasure with gate suppression cannot defeat the detector blinding attack. However, the paper of Lim *et al.* [34] claims that the full version of countermeasure with two non-zero detection efficiencies is effective against a large class of detector side-channel attacks including the blinding attack [26]. Even though this full countermeasure has not been implemented by ID Quantique, we have tested some properties of the detectors in Clavis2 to show two possible methods to hack the full countermeasure, based on certain assumptions about a future implementation.

Bob could choose randomly between $P/2$ and P detection efficiency by changing either gate voltage amplitude V_{gate} or high-voltage supply V_{bias} [34]. Since in Clavis2 hardware V_{gate} is fixed (see Section A), we assume an engineer will change V_{bias} to achieve different non-zero detection efficiencies. To achieve half of original detection efficiency, we lower V_{bias} manually. When $V_{\text{bias},0}$ of D0 drops from -55.26 V to -54.86 V, the detection efficiency P_0 reduces from 22.6% to 12.8%. Similarly, we decrease $V_{\text{bias},1}$ of D1 from -54.70 V to -54.40 V, leading to the detection efficiency P_1 reduction from 18.9% to 9.7%. After that, we test Eve's controllability of these two detectors.

First, we blind the detectors and then measure the relation between the energy of trigger pulse and probability to cause a click. The position of trigger pulse is fixed in the middle of gate signal. Figure 7 shows the testing result which indicates there is a transition range between 0% and 100% click probability.

From the measurement result, Eve can randomly select different levels of trigger pulse energy (shown as dotted lines in Fig. 7) to attack the full version of countermeasure. As we know, only when Bob chooses the same measurement basis as Eve, all the energy of trigger pulse arrives targeted detector and achieves a click. For target D0, if trigger pulse energy E_1 is chosen, D0 always clicks, while at E_2, the detector only clicks if higher V_{bias} is applied. When E_1 and E_2 are chosen randomly with the same probability $P_0/2$, the detection probability for higher V_{bias} is P_0 and the detection probability for lower V_{bias} is only $P_0/2$. Therefore, the attack reproduces correct detection probabilities as the protocol requires. Similarly, for target D1, Eve can choose E_3 to trigger click always and choose E_4 to get a click only if higher V_{bias} is applies. This reproduces correct detection probabilities, $P_1/2$ and P_1. At the same time, E_1 and E_3 remain safely below $E^{\text{no gate}}_{\text{never},0,1}$ shown in Fig. 6, so clicks are never produced in the absence of the gate and alarm is not triggered. This allows Eve to hack the countermeasure tracelessly.

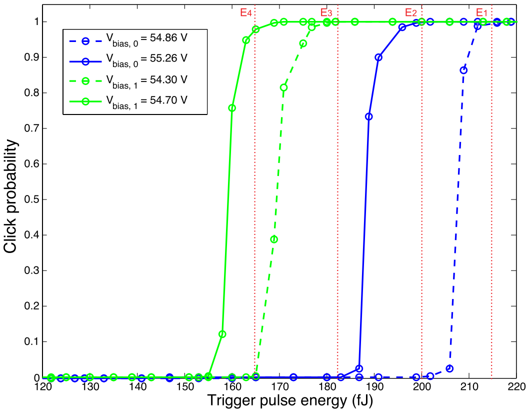


Fig. 7. Click probabilities under blinding attack versus energy of trigger pulse. Solid curves show the energy of trigger pulse for original V_{bias}, while dashed curves for reduced V_{bias} lowering photon detection efficiency by about a factor of 2. The blinding power is 0.38 mW and the timing of trigger pulse is aligned in the middle of the gate by minimizing its energy required to make a click.

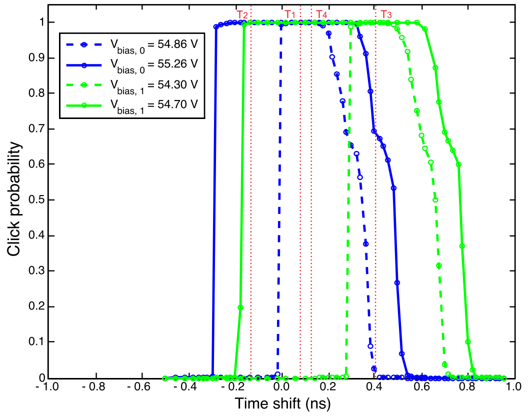


Fig. 8. Click probabilities under blinding attack versus relative time shift of trigger pulse. Solid curves give the detection probability at the original V_{bias}, and dashed curves give the detection probability at lower V_{bias}. Note that the latter extends over a relatively narrower time window. The blinding power is 0.38 mW. The energy of trigger pulse for D0 is 0.22 pJ and for D1 is 0.19 pJ. These energy levels are marked as red × in Fig. 6.

Second, we test the correlation between time shift of trigger pulse and click probability of blinded detector. The trigger pulse energy we use in this test for D1 is slightly lower than that of D0, but both levels of energy are above $E^{\text{gate}}_{\text{always},0,1}$ in Fig. 6 marked as red ×. The measurement result is shown in Fig. 8.

This testing result illustrates another method to attack the countermeasure: randomly adjusting the time shift of the trigger pulse. For D0, after fixing the suitable energy level of trigger pulse, Eve can always trigger a click by choosing time shift T_1, but only trigger a click at higher V_{bias} by choosing T_2. Similarly, if target detector is D1, the detector always clicks at T_3, but only clicks at higher V_{bias} at T_4. Then, when Eve sends trigger pulse to control D0, she randomly selects T_1 and T_2 with equal probability $P_0/2$ to reproduce the correct detection efficiencies of D0. Eve utilizes the same strategy for D1 to achieve correct detection probabilities, $P_1/2$ and P_1. In this way, Eve also hacks Clavis2 system tracelessly.

Generally, a finite set of decoy detection efficiency levels $\eta_1 < \eta_2 < \eta_3 < ... < \eta_n$ can be hacked by properly setting probabilities of different attacking energy levels or time-shifts. We take energy levels of trigger pulse as an example. According to the result in Fig. 7, it is reasonable to extrapolate that we can find n distinct levels of trigger pulse energy $E_1 > E_2 > E_3 > ... > E_n$ in this situation. Then Eve can apply E_k $(k = 1, ..., n)$ with probability q_k to satisfy $\eta_k = \sum_{i=1}^{k} q_i$. This would reproduce every expected value of η_k and hack the system. We have so far assumed that applying energy level E_k causes zero click probability for decoy levels up to η_{k-1}, and 100% click probability for η_k and above. However this is not a necessary condition. More generally, under energy E_k, the click probability for efficiency level η_i is $\beta_{\eta_i}^{E_k}$. To reproduce the expected efficiencies, we need to satisfy the following set of equations:

$$\begin{aligned}
&q_1\beta_{\eta_1}^{E_1} + q_2\beta_{\eta_1}^{E_2} + ... + q_n\beta_{\eta_1}^{E_n} = \eta_1 \\
&q_1\beta_{\eta_2}^{E_1} + q_2\beta_{\eta_2}^{E_2} + ... + q_n\beta_{\eta_2}^{E_n} = \eta_2 \\
&...... \\
&q_1\beta_{\eta_n}^{E_1} + q_2\beta_{\eta_n}^{E_2} + ... + q_n\beta_{\eta_n}^{E_n} = \eta_n. \qquad (5)
\end{aligned}$$

We might solve these equations to get values $0 \leq q_k < 1$. A worse case would be if Eve cannot find values of all q_k, which means she may only have a partial control of Bob's η_k. However, it still breaks the assumption in the security proof [34] that Eve cannot form faked states with click probability conditional on Bob's randomly chosen efficiency. For quantitative analysis, an updated security proof would be needed first.

From the above testing and analysis of the implementation that changes V_{bias}, we can guess that an alternative implementation that changes V_{gate} [34] or adds an intensity modulator in front of the detectors [39], may leave a similar loophole. If we apply the intensity modulator, the energy of the trigger pulse arriving at the detector is not constant but depends on the modulation. However, this case is similar to gate voltage modulation, as we only consider the total energy from the gate signal and trigger pulse. Therefore, we will get similar results as Figs. 7 and 8, but the amount of trigger pulse energy and time shift might be different.

The reason for this practical loophole is a wrong assumption made by Lim and his colleagues [34]. They assume Eve cannot generate faked states that trigger detections with probabilities that are *proportional* to the original photon detection efficiency. Here we have proved this is in fact possible. Therefore, the model of a practical detector should be more precise in security analysis, if one wishes to close the detector control loophole without resorting to measurement-device-independent QKD.

VI. Our Attacks in a Black-Box Setting

According to Kerckhoffs' principle [40], Eve always knows everything about the algorithms and hardware of Alice's and Bob's boxes, including the precise values of equipment parameters. The classical security community practices Kerckhoffs' principle since 1970's, and widely agrees that this is a good approach to implementation security [1]. This is supported by many examples of cryptographic systems that did not follow this principle and were compromised [41]. The quantum academic community certainly agrees that QKD should be made secure in this setting, which is necessary for QKD being unconditionally secure [10]–[15].

However, it is also a practically interesting question if any proposed attack can be mounted on today's commercial QKD systems in a black-box setting, when Eve only has access to the public communication lines but cannot directly measure signals and values of analog parameters inside Alice's and Bob's boxes [42]. In this realistic scenario, Eve may purchase (or acquire by other means) a sample of the system hardware, open it, make internal measurements and rehearse her attacks on it. Then she has to eavesdrop on her actual target, an installed system sample in which she has not had physical access to the boxes. Although the latter sample can be of the same model and design, it will generally have different values of internal analog parameters, owing to sample-to-sample variation in system components. A full implementation of our attacks in this scenario remains to be tested. In this setting it will be of utmost importance for Eve to avoid triggering clicks in the absence of the gate, because this would very quickly brick the system and risk revealing her attack attempt. The original blinding attack that applies the trigger after the gate becomes very sensitive to precise values of thresholds in the presence of the first version of countermeasure (Fig. 2). For this reason we think the countermeasure will likely be triggered by the original attack in the realistic black-box setting.

Our modified attack that applies the trigger inside the gate will likely avoid triggering the alarm, because the no-gate threshold energies are much higher that the energies required for detector control (Fig. 6). It also tolerates some fluctuation in experimental parameters for detector control. For example, when Eve applies 0.38 mW blinding power, 252 fJ trigger pulse energy, and times her trigger pulse at the middle of the gate, we have verified that the attack still works perfectly for up to ±21% change in the trigger energy (see Fig. 6) or up to ±1.3 ns change in the trigger timing. This makes it robust against reasonably expected fluctuations and imprecision of the system parameters. In particular, the timing accuracy required for our attack in much coarser than the several tens of picoseconds precision Alice and Bob use in normal operation [43]. The trigger energy setting precision is similar to the original attack that required ±16% [26].

Eve may need a few attempts to set a correct trigger energy when attacking a new copy of the system. She can do this by starting at a low trigger energy and attempting several increasing values of energy while watching the classical traffic Alice-Bob for the success or failure of the QKD session she has attacked [44]. A QKD session that fails because of too low detection efficiency is a naturally occurring event that is part of normal system operation, does not raise an alarm and is recovered from automatically in Clavis2 [43], [45].

A full two-level implementation of the countermeasure may require Eve to run more attempts, because of a finer degree of control required over the trigger pulse energy and timing. Yet, similarly to the first countermeasure implementation, the no-gate trigger energy that would raise alarm remains safely well above the energies required for detector control. The practicality of attack in the black-box setting is thus difficult to predict without having the actual industrial implementation of the full countermeasure, and actually demonstrating the full attack, which can be a future study.

VII. Conclusion

We have tested the first implementation of the countermeasure against the blinding attack in the commercial QKD system Clavis2. Our testing result demonstrates that presently implemented countermeasure is effective against the original blinding attack but not effective against a modified blinding attack. The modified attack fully controls Bob's single-photon detectors but does not trigger the security alarm. The modified attack is similar to the original detector blinding attack [26] with the only difference that the trigger pulses are time-aligned to coincide with the detector gates, instead of following it. We argue that this attack should be implementable in practice against an installed QKD communication line where Eve does not have physical access to characterising Alice and Bob, however such full demonstration has not yet been done, to our knowledge.

We have also tested the full proposed implementation of countermeasure with two non-zero efficiency levels, and found its security to be unreliable despite predictions of the theory proposal [34]. From the current testing results, bright-pulse triggering probabilities of the blinded detectors depend on several factors including V_{bias}, timing and energy of the trigger pulse (see Section V). This in principle allows Eve to compromise the full countermeasure implementation.

We have tested the countermeasure implemented with the gated single-photon detectors (SPDs). The idea of random detection efficiency can be applied to other types of SPDs that are also sensitive to the blinding attack: free-running SPDs [46] and superconducting nanowire SPDs [28]. However, the countermeasure based on these detectors might still be hackable. Since the efficiencies of these types of SPDs depend on the bias voltage or current, varying these bias signals likely changes other parameters inside the SPD and its electronics. Therefore, when we randomize the detection efficiency, other degrees of freedom might be changed as well. Eve has a chance to exploit these side channels to hack the countermeasure. Of course, the exact outcome cannot be known until the countermeasures in different types of detectors are experimentally tested.

According to our testing result, this countermeasure is not as reliable as would be expected in a high-security environment of QKD. Although an ideal industrial countermeasure has not been achieved, everybody now has a more clear concept about the detector loopholes. This procedure emphasizes the necessity of security testing every time practical QKD systems are developed or updated. We only can reach the final practical security of any QKD system after several iterations of implementation development and testing verification. Our countermeasure testing also illustrates that patching a loophole is still time-consuming and difficult. However, addressing practical vulnerabilities at the design stage of a QKD system is both cheaper and less messy than trying to retrofit patches on an existing deployed solution. Addressing security at the design stage should be the goal whenever possible.

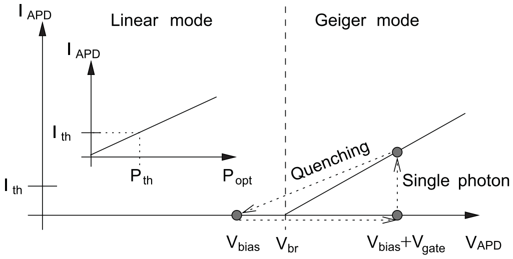


Fig. 9. Linear-mode and Geiger-mode APD operation (reprinted from [26]).

Appendix A
Background

In this section, we recap the operating principle of the single-photon detector, its implementation in Clavis2, and the original blinding attack [26]. Most available single-photon detectors are APDs operating in Geiger mode, in which they are sensitive to single photons [47]. As shown in Fig. 9, when the APD is reverse-biased above its breakdown voltage V_{br}, a single photon can cause a large current I_{APD}. If this current exceeds the threshold I_{th}, electronics registers this as a photon detection (a 'click'). After that, an external circuit quenches the avalanche by lowering the bias voltage V_{APD} below V_{br}, and the APD comes into a linear mode. If the APD is illuminated by bright light (which does not happen in normal single-photon operation but can happen during an eavesdropping attack), I_{APD} in the linear mode is proportional to the incident bright optical power P_{opt}. I_{th} then becomes a threshold on the incident optical power P_{th} that makes a click.

From an engineering view, the detector can be analyzed by its circuit. Figure 10 shows an equivalent circuit diagram of the two detectors used in Clavis2. When no gate signal is applied, the APDs are biased slightly below their V_{br} by the negative high-voltage supply $V_{\text{bias},0} = -55.26$ V, $V_{\text{bias},1} = -54.70$ V.[1] To bring the APD into Geiger mode, an additional 3 V high, 2.8 ns long pulse is applied through a logic level converter DD1. The anode of the APD is AC-coupled to a fast comparator DA1. Since the capacitor C1 blocks the DC component, only when the current flowing through the APD changes, it generates a pulse as the input of DA1. If the peak voltage of this pulse is greater than the positive threshold $V_{\text{th}} = 70$ mV, the comparator produces a logic output signal indicating a click. Once a click in either of the two Bob's detectors is registered, the next 50 gates will not be applied to both detectors, which constitutes a deadtime to reduce afterpulsing.

[1] Using values from the sample of Clavis2 tested in our present study at the University of Waterloo, which is a different sample than in [26], [31], and [32].

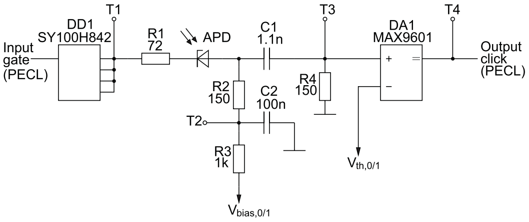


Fig. 10. Equivalent detector bias and comparator circuit, as implemented in Clavis2 (reprinted from [26]).

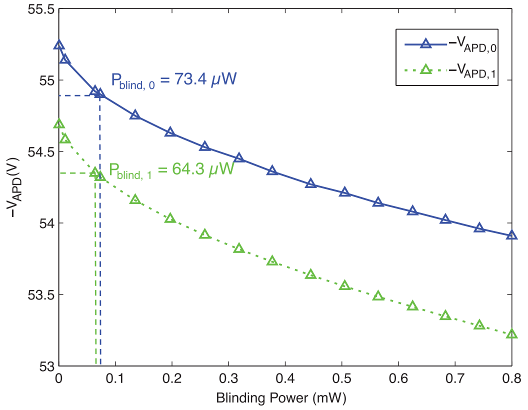


Fig. 11. Bias voltage of APDs versus c.w. blinding power.

If Eve sends a bright c.w. illumination to the gated detectors, the bright light makes the APD generate a significant photocurrent that monotonically increases with the optical power P_{opt}. When we consider effects of this current on the whole detector circuit (Fig. 10), the most useful one is a reduction of the voltage across the APD V_{APD}. Although the high-voltage supply V_{bias} stays constant, the photocurrent causes a significant voltage across R3 $= 1$ kΩ, thus V_{APD} drops. If we apply enough illumination power, V_{APD} will be less than V_{br} even inside the gate, and the APD then always stays in the linear mode. The detector becomes blind to single photons. In our testing, we measure the voltage at test point T2 V_{T2} in Fig. 10 and refer to this voltage as V_{APD} in the text. V_{T2} is close to real V_{APD}, because $\text{R1} + \text{R2} \ll R3$ [precisely, $V_{\text{APD}} = V_{\text{T2}} + (V_{\text{T2}} - V_{\text{bias}})(\text{R1} + \text{R2})/\text{R3}$].

After blinding Bob's detectors, Eve can conduct a faked-state attack. Eve first intercepts all photons sent by Alice. Whenever Eve detects a photon, she sends the same state to Bob via a bright trigger pulse of a certain energy, superimposed on her blinding illumination. Only if Bob chooses the same measurement basis as Eve and applies the gate, one of Bob's detectors will click and he will get the same bit value as Eve. Otherwise, there is no click at Bob's side. During the sifting procedure, Alice and Bob keep the bit values when they have chosen the same basis, and so does Eve. Therefore Eve has identical bit values with Bob, introduces no extra QBER, and does not increase the alarm counter. Eve then listens to the public communication between Alice and Bob and performs the same error correction and privacy amplification procedures as them, to obtain an identical copy of their secret key [26].

Appendix B

Analysis of Processes in the Detector

For further understanding of the detector behaviour under successful blinding attack, we attempt to quantitatively model electrical and thermal processes in it. As we mentioned previously, the bias voltage decreases when the blinding power is applied. A measured relationship between V_{APD} and continuous blinding power is shown in Fig. 11. Detector 0 is blinded at $P_{\text{opt}} > P_{\text{blind,0}} = 73.4$ μW and detector 1 is blinded at $P_{\text{opt}} > P_{\text{blind,1}} = 64.3$ μW. Higher blinding illumination leads to lower bias voltage. This is consistent with the same measurement done for the original blinding attack [26].

In a detector blinded by c.w. laser illumination, the gain factor is affected by not only the power of blinding laser, but also the gate signal. When the APD is blinded and forced to work in the linear mode, it can be treated as an ordinary photodiode with a finite internal gain. Photoelectrons and holes are accelerated by a high electric field and initiate a chain of impact ionizations that generates secondary electron-hole pairs. Thus, the APD has an internal multiplication gain factor $M > 1$, since one photon can yield many electrons of photocurrent flowing in the circuit. When V_{APD} is much lower than V_{br}, M will be close to 1. However, the APD may not have any significant photosensitivity below so-called punch-through voltage, below which the electrical field does not extend into the absorption layer of InGaAs/InP heterostructure [48].

We have done a measurement of small-signal gain G of the APDs in Clavis2 by measuring their photocurrent response to a short optical pulse input. The results are shown in Fig. 12. There is virtually no photosensitivity below the punch-through voltage of about 31 V. Above that voltage G starts at ~ 0.7 A/W (corresponding to $\sim 60\%$ quantum efficiency assuming $M = 1$), then rises above 100 A/W closer to V_{br}. The gain values measured at $V_{\text{br}} - 2$ V are ~ 7 and ~ 10 A/W, which is consistent with values from data sheets of commercial APDs. From the above measurements, we know that Eve can vary the amount of blinding power to the detectors to control the bias voltage and thus the gain factor.

After we blind Bob's detectors in Clavis2, the gain factor is greater during the 2.8 ns gate duration, because the gate signal raises V_{APD}. Thus the electrical charge generated by the APD in response to a trigger pulse applied in the gate is greater than when it's applied outside the gate. For example, in Fig. 4(c), the gate pulse alone contributes 1.053 pC extra charge on top of the current that would be generated without the gate. When the trigger pulse is applied after the gate [Fig. 4(d)], the total charge of the two pulses is 1.467 pC; however, when the trigger pulse is moved into the gate [Fig. 4(e)], the total charge rises to 1.613 pC. Therefore, a greater gain factor during the gated time helps the pulse to cross the threshold.

We have attempted to model the increased gain due to the gate. In our model, we consider a thermal effect and an internal resistance of the APD. On the one hand, an increased temperature raises V_{br} [49]. Electrical heating $(V_{\text{APD}} \cdot I_{\text{APD}})$

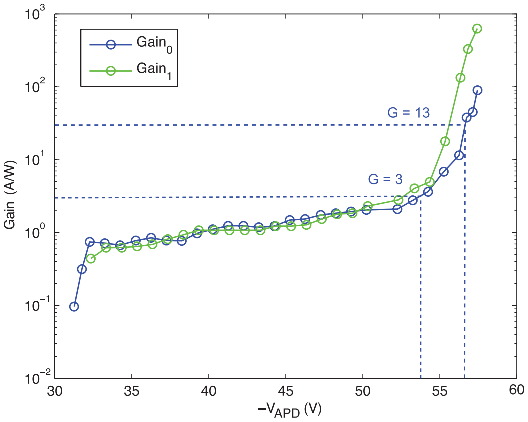


Fig. 12. Gain versus APD bias voltage. Values of gain for bias voltages below 31 V were negligibly low for a practical attack, and below the sensitivity of our measurement method.

and the absorption of the blinding power result in a heat dissipation: 61.2 mW for detector 0 and 66.03 mW for detector 1.[2] Then, an estimated 190 K/W thermal resistance [31] between each APD chip and the cold plate converts the power dissipation into the increased temperature. The temperature-dependent breakdown voltage increases with the coefficient of about 0.1 V/K [31]. As a result, V_{br} increases by 1.16 V (1.25 V) for detector 0 (1). Figure 12 shows the relation between gain factor and the actual V_{APD} in the linear mode. When V_{APD} is close to V_{br}, the gain factor increases rapidly. On the other hand, we suppose the APD has a passive internal resistance, so the internal bias voltage across the ideal photodiode is less than the value of V_{APD} we test. By measuring the voltage of a stable avalanche pulse and calculating the current trough the detector circuit when avalanche happens, we obtain the internal resistance of 330 Ω in detector 0 and 275 Ω in detector 1. Therefore, the real bias voltage under blinding attack shown in Fig. 4(c–f) is 53.77 V, which corresponds to $G = 3$ A/W in detector 0 as shown in Fig. 12. When 3 V gate is applied, the bias voltage becomes 56.77 V which corresponds to $G = 13$ A/W in Fig. 12. However, the measured charges in Fig. 4(d) and (e) illustrate much less gain change: $G = 1.3$ A/W at 53.77 V and $G = 1.76$ A/W at 56.77 V.[3] The discrepancy may be explained by a larger actual thermal resistance between the APD and the cold plate than we estimate, which should be verified in future research.

[2] Under 0.564 mW blinding power, $V_{\mathrm{APD},0} = 54.14$ V, $I_{\mathrm{APD},0} = 1.12$ mA. Heat dissipation of detector 0: 54.14 V $\cdot$ 1.12 mA + 0.564 mW= 61.2 mW; $V_{\mathrm{APD},1} = 53.484$ V, $I_{\mathrm{APD},1} = 1.224$ mA, Heat dissipation of detector 1: 53.484 V $\cdot$ 1.224 mA + 0.564 mW = 66.03 mW.

[3] When we apply a 0.32 pJ trigger pulse after the gate, this single trigger pulse contributes 0.414 pC charge which is the difference between the total charges in Fig. 4(c) and (d). $G = 0.414$ pC/ 0.32 pJ = 1.3 A/W. When we apply a 0.32 pJ trigger pulse during the gate, this single trigger pulse contributes 0.56 pC charge which is the difference between the total charges in Fig. 4(c) and (e). $G = 0.56$ pC/ 0.32 pJ = 1.76 A/W.

Acknowledgment

The authors thank C. C. W. Lim, N. Gisin, and E. Anisimova for discussions. This work was supported by Industry Canada, NSERC (programs Discovery and CryptoWorks21), CFI, Ontario MRI, US Office of Naval Research, ID Quantique, European Commission's FET QICT SIQS and EMPIR 14IND05 MIQC2 projects. P.C. was supported from Thai DPST scholarship.

References

[1] M. Naor, "On cryptographic assumptions and challenges," in *Advances in Cryptology—CRYPTO* . Berlin, Germany: Springer, 2003, pp. 96–109.

[2] *ETSI White Paper no. 8: Quantum Safe Cryptography and Security*, ETSI, Sophia Antipolis, France, 2015.

[3] C. H. Bennett and D. P. DiVincenzo, "Quantum information and computation," *Nature*, vol. 404, pp. 247–255, Mar. 2000.

[4] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, 1994, pp. 124–134.

[5] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Comput., Syst., Signal Process.*, Bengaluru, India, 1984, pp. 175–179.

[6] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Phys. Rev. Lett.*, vol. 67, no. 6, pp. 661–663, Aug. 1991.

[7] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Modern Phys.*, vol. 74, no. 1, pp. 145–195, Mar. 2002.

[8] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, no. 3, pp. 1301–1350, 2009.

[9] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, pp. 802–803, Oct. 1982.

[10] H.-K. Lo and H. F. Chau, "Unconditional security of quantum key distribution over arbitrarily long distances," *Science*, vol. 283, no. 5410, pp. 2050–2056, Mar. 1999.

[11] P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 441–444, 2000.

[12] N. Lütkenhaus, "Security against individual attacks for realistic quantum key distribution," *Phys. Rev. A*, vol. 61, no. 5, p. 052304, 2000.

[13] D. Mayers, "Unconditional security in quantum cryptography," *J. ACM*, vol. 48, no. 3, pp. 351–406, May 2001.

[14] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," *Quantum Inf. Comput.*, vol. 4, no. 5, pp. 325–360, 2004.

[15] R. Renner, N. Gisin, and B. Kraus, "Information-theoretic security proof for quantum-key-distribution protocols," *Phys. Rev. A*, vol. 72, no. 1, p. 012332, 2005.

[16] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," *Phys. Rev. Lett.*, vol. 68, no. 21, pp. 3121–3124, 1992.

[17] T. Schmitt-Manderbach, *et al.* "Experimental demonstration of free-space decoy-state quantum key distribution over 144 km," *Phys. Rev. Lett.*, vol. 98, no. 1, p. 010504, 2007.

[18] D. Stucki *et al.*, "High rate, long-distance quantum key distribution over 250 km of ultra low loss fibres," *New J. Phys.*, vol. 11, no. 7, p. 075003, 2009.

[19] Y.-L. Tang *et al.*, "Field test of measurement-device-independent quantum key distribution," *IEEE J. Sel. Top. Quantum Electron.*, vol. 21, no. 3, p. 1, May/Jun. 2015.

[20] Several companies sell QKD systems: ID Quantique (Switzerland). [Online]. Available: http://www.idquantique.com/; The Austrian Institute of Technology (Austria). [Online]. Available: http://www.ait.ac.at/; QuantumCTek (China). [Online]. Available: http://www.quantum-info.com/; and Qasky (China). [Online]. Available: http://www.qasky.com/

[21] A. Vakhitov, V. Makarov, and D. R. Hjelme, "Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography," *J. Mod. Opt.*, vol. 48, no. 13, pp. 2023–2038, 2001.

[22] V. Makarov, A. Anisimov, and J. Skaar, "Effects of detector efficiency mismatch on security of quantum cryptosystems," *Phys. Rev. A*, vol. 74, no. 2, p. 022313, 2006.

[23] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, "Trojan-horse attacks on quantum-key-distribution systems," *Phys. Rev. A*, vol. 73, no. 2, p. 022320, 2006.

[24] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," *Quantum Inf. Comput.*, vol. 7, nos. 1–2, pp. 73–82, 2007.

[25] Y. Zhao, C.-H. Fung, B. Qi, C. Chen, and H.-K. Lo, "Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems," *Phys. Rev. A*, vol. 78, no. 4, p. 042333, 2008.

[26] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Hacking commercial quantum cryptography systems by tailored bright illumination," *Nature Photon.*, vol. 4, no. 10, pp. 686–689, 2010.

[27] S.-H. Sun, M.-S. Jiang, and L.-M. Liang, "Passive Faraday-mirror attack in a practical two-way quantum-key-distribution system," *Phys. Rev. A*, vol. 83, no. 6, p. 062331, 2011.

[28] L. Lydersen, M. K. Akhlaghi, A. H. Majedi, J. Skaar, and V. Makarov, "Controlling a superconducting nanowire single-photon detector using tailored bright illumination," *New J. Phys.*, vol. 13, p. 113042, Nov. 2011.

[29] P. Jouguet, S. Kunz-Jacques, and E. Diamanti, "Preventing calibration attacks on the local oscillator in continuous-variable quantum key distribution," *Phys. Rev. A*, vol. 87, p. 062313, Jun. 2013.

[30] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, "Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch," *Phys. Rev. A*, vol. 91, p. 062301, Jun. 2015.

[31] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Thermal blinding of gated detectors in quantum cryptography," *Opt. Exp.*, vol. 18, no. 26, p. 27938–27954, 2010.

[32] C. Wiechers *et al.*, "After-gate attack on a quantum cryptosystem," *New J. Phys.*, vol. 13, no. 1, p. 013043, 2011.

[33] M. Legre and G. Ribordy, "Apparatus and method for the detection of attacks taking control of the single photon detectors of a quantum cryptography apparatus by randomly changing their efficiency," WO Patent 2012 046 135 A2, Apr. 12, 2012.

[34] C. C. W. Lim, N. Walenta, M. Legré, N. Gisin, and H. Zbinden, "Random variation of detector efficiency: A countermeasure against detector blinding attacks for quantum key distribution," *IEEE J. Sel. Topics Quantum Electron.*, vol. 21, no. 3, May/Jun. 2015, Art. no. 6601305.

[35] (Sep. 22, 2016). *Clavis2 Specification Sheet*. [Online]. Available: http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf

[36] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, "Quantum key distribution over 67 km with a plug&play system," *New J. Phys.*, vol. 4, no. 1, p. 41, 2002.

[37] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, "'Plug and play' systems for quantum cryptography," *Appl. Phys. Lett.*, vol. 70, no. 7, pp. 793–795, 1997.

[38] L. Lydersen *et al.*, "Superlinear threshold detectors in quantum cryptography," *Phys. Rev. A*, vol. 84, p. 032320, Sep. 2011.

[39] T. Moroder, M. Curty, and N. Lütkenhaus, "Detector decoy quantum key distribution," *New J. Phys.*, vol. 11, p. 045008, Apr. 2009.

[40] A. Kerckhoffs, "La cryptographie militaire," *J. Sci. Militaires*, vol. 9, pp. 5–38, Jan. 1883.

[41] S. Singh, *The Secret History of Codes and Code-breaking*. London U.K.: Four Estate, 1999.

[42] N. Gisin, "Quantum Cryptography: where do we stand," in *Proc. QCrypt*, Tokyo, Japan, Sep./Oct. 2015. [Online]. Available: http://arxiv.org/abs/1508.00341

[43] N. Jain *et al.*, "Device calibration impacts security of quantum key distribution," *Phys. Rev. Lett.*, vol. 107, p. 110501, Sep. 2011.

[44] V. Makarov and D. R. Hjelme, "Faked states attack on quantum cryptosystems," *J. Mod. Opt.*, vol. 52, no. 5, pp. 691–705, 2005.

[45] V. Makarov, *et al.* "Creation of backdoors in quantum communications via laser damage," *Phys. Rev. A*, vol. 94, no. 3, pp. 030302, 2016.

[46] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, "Controlling an actively-quenched single photon detector with bright light," *Opt. Exp.*, vol. 19, no. 23, p. 23590–23600, 2011.

[47] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *J. Mod. Opt.*, vol. 51, nos. 9–10, pp. 1267–1288, 2004.

[48] P. A. Hiskett *et al.*, "Performance and design of ingaas/inp photodiodes for single-photon counting at 1.55 μm," *Appl. Opt.*, vol. 39, no. 36, p. 6818, 2000.

[49] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. New York, NY, USA: Wiley, 2007.

Anqi Huang was born in Zixing, China. She received the B.Sc. and M.Sc. degrees from the College of Electronics Science and Engineering, National University of Defense Technology, in 2012 and 2014, respectively. Since 2014, she has been pursuing the Ph.D. degree in quantum information with the Institute for Quantum Computing, University of Waterloo, Canada. Her research interest includes practical security evaluation of quantum communication.

Shihan Sajeed was born in Dhaka, Bangladesh. He received the B.Sc. and M.Sc. degrees from the Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka, in 2007 and 2009, respectively. Since 2013, he has been pursuing the Ph.D. degree in quantum information with the Institute for Quantum Computing, University of Waterloo, Canada. He was a Faculty Member with the Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka, from 2010 to 2013. His research interest includes the security of quantum communication.

Poompong Chaiwongkhot was born in Nakhonsawan, Thailand. He received the B.S. degree in physics from Chiang Mai University, Thailand, in 2013, and the M.S. degree in physics (quantum information) from the University of Waterloo, Canada, in 2015. He is currently pursuing the Ph.D. degree in physics with the University of Waterloo. His research interest is around the security verification of practical quantum cryptography system.

Mathilde Soucarros received the M.Eng. degree in electronics and computer engineering from the National Institute of Applied Sciences of Rennes, France, the M.Eng. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2009, and the Ph.D. degree in mathematics from Joseph Fourier University, Grenoble, in 2012, with a focus on random number generators. Since 2012, she has been with ID Quantique SA, Geneva, Switzerland, where she is involved in the development of quantum and classical cryptographic devices and on topics related to random number generators.

Matthieu Legré received the master's degree in physics, with specialization in optic, from the Institut d'Optique Graduate School, Paris, France, and the Ph.D. degree. He was with the Group of Applied Physics, Geneva, where he was involved in thesis under the supervision of Prof. N. Gisin, Geneva, in 2001. He was with ID Quantique R&D team in 2007, where he was involved in optical fibre metrology and quantum physics. He is responsible of all development aspects linked to physical issues. He held a postdoctoral position with the Group of Applied for a two years, where he managed the section of optical metrology.

Vadim Makarov was born in Leningrad, USSR, in 1974. He received the B.S. and M.S. degrees in radiophysics from St. Petersburg State Polytechnical University, Russia, in 1998, and the Ph.D. degree in quantum cryptography from the Norwegian University of Science and Technology, Trondheim, in 2007. He was a Post-Doctoral Researcher with the Pohang University of Science and Technology, South Korea, from 2007 to 2008, and Trondheim from 2008 to 2011. Since 2012, he has been a Research Assistant Professor with the Quantum Hacking Laboratory, Institute for Quantum Computing, University of Waterloo, Canada. His research interests center around the practical security of quantum cryptography systems, and technology of free-space and satellite-based quantum communications.

Appendix E

Experimental quantum key distribution with source flaws

Experimental quantum key distribution with source flaws

Feihu Xu,[1,*] Kejin Wei,[1,2] Shihan Sajeed,[3,4] Sarah Kaiser,[3,5] Shihai Sun,[6] Zhiyuan Tang,[1] Li Qian,[1]
Vadim Makarov,[3,4,5] and Hoi-Kwong Lo[1]

[1] *Centre for Quantum Information and Quantum Control (CQIQC), Department of Electrical & Computer Engineering and Department of Physics, University of Toronto, Toronto, Ontario M5S 3G4, Canada*

[2] *School of Science and State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China*

[3] *Institute for Quantum Computing (IQC), University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

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

[5] *Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

[6] *College of Science, National University of Defense Technology, Changsha 410073, China*

(Received 15 May 2015; revised manuscript received 8 July 2015; published 4 September 2015)

Decoy-state quantum key distribution (QKD) is a standard technique in current quantum cryptographic implementations. Unfortunately, existing experiments have two important drawbacks: the state preparation is assumed to be perfect without errors and the employed security proofs do not fully consider the finite-key effects for general attacks. These two drawbacks mean that existing experiments are not guaranteed to be proven to be secure in practice. Here, we perform an experiment that shows secure QKD with imperfect state preparations over long distances and achieves rigorous finite-key security bounds for decoy-state QKD against coherent attacks in the universally composable framework. We quantify the source flaws experimentally and demonstrate a QKD implementation that is tolerant to channel loss despite the source flaws. Our implementation considers more real-world problems than most previous experiments, and our theory can be applied to general discrete-variable QKD systems. These features constitute a step towards secure QKD with imperfect devices.

DOI: 10.1103/PhysRevA.92.032305 PACS number(s): 03.67.Dd, 03.67.Hk, 42.50.Ex

I. INTRODUCTION

Quantum key distribution (QKD), offering information-theoretic security in communication, has aroused great interest among both scientists and engineers [1]. The most important question in QKD is its security. This fact has finally been proven based on the laws of quantum mechanics [2,3]. However, for real-life implementations that are mainly based on attenuated laser pulses, the occasional production of multiphotons and channel loss make QKD vulnerable to various subtle attacks [4]. Fortunately, the decoy-state method [5] has solved this security issue and dramatically improved the performance of QKD with faint lasers. Several experimental groups have demonstrated that decoy-state BB84 is secure and feasible under real-world conditions [6–12]. As a result, the decoy-state method has become a standard technique in many current QKD implementations [13].

Until now, however, decoy-state QKD experiments [6–13] have had two important drawbacks. The first one is that in the key rate formula of all existing experiments, it is commonly assumed that the phase and polarization encoding is done *perfectly* without errors. Thus the state preparation is assumed to be basis independent. That is, the density matrices for the two conjugate bases are assumed to be the same. This is a highly unrealistic assumption and may mean that the key generation is actually *not* proven to be secure in previous QKD experiments [6–13].

What if we use a key rate formula that takes imperfect encodings into account? The standard Gottesman-Lo-Lütkenhaus-Preskill (GLLP) security proof [3] (see also [14]) does allow one to do so. Unfortunately, GLLP formalism is very conservative in assuming that the dimensionality of the prepared states is unbounded. Then the eavesdropper (Eve) could perform an unambiguous-state-discrimination (USD) attack [15]. Consequently, the secret key rate will be reduced substantially (e.g., a commercial system is secure below 10-km fiber only). We remark that source flaws are a serious concern, not only in decoy-state BB84 but also in other quantum information processing protocols [16,17].

To address the source flaw problem, Tamaki *et al.* put forward a theoretical proposal—the loss-tolerant protocol [18]—that outperforms GLLP analysis significantly. The loss-tolerant protocol considers a realistic situation where the dimension of the prepared states is bounded to two (which we call a *qubit assumption*). Then it is impossible for Eve to perform the USD attack. Eve's information can be bounded from the rejected-data analysis (i.e., using the basis-mismatch events to bound the phase error rate) proposed in [19]. Nevertheless, Ref. [18] is only valid in the asymptotic limit with unlimited resources. The practicality of the loss-tolerant protocol remains unknown.

Recently, though an elegant proposal has implied that Eve's information can be bounded without monitoring signal disturbance [20], source flaw was still not considered in the theory and experiment [21]. Therefore, all previous QKD experiments ignore the source flow problem, and all papers addressing this problem are theoretical. For these reasons, until now, the feasibility of long-distance QKD implementations with imperfect encodings has remained undemonstrated.

The second drawback in previous experiments [6–13] is that the finite-key security claims were made with the assumption that Eve was restricted to particular types of attacks

[*] Present address: Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; feihu.xu@utoronto.ca

(e.g., collective attacks). Unfortunately, such assumptions cannot be guaranteed in practice. Very recently, based on the frameworks proposed in [22,23], Hayashi *et al.* and Lim *et al.* independently provide, for the first time, tight and rigorous security bounds against general quantum attacks (i.e., coherent attacks) for decoy-state QKD [24,25]. Nonetheless, a QKD experiment that implements such an advanced theory has yet to be completed.

In this work, we present experimental realization of the loss-tolerant protocol [18] and the finite-key analysis [24]. By modifying a commercial plug&play QKD system, we experimentally show that with imperfect source encodings, it is still able to perform secure QKD over long distances. In particular, with our security analysis, we successfully generate secure keys over different channel lengths, up to 50-km telecom fibers. In contrast, not even a single bit of secure key can be extracted with GLLP security proof. We note in passing that our experiment requires only three encoding states. Thus it can simplify conventional BB84 implementations.[1] Moreover, we study how to apply the finite-key analysis of [24,25] in real implementations. We generate secure keys that can be secure against coherent attacks in the universally composable framework [26]. Our implementation, security analysis, and parameter estimation procedure can be applied to general discrete-variable QKD systems. Our results break ground for future QKD experiments with imperfect sources.

The rest of this paper is organized as follows. We introduce the protocol in Sec. II. In Sec. III, we present the security analysis. In Sec. IV, we present the decoy-state analysis for parameter estimation. In Sec. V, we verify the qubit assumption. In Sec. VI, we present our experimental setup and experimental results. Finally, we conclude this paper in Sec. VII.

II. PROTOCOL

The loss-tolerant protocol is a general method that works not only for the standard BB84 protocol, but even for the three-state protocol [27] where there is a strong asymmetry between the two bases. The three-state QKD runs almost the same as BB84, except that (i) Alice sends Bob only three pure states $\{|0_z\rangle, |0_x\rangle, |1_z\rangle\}$, where $|i_j\rangle$ ($i \in \{0,1\}$ and $j \in \{\mathsf{Z}, \mathsf{X}\}$) denotes the state associated with bit "i" in j basis, and (ii) the rejected data (i.e., the detection events when Alice and Bob use different basis) are used for the estimation of the phase error rate [19].

Here we focus on the three-state protocol and consider an asymmetric coding, where the secret key is extracted only from the events whereby Alice and Bob both choose the Z basis. To implement the loss-tolerant protocol, we extend it to a general practical setting with finite keys and finite decoy states. The concrete description of the different steps of our protocol is presented below.

a. Transmission. Alice chooses a bit value uniformly at random, selects a basis choice $\lambda \in \{\mathsf{Z},\mathsf{X}\}$ with probabilities $P_\lambda \in \{P_\mathsf{Z}, P_\mathsf{X}\}$, and an intensity choice $k \in \{\mu,\nu,\omega\}$ ({signal, decoy, vacuum}) with probabilities $P_k \in \{P_\mu, P_\nu, P_\omega\}$. Finally, she prepares a phase-randomized weak coherent pulse, chosen from three states $\{|0_\mathsf{Z}\rangle, |0_\mathsf{X}\rangle, |1_\mathsf{Z}\rangle\}$, where $|i_\lambda\rangle$ denotes the state associated with bit "i" in λ basis, and sends it to Bob via the quantum channel.

b. Detection. Bob chooses a basis from $\{\mathsf{Z},\mathsf{X}\}$ with probabilities $\{P_\mathsf{Z}, P_\mathsf{X}\}$ and measures the pulses. Then he records the detection or nondetection, his basis choice, and the measured bit value. (For double clicks, he assigns a random bit value.)

c. Basis reconciliation. Alice and Bob announce their basis and intensity choices over an authenticated public channel. Then they decide the number of the detected pulses (gain counts) $n_{\lambda,k}$, when both Alice and Bob use basis λ for intensity k.

d. Parameter estimation. First, Alice and Bob announce the bit information for all the pulses that are detected in X by Bob. Second, they compute: (i) the number of error pulses $n_{e_\mathsf{x},k}$ where both Alice and Bob use X and they obtain the disagreement bit values; (ii) the number of basis-mismatch pulses $n_{i_\mathsf{x}|j_z,k}$ where Bob detects the pulse in X and obtains the bit value i, given that Alice prepares bit j in Z basis. Third, according to the formulas shown in Table I, they calculate $s^{\mathrm{L}}_{\mathsf{X},0}$, $s^{\mathrm{L}}_{\mathsf{X},1}$, and $e^{\mathrm{U}}_{\mathsf{X},1}$, which are the lower bound of vacuum events, the lower bound of single-photon events, and the upper bound of the phase error rate, associated with the single-photon events in Z basis, respectively.

e. Error correction and verification. Alice and Bob reveal $\mathrm{leak}_{\mathrm{EC}} = n_{\mathsf{Z},\mu} f_e h(e_{\mathsf{Z},\mu})$ bit of information to perform an error correction step that can correct errors for the expected quantum bit error rate (QBER) e_Z. (f_e is the error correction inefficiency function that is chosen as 1.16 in this paper.) To ensure that they share a pair of identical keys with $\varepsilon_{\mathrm{cor}}$ correct [23], they perform an error-verification step using two universal hash functions that publish $\lceil \log_2 1/\varepsilon_{\mathrm{cor}} \rceil$ bits of information [28].

f. Privacy amplification. Using the results from steps d and e, Alice and Bob estimate the sacrificed bit length S_{PA} [see Eq. (1)] [24,25] and apply a universal hash function to their corrected strings to produce the final secret key of length ℓ [see Eq. (2)].

III. SECURITY ANALYSIS

We first define the security criteria that we are using [29]. For some small errors, $\varepsilon_{\mathrm{cor}},\varepsilon_{\mathrm{sec}} > 0$, we say that our protocol is $\varepsilon_{\mathrm{cor}} + \varepsilon_{\mathrm{sec}}$-secure if it is $\varepsilon_{\mathrm{cor}}$-correct and $\varepsilon_{\mathrm{sec}}$-secret. The former is satisfied if the secret keys are identical except with a small probability $\varepsilon_{\mathrm{cor}}$. The latter is satisfied if $\|\rho_{\mathrm{AE}} - U_\mathrm{A} \otimes \rho_\mathrm{E}\|_1/2 \leqslant \varepsilon_{\mathrm{sec}}$, where ρ_{AE} is the classical-quantum state describing the joint state of $\mathbf{S}_\mathrm{A}$ and $\mathbf{E}$, and U_A is the uniform mixture of all possible values of $\mathbf{S}_\mathrm{A}$. Importantly, this secrecy criterion guarantees that the protocol is universally composable: the pair of secret keys can be safely used in any cryptographic task [29].

The secrecy analysis is based on the framework of [23], which was extended to the case with decoy states [24]. We use the entropic uncertainty relations to establish bounds on the smooth min-entropy of the raw key conditioned on Eve's information. Conditional on passing the checks in the error-verification step, the sacrificed bit length S_{PA} [25] in privacy

[1] For those free-space systems based on four laser diodes, one could simply keep one laser just as backup in case a certain laser fails, without any decrease in performance.

TABLE I. Concrete descriptions and formulas for the parameter estimation.

Definitions:

λ: basis choice, $\lambda \in \{\mathsf{Z},\mathsf{X}\}$.
k: intensity choice, $k \in \{\mu,\nu,\omega\}$ ({signal,decoy,vacuum}).
P_λ: probability choice for basis λ, $P_\lambda \in \{P_\mathsf{Z},(1-P_\mathsf{Z})\}$.
P_k: probability choice for intensity k, $P_k \in \{P_\mu,P_\nu,P_\omega\}$.
δ_l: phase modulation errors for $l \in \{1,2,3\}$, see Eq. (6).

Measured quantities:

$n_{\lambda,k}$: the number of the detected pulses–both Alice and Bob use basis λ for intensity k.
$n_{e_\mathsf{X},k}$: the number of error pulses–both Alice and Bob use X for intensity k and they obtain the disagreement bit values.
$n_{i_\mathsf{X}|j_z,k}$: the number of basis-mismatch pulses–Bob detects the pulse in X and obtains the bit value i, given that Alice prepares bit j in Z for intensity setting k ($i,j \in \{0,1\}$).

Statistical fluctuations:

Δ: statistics [31], $\Delta(n_z,\varepsilon_1) = \sqrt{n_z/2\ln(1/\varepsilon_1)}$.
$n^\mathrm{U}_{z,k}$: the upper bound of $n_{z,k}$, $n^U_{z,k} = n_{z,k} + \Delta(n_{z,k},\varepsilon_1)$.
$n^\mathrm{L}_{z,k}$: the lower bound of $n_{z,k}$, $n^L_{z,k} = n_{z,k} - \Delta(n_{z,},\varepsilon_1)$.
τ_n: n-photon-state probability, $\tau_n = \sum_{k\in\{\mu,\nu,\omega\}} P_k e^{-k}k^n/n!$.

Decoy-estimation results:

$s^\mathrm{L}_{z,0}$: the lower bound of vacuum events–Eq. (4).
$s^\mathrm{L}_{z,1}$: the lower bound of single-photon events–Eq. (5).
$e^\mathrm{U}_{\mathsf{X},1}$: the upper bound of the phase error rate–Eq. (7).

amplification (PA) is given by [24]

$$S_\mathrm{PA} = n_{z,\mu} - s^\mathrm{L}_{z,0} - s^\mathrm{L}_{z,1}\left[q - h\left(e^\mathrm{U}_{x,1}\right)\right] + 6\log_2\frac{26}{\varepsilon_\mathrm{sec}}, \quad (1)$$

where $h(x)$ is the binary entropy function, q is the maximum fidelity for states prepared in the Z and X basis, which characterizes the quality of the source [23], and ε_sec is the secret level that can be guaranteed by PA (i.e., ε_sec-secret [26]). $\{s^\mathrm{L}_{\mathsf{Z},0},\ s^\mathrm{L}_{\mathsf{Z},1},\ e^\mathrm{U}_{\mathsf{X},1}\}$ can be calculated from the measured quantities of $\{n_{\mathsf{Z},k},\ n_{e_\mathsf{X},k},\ n_{i_\mathsf{X}|j_z,k}\}$, and the concrete formulas for such calculations are summarized in Sec. IV.

Finally, the ε_sec-secret key length in the Z basis is given by

$$\ell \geqslant s^\mathrm{L}_{\mathsf{Z},0} + s^\mathrm{L}_{\mathsf{Z},1}\left[q - h\left(e^\mathrm{U}_{\mathsf{X},1}\right)\right] - \mathrm{leak}_\mathrm{EC} - 6\log_2\frac{26}{\varepsilon_\mathrm{sec}} - \log_2\frac{2}{\varepsilon_\mathrm{cor}}, \quad (2)$$

with an overall security level $\varepsilon_\mathrm{tot} = \varepsilon_\mathrm{sec} + \varepsilon_\mathrm{cor}$. Here, following the analysis in Appendix B of [24], the secret level is given by

$$\varepsilon_\mathrm{sec} = 2[\alpha_2 + \alpha_3] + \overline{\nu} + 21\varepsilon_1. \quad (3)$$

To get the secret level given in Eq. (2), we set each error term to a common value ε; thus $\varepsilon_\mathrm{sec} = 26\varepsilon$.

With ℓ, the secret key rate (per optical pulse) is given by $R^\mathrm{L} = \ell/N$ with N denoting the total number of signals (optical pulses) sent by Alice.

IV. PARAMETER ESTIMATION

Our decoy-state analysis for parameter estimation builds on [24]. Our contribution is estimating the phase error rate $e^\mathrm{U}_{x,1}$ by incorporating source flaws. In decoy-state BB84, $e^\mathrm{U}_{x,1}$ is estimated from the counts in X basis [24]. In the loss-tolerant protocol [18], however, $e^\mathrm{U}_{x,1}$ is estimated from the rejected counts, i.e., considering the detection events associated with single photons when Alice and Bob use *different* bases. Moveover, our estimation focuses directly on the detection *counts* announced by Bob, which is different from previous analysis that is based on detection probabilities [5,30]. The results are summarized in Table I.

A. Lower bounds of vacuum counts and single-photon counts

In the original decoy-state method [5,30], Alice first randomly chooses an intensity setting (signal state or decoy state) to modulate each laser pulse and then she announces her intensity choices after Bob's detections. One can imagine a *virtual* but equivalent protocol: *Alice has the ability to first send n-photon states and then she decides only on the choice of intensity after Bob has a detection*. Let $s_{z,n}$ be the number of detection counts observed by Bob given that Alice sends n-photon states in Z basis. Note that $\sum_{n=0}^{\infty} s_{z,n} = n_z$ is the total number of detections (gain counts). In the asymptotic limit with two decoy states, we have

$$\hat{n}_{z,k} = \sum_{n=0}^{\infty} P_{k|n} s_{z,n}, \quad \forall k \in \{\mu,\nu,\omega\},$$

where $P_{k|n}$ is the conditional probability of choosing the intensity k given that Alice prepares an n-photon state. For finite-data size, from Hoeffding's inequality [31], the experimental measurement $n_{z,k}$ satisfies

$$|\hat{n}_{z,k} - n_{z,k}| \leqslant \Delta(n_z,\varepsilon_1),$$

with probability at least $1-2\varepsilon_1$, where $\Delta(n_z,\varepsilon_1) = \sqrt{n_z/2\ln(1/\varepsilon_1)}$ and $\hat{n}_{z,k}$ is the expected value of $n_{z,k}$. Note that our analysis considers the most *general* type of attack—a joint attack—consistent with quantum memories. The above equation allows us to establish a relation between the asymptotic values and the observed statistics. Specifically,

$$\hat{n}_{z,k} \leqslant n_{z,k} + \Delta(n_z,\varepsilon_1) = n^\mathrm{U}_{z,k},$$
$$\hat{n}_{z,k} \geqslant n_{z,k} - \Delta(n_z,\varepsilon_1) = n^\mathrm{L}_{z,k},$$

are respectively the upper and lower bound of the gain counts $n_{z,k}$ for a given intensity setting $k \in \{\mu,\nu,\omega\}$.

An analytical lower bound on $s_{z,0}$ can be established by exploiting the structure of the conditional probabilities $P_{k|n}$ based on Bayes' rule: $P_{k|n} = \frac{P_k}{\tau_n}\frac{e^{-k}k^n}{n!}$, where $\tau_n = \sum_{k\in\{\mu,\nu,\omega\}} P_k e^{-k}k^n/n!$ is the probability that Alice prepares an n-photon state. Based on an estimation method in [30], we have

$$s_{z,0}^{\mathrm{L}} = \frac{\tau_0}{(\nu-\omega)}\left(\frac{\nu e^{\omega} n_{z,\omega}^{\mathrm{L}}}{P_\omega} - \frac{\omega e^{\nu} n_{z,\nu}^{\mathrm{U}}}{P_\nu}\right), \tag{4}$$

$$s_{z,1}^{\mathrm{L}} = \frac{\mu\tau_1}{\mu(\nu-\omega)-(\nu^2-\omega^2)}\left[\frac{e^{\nu} n_{z,\nu}^{\mathrm{U}}}{P_\nu} - \frac{e^{\omega} n_{z,\omega}^{\mathrm{L}}}{P_\omega} + \frac{\nu^2-\omega^2}{\mu^2}\left(\frac{s_{z,0}^{\mathrm{L}}}{\tau_0} - \frac{e^{\mu} n_{z,\mu}^{\mathrm{U}}}{P_\mu}\right)\right]. \tag{5}$$

B. Upper bound of phase error rate

In the asymptotic case, we follow [18] to estimate the phase error rate. The details are shown in Appendix A. Here we extend [18] to the finite-key case.

We focus on phase encoding BB84 and assume $\{\delta_1, \delta_2, \delta_3\}$ to be Alice's phase modulation errors for $\{\pi/2, \pi, 3\pi/2\}$; thus the four BB84 imperfect states sent by Alice are given by

$$\begin{aligned} |\phi_{0_z}\rangle &= |0_z\rangle, \\ |\phi_{1_z}\rangle &= \sin\delta_2|0_z\rangle + \cos\delta_2|1_z\rangle, \\ |\phi_{0_x}\rangle &= \cos\delta_1|0_x\rangle + \sin\delta_1|1_x\rangle, \\ |\phi_{1_x}\rangle &= \sin\delta_3|0_x\rangle + \cos\delta_3|1_x\rangle. \end{aligned} \tag{6}$$

After considering the finite-data analysis, $e_{x,1}^{U}$ is given by

$$e_{x,1}^{\mathrm{U}} = \frac{s_{0_x|1_x,1}^{vir,\mathrm{U}} + s_{1_x|0_x,1}^{vir,\mathrm{U}}}{s_{0_x|0_x,1}^{vir,\mathrm{L}} + s_{0_x|1_x,1}^{vir,\mathrm{L}} + s_{1_x|0_x,1}^{vir,\mathrm{L}} + s_{1_x|1_x,1}^{vir,\mathrm{L}}}. \tag{7}$$

Here

$$\begin{bmatrix} P_z s_{0_x|j_x,1}^{vir,\mathrm{U}} \\ P_z s_{1_x|j_x,1}^{vir,\mathrm{U}} \end{bmatrix} = B \times A^{-1} \begin{bmatrix} 2P_x s_{j_x|0_z,1}^{\mathrm{U}} \\ 2P_x s_{j_x|1_z,1}^{\mathrm{U}} \\ P_z s_{j_x|0_x,1}^{\mathrm{U}} \end{bmatrix}, \tag{8}$$

$$\begin{bmatrix} P_z s_{0_x|j_x,1}^{vir,\mathrm{L}} \\ P_z s_{1_x|j_x,1}^{vir,\mathrm{L}} \end{bmatrix} = B \times A^{-1} \begin{bmatrix} 2P_x s_{j_x|0_z,1}^{\mathrm{L}} \\ 2P_x s_{j_x|1_z,1}^{\mathrm{L}} \\ P_z s_{j_x|0_x,1}^{\mathrm{L}} \end{bmatrix}, \tag{9}$$

where P_z and P_x are the probabilities that Alice and Bob choose the Z and X basis, $j \in \{0,1\}$, and A and B are given by

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -\cos(2\delta_2) & \sin(2\delta_2) \\ 1 & \sin(2\delta_1) & \cos(2\delta_1) \end{bmatrix}, \quad B = \frac{1}{12}\begin{bmatrix} (1+\sin\delta_2) & \sin\delta_2(1+\sin\delta_2) & \cos\delta_2(1+\sin\delta_2) \\ (1-\sin\delta_2) & -\sin\delta_2(1-\sin\delta_2) & -\cos\delta_2(1-\sin\delta_2) \end{bmatrix}. \tag{10}$$

$s_{j_x|i_z,1}^{\mathrm{U}}$ ($s_{j_x|i_z,1}^{\mathrm{L}}$) denotes the upper (lower) bound of single-photon events when Bob has detections associated with bit "j" in the X basis, *given that* Alice sends a state of i_z with $i \in \{0,1\}$.

$s_{j_x|i_z,1}^{\mathrm{L}}$ and $s_{j_x|0_x,1}^{\mathrm{L}}$ can be estimated equivalently by plugging $\{n_{j_x|i_z,k}^{\mathrm{L}}, n_{j_x|i_z,k}^{\mathrm{U}}\}$ and $\{n_{j_x|0_x,k}^{\mathrm{L}}, n_{j_x|0_x,k}^{\mathrm{U}}\}$ into Eqs. (4) and (5). $s_{j_x|i_z,1}^{\mathrm{U}}$ and $s_{j_x|0_x,1}^{\mathrm{U}}$ can be estimated by

$$\begin{aligned} s_{j_x|i_z,1}^{\mathrm{U}} &= \tau_1 \frac{n_{j_x|i_z,\nu}^{\mathrm{U}} - n_{j_x|i_z,\omega}^{\mathrm{L}}}{\nu-\omega}, \\ s_{j_x|0_x,1}^{\mathrm{U}} &= \tau_1 \frac{n_{j_x|0_x,\nu}^{\mathrm{U}} - n_{j_x|0_x,\omega}^{\mathrm{L}}}{\nu-\omega}. \end{aligned} \tag{11}$$

V. VERIFYING QUBIT ASSUMPTION

The qubit assumption is normally required in the security proofs [1,2] to simplify the analysis. With the qubit assumption, using large deviation techniques (e.g., quantum de Finetti theorem), one can show that Eve can effectively apply only the same superoperator on each transmitted qubit. This greatly simplifies the security proofs. In practice, however, *no* previous works have verified this assumption in practice. Note that a specific attack to exploit the higher dimensionality of state preparation has been proposed in [32]. Here we perform a comprehensive analysis to theoretically verify the qubit assumption (with high accuracy) in a practical QKD system, even with device imperfections. These results are shown in Appendix D.

VI. EXPERIMENT

We implement the protocol, presented in Sec. II, with a modified commercial ID-500 plug&play QKD system, manufactured by ID Quantique (see Fig. 1) [33,34]. Nonetheless, we remark that our methods of parameter optimizations, finite-key analysis, the quantification of phase modulation errors, and the implementation can also be applied to standard QKD systems. Here, we use the plug&play QKD system simply as an example to illustrate our *general* methods.

A. Setup

The initial plug&play system employs the phase-coding QKD scheme and it works as follows (see Fig. 1) [34]. Bob first sends two laser pulses (i.e., signal and reference pulse) to Alice. Alice uses the reference pulse as a synchronization signal (detected by her classical photodetector) to activate her phase modulator (PM). Then Alice modulates the phase of the signal pulse only, attenuates the two pulses to single-photon level, and sends them back to Bob. Bob randomly chooses his measurement basis by modulating the phase of the returning reference pulse and detects the interference signals with his two single-photon detectors (SPDs).

Our modifications on top of ID-500 are as follows. To implement the decoy-state protocol, we add two acousto-optic modulators (AOMs, Brimrose) to achieve polarization-insensitive intensity modulation. AOM$_1$—driven by a waveform with random pattern generated from a function generator

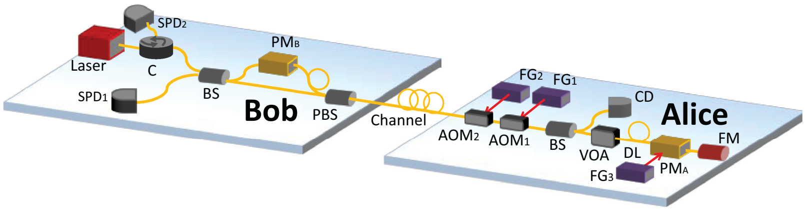


FIG. 1. (Color online) Experimental setup. SPD$_1$/SPD$_2$, single-photon detector; C, circulator; PM$_A$/PM$_B$, phase modulator; BS, beam splitter; PBS, polarization beam splitter; CD, classical photodetector; VOA, variable optical attenuator; AOM$_1$/AOM$_2$, acousto-optic modulator; FG, function generator; DL, delay line; FM, Faraday mirror. PM$_A$ randomly selects a phase from $\{0,\pi/2,\pi\}$ for the three-state modulations. AOM$_1$ randomly modulates the intensity of each pulse to be either signal-state level or decoy-state level, while AOM$_2$ compensates the phase shift due to AOM$_1$.

(FG$_1$, Agilent 88250A)—is used for the decoy modulation, while AOM$_2$—driven by a fixed waveform generated from FG$_2$—is used to compensate the phase shift caused by the frequency shift of the AOM [6]. To implement the three-state protocol, we adopt another FG, i.e., FG$_3$ in Fig. 1, to control Alice's PM. FG$_1$ and FG$_3$ are loaded with random numbers generated from a quantum random number generator [35]. We have measured the system parameters as shown in Table II.

B. Quantifying modulation error

We quantify the modulation error δ_θ in the source through calibrating Alice's PM, a LiNbO$_3$ waveguide-based electro-optical modulator, on two plug&play QKD systems—ID 500 and Clavis2 [34]. δ_θ is defined as the difference between the actual phase and the expected phase $\theta \in \{0, \pi/2, \pi \quad 3\pi/2\}$. We find that in ID-500, the voltages $\{0,\ 0.30V_m,\ 0.62V_m,\ 0.92V_m\}$ modulate the expected phases $\{0,\ \pi/2,\ \pi\ ,\ 3\pi/2\}$, where $V_m \approx 3.67$ V is a maximal value allowed on Alice's PM. The calibration process is as follows. Alice is directly connected to Bob with a short fiber (about 1 m), Alice scans the voltages applied to her PM, Bob sets his own PM at a fixed unmodulated phase $\{0\}$, and then records the detection counts of his two SPDs. These counts are denoted by $D_{1,\theta}$ and $D_{2,\theta}$. The detection counts on ID-500 and Clavis2 are shown in Table III.

In ID-500, to quantify δ_θ, we first determine the detector efficiencies (η_{d1}, η_{d2}) and the dark count rates ($Y_{0,d1}, Y_{0,d2}$) for Bob's two SPDs and find that $\eta_{d1} = 5.05\%$ and $\eta_{d2} = 4.99\%$ and $Y_{0,d1} \approx Y_{0,d2} = 4.01\times10^{-5}$. In Table III, $D_{1,0}$ quantifies the amount of global misalignment between Alice and Bob (i.e., the summation of the dark counts and the imperfect visibility). This global misalignment can increase QBER, but it is irrelevant to bound Eve's information in the loss-tolerant protocol [18]. Only the relative orientation between the three states prepared by Alice quantifies the source flaws that can be potentially exploited by Eve. Hence, we subtract $D_{1,0}$ in the quantification of δ_θ. In our analysis of the statistics, we use Hoeffding's inequality [31] to guarantee the definition of composable security. The upper bound of δ_θ is then given by

$$\delta_\theta \leqslant \bar{\delta}_\theta = \left|\theta - 2\arctan\left(\sqrt{\frac{[(D_{1,\theta}+\Delta(D_{1,\theta},\varepsilon)] - [D_{1,0}-\Delta(D_{1,0},\varepsilon)]/\eta_{d1}}{[(D_{2,\theta}-\Delta(D_{2,\theta},\varepsilon)] - [D_{1,0}+\Delta(D_{1,0},\varepsilon)]/\eta_{d2}}}\right)\right|, \tag{12}$$

where $\Delta(D_{i,\theta},\varepsilon)$=$\sqrt{D_{i,\theta}/2\ln(1/\varepsilon)}$ (with $i \in \{0,1\}$) [31]. In general, if $Y_{0,d1} \neq Y_{0,d2}$ in a practical system, in Eq. (12), we can use $D_{i,\theta}$ to subtract the dark counts of detector d_i. Here, we choose a failure probability $\varepsilon = 10^{-10}$ (i.e., a confidence level $1 - 2\times10^{-10}$). The upper bounds of δ_θ are shown in Table III. From this table, the error δ in ID-500 is upper bounded by the case of δ_π, i.e., $\delta \leqslant \bar{\delta}_\pi = 0.134$.

TABLE II. Parameters measured in an ID-500 commercial QKD system, including laser wavelength λ, optical misalignment error e_d (the probability that a photon hits the erroneous detector), Bob's overall quantum efficiency η_{Bob}, dark count rate per pulse Y_0 for each detector, and system repetition rate f.

λ	e_d	η_{Bob}	Y_0	f
1551.71 nm	2.35%	5.05%	4.01×10^{-5}	5 MHz

Using the same method for Clavis2, we find that δ is upper bounded by $\delta \leqslant \bar{\delta}_\pi = 0.145$. Notice that δ can also be

TABLE III. Raw counts and modulation errors for Alice's phase modulator in ID-500 and Clavis2 commercial plug&play systems. $D_{1,\theta}$ ($D_{2,\theta}$) represents the detections counts of SPD$_1$ (SPD$_2$). $\bar{\delta}_\theta$, given by Eq. (12), is the upper bound of modulation error for a given phase θ.

System	θ	$D_{1,\theta}$	$D_{2,\theta}$	$\bar{\delta}_\theta$
ID-500	0	630	867678	–
	$\pi/2$	456735	444336	0.013
	π	856245	4744	0.134
	$3\pi/2$	464160	436962	0.030
Clavis2	0	727	1075320	–
	$\pi/2$	546724	527735	0.023
	π	1111574	6990	0.145
	$3\pi/2$	566813	531417	0.037

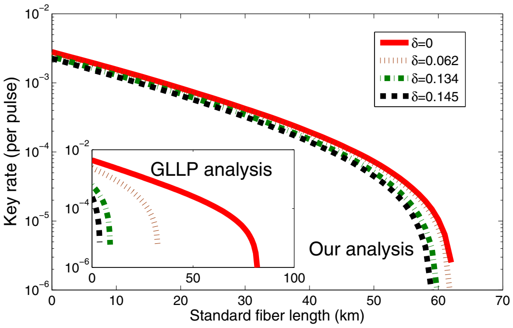


FIG. 2. (Color online) Practical key rates with parameters of Table II, $N = 5\times10^{10}$ and $\varepsilon_{tot} = 10^{-10}$. The main figure is for our analysis, while the inset figure is for the decoy-state BB84 with the GLLP security analysis. With GLLP, the maximal distance for our ID-500 system is about 9 km (green dashed-dotted curve in the inserted figure). In contrast, our analysis can substantially outperform GLLP in that the ID-500 system can be made secure over 60 km and the secure key rate is almost the same as the case without considering source flaws (i.e., assuming $\delta = 0$).

estimated using the interference visibility or the extinction ratio of the PM [36]. In a system with an advanced phase-stabilized interferometer [37], the value of $\delta \leqslant 0.062$ corresponds to about 99.9% visibility or a 30-dB extinction ratio.

C. Numerical evaluation

With δ_θ and the parameters in Table II, Fig. 2 shows the simulation results, where we choose the total number of pulses $N = 5\times10^{10}$ and the security level $\varepsilon_{\rm tot} = 10^{-10}$. We use the model proposed in [30] to simulate the virtual data. For comparison, this figure also includes the key rate for the decoy-state BB84 based on the GLLP security analysis (see Appendix B for the model). The power of our security analysis is explicitly shown by the fact that GLLP delivers a key rate that decreases rapidly when δ increases. The maximal tolerant distance is about 9 km for our QKD system. Our security analysis, however, can substantially outperform GLLP. Our QKD setup can be made secure over 60 km, and the secure key rate is almost the same as the case without source flaws. Using simulation, we also determine the implementation parameters to achieve the optimal system performance. The optimized parameters are shown in Table IV.

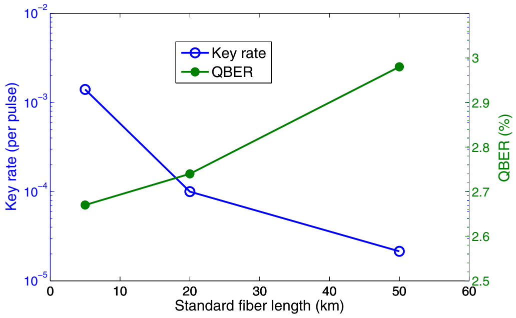


FIG. 3. (Color online) Experimental secret key rates (blue circle) and QBERs (green dot) over fiber lengths of 5, 20, and 50 km.

D. Experimental results

In our demonstration, we implement the loss-tolerant protocol in the finite-key regime over standard fiber lengths (L) of 5, 20, and 50 km. In the 5- and 20-km experiments, we performed a real decoy-state QKD implementation with optimized parameters. We use FG1 to randomly modulate the signal and decoy states and use FG3 to randomly modulate the three states of $\{|0_z\rangle, |0_x\rangle, |1_z\rangle\}$. In the 50-km experiment, we removed the two AOMs due to their high loss (over 3 dB each) and used the variable optical attenuator (VOA) in Alice to modulate the decoy intensities for a proof-of-concept decoy-state modulation.

Our measurement and postprocessing are different from previous experiments in that we directly measure the detection *counts* instead of the so-called gains (i.e., probabilities) [6–12], and we also record the basis-mismatch counts. In the 5-km and 20-km experiments, we chose to operate the system for a few hours and collected about 75 sets of data, with each set of about 104.5 million pulses, which corresponds to a total number of pulses $N = 7.84\times10^9$. In the 50-km experiment, we collected about 500 sets of data and sent a total number of $N = 5.23\times10^{10}$ pulses. The details of the experimental counts are shown in Appendix C.

In our analysis of experimental data, we consider a security level $\varepsilon_{tot} = 10^{-10}$. With δ_θ, we find that $q = 0.79$. By plugging the experimental counts into the decoy-state estimations and using Eq. (2), we obtain the experimental results listed in Table IV and Fig. 3. The system's QBER is below 3%. Based on the loss-tolerant analysis, a secure key rate (per optical

TABLE IV. Implementation parameters and experimental results. N is the total number of pulses sent by Alice. P_μ, P_ν are the probabilities to choose different intensities. $P_{\rm z}$ is the probability to choose the Z basis. ω equals about 0.001 for 5- and 50-km experiments, and it equals about 0.003 for 20-km experiment. The estimation results are obtained by plugging the experimental counts into the decoy-state estimation equations (see Table I). The key rate is obtained from Eq. (2).

Channel		Parameters						Estimation			Performance		
L (km)	Attn (dB)	N	μ	ν	P_μ	P_ν	$P_{\rm z}$	$s^L_{\rm z,0}$	$s^{\rm L}_{\rm z,1}$	$e^{\rm U}_{\rm x,1}$	$e_{{\rm z},\mu}$	l	$R^{\rm L}$
5	1.4	7.84×10^9	0.41	0.05	0.64	0.27	0.70	7.40×10^4	3.02×10^7	6.28%	2.67%	1.06×10^7	1.40×10^{-3}
20	4.5	7.84×10^9	0.37	0.06	0.40	0.50	0.60	6.15×10^4	6.58×10^6	8.67%	2.74%	8.07×10^5	1.03×10^{-4}
50	10.5	5.23×10^{10}	0.55	0.06	0.74	0.18	0.50	3.36×10^5	1.33×10^7	8.46%	2.98%	1.07×10^6	2.14×10^{-5}

pulse) of 1.40×10^{-3} was generated at 5 km, while at 50 km it was 2.14×10^{-5}. Given the 5-MHz repetition rate, the key rates per second are 7 kbps and 107 bps, respectively. Over 1 kbit of unconditionally secure keys are exchanged between Alice and Bob. The security of these keys considers source flaws and satisfies the composable security definition, and it can withstand general attacks by Eve. With a state-of-the-art high-speed QKD system working at a gigahertz repetition rate, our loss-tolerant analysis can easily enable a key rate of megabits per second.

As a comparison to previous security analysis (e.g., GLLP [3]) with the source flaw $\delta = 0.134$, no matter how many decoy states we choose or how large the data size we use, the key generation rate will hit zero at only about 10 km. That is, at 20 and 50 km, using previous GLLP security proof, not even a single bit could be shared between Alice and Bob with guaranteed security. This means that if considering source flaws in previous long-distance decoy-state experiments [6–13], the key generation might *not* be proven to be secure. In contrast, our analysis can easily achieve a high secure key generation rate over long distances, even in the presence of source flaws.

VII. CONCLUSION

We have demonstrated decoy-state QKD with imperfect state preparations and employed tight finite-key security bounds with composable security against coherent attacks. Our experiment demonstrates that the perfect state-preparation assumption can be removed, and it is still able to perform QKD over long distances. In our paper, we ignore certain imperfections in the source such as the intensity fluctuations of signal and decoy states, which have a small effect and can be taken care of using previous results [25]. Moreover, it will be interesting to consider the source flaw problem in the new protocol of [20]. Future research can also combine our results with measurement-device-independent QKD [16] to remove the security loopholes, both in the source and in the detectors.

Note added in proof. Recently, we noticed a paper which addresses the finite-key effect of the loss-tolerant protocol [38]. In contrast to our present manuscript, that paper is strictly theoretical.

ACKNOWLEDGMENTS

We thank O. Bedroya, M. Curty, M. Jiang, C. Lim, N. Lütkenhaus, M. Legré, X Ma, G. Ribordy, and particularly, M. Lucamarini and K. Tamaki for valuable discussions. Support from NSERC, the CRC program, the Connaught Innovation Fund, Industry Canada, the OGS Visa Award, Air Force Office of Scientific Research (AFOSR), CryptoWorks21, U.S. Office of Naval Research, a Mike & Ophelia Lazaridis Fellowship, the National Natural Science Foundation of China (Grants No. 61178010 and No. 11304391), and the China Scholarship Council (No. 201406470051) is gratefully acknowledged.

APPENDIX A: PHASE ERROR RATE IN THE ASYMPTOTIC CASE

We follow [18] to estimate the phase error rate. To make our paper self-contained, we present the main results from [18] in this section. For simplicity, we consider three pure states, described in Eq. (6). The density matrices for the three states $|\phi_{0_z}\rangle$, $|\phi_{1_z}\rangle$, $|\phi_{0_x}\rangle$ are as follows:

$$\rho_{0_z} = |\phi_{0_z}\rangle\langle\phi_{0_z}| = (I + \sigma_z)/2, \tag{A1}$$

$$\rho_{1_z} = |\phi_{1_z}\rangle\langle\phi_{1_z}| = \begin{bmatrix} \sin^2\delta_2 & \sin\delta_2\cos\delta_2 \\ \sin\delta_2\cos\delta_2 & \cos^2\delta_2 \end{bmatrix} = \frac{1}{2}I - \frac{1}{2}\cos(2\delta_2)\sigma_z + \frac{1}{2}\sin(2\delta_2)\sigma_x, \tag{A2}$$

$$\rho_{0_x} = |\phi_{0_x}\rangle\langle\phi_{0_x}| = \frac{1}{2}\begin{bmatrix} 1+\sin(2\delta_1) & \cos(2\delta_1) \\ \cos(2\delta_1) & 1-\sin(2\delta_1) \end{bmatrix} = \frac{1}{2}I + \frac{1}{2}\sin(2\delta_1)\sigma_z + \frac{1}{2}\cos(2\delta_1)\sigma_x. \tag{A3}$$

Here $\sigma_{x,y,z}$ denote Pauli matrices and I is the identity matrix. The equivalent entanglement states between Alice and Bob are [18]

$$|\Psi_z\rangle = \big(|0_z\rangle|\phi_{0_z}\rangle + |1_z\rangle|\phi_{1_z}\rangle\big)\big/\sqrt{2} \quad |\Psi_x\rangle = |0_x\rangle|\phi_{0_x}\rangle. \tag{A4}$$

Let $Y^{\omega}_{s_\beta,j_\alpha}$ with $\omega \in \{Z,X\}$ and $s,j \in \{0,1\}$ denote the joint probability that Alice (Bob) obtains a bit value $j(s)$ conditional on the state preparation of $|\Psi_\omega\rangle$ and her (his) basis choice α (β); then the joint probabilities for different states are [18]

$$Y^z_{s_x,0_z} = \tfrac{2}{6}\mathrm{Tr}\big[D_{sx}\sigma^z_{B,0_z}\big] = \tfrac{1}{6}\mathrm{Tr}\big[D_{sx}\rho_{0_z}\big] = (q_{s_x|I} + q_{s_x|z})/6, \tag{A5}$$

where $\sigma^z_{B,0_z} = \mathrm{Tr}_A[|0_z\rangle\langle 0_z| \otimes I|\Psi_z\rangle\langle\Psi_z|] = \frac{1}{2}|\phi_{0_z}\rangle\langle\phi_{0_z}|$, and $q_{s_x|(I,x,z)} = \mathrm{Tr}[D_{s_x}\sigma_{I,x,z}]/2$;

$$Y^z_{s_x,1_z} = \tfrac{2}{6}\mathrm{Tr}\big[D_{sx}\sigma^z_{B,1_z}\big] = \tfrac{1}{6}\mathrm{Tr}[D_{sx}\rho_{1_z}] = [q_{s_x|I} - \cos(2\delta_2)q_{s_x|z} + \sin(2\delta_2)q_{s_x|x}]/6, \tag{A6}$$

where $\sigma^z_{B,1_z} = \mathrm{Tr}_A[|1_z\rangle\langle 1_z| \otimes I|\Psi_z\rangle\langle\Psi_z|] = \frac{1}{2}|\phi_{1_z}\rangle\langle\phi_{1_z}|$;

$$Y^x_{s_x,0_x} = \tfrac{1}{6}\mathrm{Tr}\big[D_{sx}\sigma^x_{B,0_x}\big] = \tfrac{1}{6}\mathrm{Tr}\big[D_{sx}\rho_{0_x}\big] = \big[q_{s_x|I} + \sin(2\delta_1)q_{s_x|z} + \cos(2\delta_1)q_{s_x|x}\big]\big/6, \tag{A7}$$

where $\sigma^x_{B,0_x} = \mathrm{Tr}_A[|0_x\rangle\langle 0_x| \otimes I|\Psi_x\rangle\langle\Psi_x|] = \frac{1}{2}|\phi_{0_x}\rangle\langle\phi_{0_x}|$.

Equations (A5)–(A7) can be rewritten as

$$\begin{bmatrix} Y^z_{s_x,0_z} \\ Y^z_{s_x,1_z} \\ Y^x_{s_x,0_x} \end{bmatrix} = \frac{1}{6}\begin{bmatrix} Y^z_{s_x|0_z} \\ Y^z_{s_x|1_z} \\ Y^x_{s_x|0_x} \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -\cos(2\delta_2) & \sin(2\delta_2) \\ 1 & \sin(2\delta_1) & \cos(2\delta_1) \end{bmatrix}\begin{bmatrix} q_{s_x|I} \\ q_{s_x|z} \\ q_{s_x|x} \end{bmatrix} \equiv \frac{1}{6}A\begin{bmatrix} q_{s_x|I} \\ q_{s_x|z} \\ q_{s_x|x} \end{bmatrix}. \tag{A8}$$

Here $Y^z_{s_x|0_z}$ denotes the conditional probability that Bob obtains bit s in basis x given that Alice sends 0_z. The same definition is applied to $Y^z_{s_x|1_z}$ and $Y^x_{s_x|0_x}$. Note that all these quantities can be measured *directly* in experiment.

To estimate the phase error rate, we consider a *virtual* protocol: Alice first prepares $|\Psi_z\rangle$ and then both Alice and Bob measure systems A and B in the X basis [18]. The joint probabilities of the virtual states $Y^{z,vir}_{s_x,j_x}$ are

$$\begin{aligned} Y^{z,vir}_{s_x,0_x} &= \tfrac{1}{12}\mathrm{Tr}\big[D_{s_x}\sigma^{z,vir}_{B,0_x}\big] = \tfrac{1}{3}\big[(1+\sin\delta_2)q_{s_x|I} + \sin\delta_2(1+\sin\delta_2)q_{s_x|x} + \cos\delta_2(1+\sin\delta_2)q_{s_x|x}\big], \\ Y^{z,vir}_{s_x,1_x} &= \tfrac{1}{12}\mathrm{Tr}\big[D_{s_x}\sigma^{z,vir}_{B,1_x}\big] = \tfrac{1}{3}\big[(1-\sin\delta_2)q_{s_x|I} - \sin\delta_2(1-\sin\delta_2)q_{s_x|x} - \cos\delta_2(1-\sin\delta_2)q_{s_x|x}\big]. \end{aligned} \tag{A9}$$

Equation (A9) can then be rewritten as

$$\begin{bmatrix} Y^{z,vir}_{s_x,0_x} \\ Y^{z,vir}_{s_x,1_x} \end{bmatrix} = \frac{1}{12}\begin{bmatrix} (1+\sin\delta_2) & \sin\delta_2(1+\sin\delta_2) & \cos\delta_2(1+\sin\delta_2) \\ (1-\sin\delta_2) & -\sin\delta_2(1-\sin\delta_2) & -\cos\delta_2(1-\sin\delta_2) \end{bmatrix}\begin{bmatrix} q_{s_x|I} \\ q_{s_x|z} \\ q_{s_x|x} \end{bmatrix} \equiv B\begin{bmatrix} q_{s_x|I} \\ q_{s_x|z} \\ q_{s_x|x} \end{bmatrix}. \tag{A10}$$

Combining it with Eq. (A8), we can obtain the rate of virtual states based on experimental results, which is

$$\begin{bmatrix} Y^{z,vir}_{s_x,0_x} \\ Y^{z,vir}_{s_x,1_x} \end{bmatrix} = B \times A^{-1}\begin{bmatrix} Y^z_{s_x|0_z} \\ Y^z_{s_x|1_z} \\ Y^x_{s_x|0_x} \end{bmatrix}. \tag{A11}$$

Finally, the phase error can be estimated by

$$e_x = \frac{Y^{z,vir}_{1_x,0_x} + Y^{z,vir}_{0_x,1_x}}{Y^{z,vir}_{0_x,0_x} + Y^{z,vir}_{1_x,0_x} + Y^{z,vir}_{0_x,1_x} + Y^{z,vir}_{1_x,1_x}}. \tag{A12}$$

The extended result of Eq. (A12) for the finite-data case is presented in Eq. (4) of the main text.

APPENDIX B: GLLP SECURITY ANALYSIS WITH SOURCE FLAWS

We discuss the standard GLLP security analysis for BB84 with source flaws [3,36], which is used for our simulation of Fig. 2.

Based on GLLP for imperfect sources, the $\varepsilon_{\mathrm{sec}}$ secret key length is similar to the key formula [i.e., Eq. (1)] in the main text, except for the phase error rate, which includes the correction due to basis-dependent flaws and is revised to [3]

$$\bar{e}^U_{x,1} \leqslant e^U_{x,1} + 4\Delta' + 4\sqrt{\Delta' e^U_{x,1}} + \epsilon_{ph}. \tag{B1}$$

Here, Δ' is called the balance of a quantum coin [3,36] and quantifies the basis-dependent flaws of Alice's signals associated with single-photon events. Δ' is given by [3]

$$\Delta' \leqslant \frac{\Delta}{Y_1}, \quad \Delta = \frac{1 - F(\rho_z,\rho_x)}{2}, \tag{B2}$$

where Y_1 (typically called the yield of single photons [5,30]) is the frequency of successful detections associated with single photons, and $F(\rho_z,\rho_x)$ is the fidelity of the density matrices for the Z and X basis. Using Eq. (6), we can easily calculate $F(\rho_z,\rho_x)$ given $\{\delta_1,\delta_2,\delta_3\}$. In our QKD system, with $\{\delta_1,\delta_2,\delta_3\}$ upper bounded by 0.127, we have $F(\rho_z,\rho_x) = 1 - 1.9\times10^{-3}$. So, from Eq. (B2), $\Delta = 9.45\times10^{-4}$.

In GLLP analysis, the imperfect fidelity $F(\rho_z,\rho_x)$ can be enhanced in principle by Eve via exploiting the channel loss, which is clearly shown in Eq. (B2), i.e., Δ is enhanced to Δ'. Combined with the decoy-state estimations discussed in [24], we can derive the key length and obtain the inset curves in Fig. 2.

APPENDIX C: EXPERIMENTAL COUNTS

In Table V we list the raw experimental counts for each distance. Note that, in the experiment results,

$$n_{1_x|0_x,k} = n_{e_x,k}, \quad n_{0_x|0_x,k} = n_{x,k} - n_{e_x,k}.$$

In the 5- and 20-km experiments, we collected about 75 sets of data, with each set of about 104.5 million pulses sent out by Alice. This corresponds to a total number of pulses $N = 7.84\times10^9$. In the 50-km experiment, we collected about 500 sets of data and sent a total number of $N = 5.23\times10^{10}$ pulses. The experimental gain counts ($n_{z,k}$, $n_{x,k}$), error counts ($n_{e_z,k}$, $n_{e_x,k}$), and rejected counts ($n_{0_x|z,k}, n_{1_x|z,k}$) are listed in the table.

APPENDIX D: QUBIT ASSUMPTION AND ITS VERIFICATION

We verify the qubit assumption, i.e., that the four BB84 states remain in two dimensions. This assumption is commonly made in various QKD protocols including decoy-state BB84 and MDI-QKD. We focus on a standard *one-way phase-encoding* system, which has been widely implemented in experiments [7,10–12]. In this system, a LiNbO$_3$ waveguide-based phase modulator (PM) is commonly

TABLE V. Experimental raw counts.

Distance	$n_{z,\mu}$	$n_{z,\nu}$	$n_{z,\omega}$	$n_{x,\mu}$	$n_{x,\nu}$	$n_{x,\omega}$
5 km	7.84×10^{7}	2.23×10^{6}	2.60×10^{4}	7.17×10^{6}	4.08×10^{5}	4.70×10^{3}
20 km	8.09×10^{6}	1.50×10^{6}	2.71×10^{4}	3.40×10^{6}	6.31×10^{5}	1.36×10^{4}
50 km	2.01×10^{7}	6.94×10^{5}	4.81×10^{4}	2.06×10^{6}	7.10×10^{5}	4.82×10^{4}
	$n_{e_{z,\mu}}$	$n_{e_{z,\nu}}$	$n_{e_{z,\omega}}$	$n_{e_{x,\mu}}$	$n_{e_{x,\nu}}$	$n_{e_{x,\omega}}$
5 km	1.01×10^{6}	6.40×10^{4}	6.80×10^{3}	1.32×10^{5}	1.25×10^{4}	1.76×10^{3}
20 km	2.22×10^{5}	6.13×10^{4}	6.78×10^{3}	5.67×10^{4}	2.68×10^{4}	2.65×10^{3}
50 km	5.98×10^{5}	8.46×10^{4}	2.28×10^{4}	6.40×10^{5}	8.89×10^{4}	2.23×10^{4}
	$n_{0_x\|0_z,\mu}$	$n_{0_x\|0_z,\nu}$	$n_{0_x\|0_z,\omega}$	$n_{1_x\|0_z,\mu}$	$n_{1_x\|0_z,\nu}$	$n_{1_x\|0_z,\omega}$
5 km	3.83×10^{6}	2.47×10^{5}	3.30×10^{3}	4.16×10^{6}	2.32×10^{5}	2.40×10^{3}
20 km	1.36×10^{6}	2.39×10^{5}	4.56×10^{3}	1.34×10^{6}	2.2×10^{5}	4.59×10^{3}
50 km	0.57×10^{7}	1.63×10^{5}	1.10×10^{4}	0.56×10^{7}	1.76×10^{5}	1.26×10^{4}
	$n_{0_x\|1_z,\mu}$	$n_{0_x\|1_z,\nu}$	$n_{0_x\|1_z,\omega}$	$n_{1_x\|1_z,\mu}$	$n_{1_x\|1_z,\nu}$	$n_{1_x\|1_z,\omega}$
5 km	3.83×10^{6}	2.46×10^{5}	3.31×10^{3}	4.15×10^{6}	2.32×10^{5}	2.41×10^{3}
20 km	1.37×10^{6}	2.38×10^{5}	4.57×10^{3}	1.34×10^{6}	2.21×10^{5}	4.60×10^{3}
50 km	0.58×10^{7}	1.62×10^{5}	1.11×10^{4}	0.56×10^{7}	1.77×10^{5}	1.25×10^{4}

used to encode/decode phase information. Figure 4 illustrates the schematic of such a PM [39]. For commercial products, see [40]. To guarantee the qubit assumption, Alice's PM is supposed to have the same timing, spectral, spatial, and polarization mode information for different BB84 states. We find that timing and spatial information can be easily guaranteed without any additional devices, while spectral and polarization information can also be guaranteed with standard low-cost optical devices such as a wavelength filter and polarizer. Therefore, based on standard devices, we can verify the qubit assumption with high accuracy. We remark that our method serves as a specific example to practically verify the qubit assumption. In the future, it will be interesting to work toward constructing a more general theory on the verification of the qubit assumption.

In the following, we discuss timing, spectral, spatial, and polarization properties for different encoding phases.

1. Temporal-spectral mode

Temporal mode. Figure 4 shows the schematic of the phase modulation based on LiNbO$_3$ crystal. When the PM modulates different phases, the electro-optical effect inside the LiNbO$_3$ waveguide changes the principal refractive index n_z. At first sight, it might appear that the timing information is indeed changed for different phase modulations. However, we show that such a change is so small that it can be neglected.

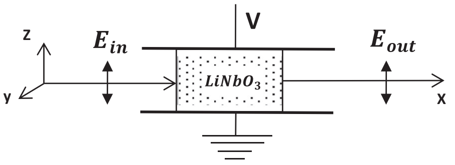


FIG. 4. Schematic of an electro-optic phase modulator based on LiNbO$_3$ crystal [39]. Commercial products can be seen in [40]. The double-headed arrows show the direction of polarization of the optical beam. The crystal is cut in a configuration so that the applied electrical field (voltage) is along the direction of the principal (z) axis. To take the advantage of the largest electro-optical coefficient in the z axis, an optical beam is propagating along the x axis, with the direction of polarization parallel to the z axis.

According to the EM theory in LiNbO$_3$ waveguides, the relations among the principal refractive index n_z, the group refractive index n_g, and the extraordinary refractive index n_e are given by [39]

$$n_g = n_z + \omega_0 \frac{dn_z(\omega)}{d\omega}\bigg|_{\omega_0}, \quad n_z = n_e - \frac{1}{2} n_e^3 r_z \frac{V}{d}, \tag{D1}$$

where ω_0 is the central frequency of the optical field, r_z is the electro-optical coefficient along the z axis, V is the voltage applied onto the crystal, and d is the thickness of the crystal. Thus the timing difference Δt between $\{0\}$ and the phase modulation $\{\pi\}$ is given by

$$\Delta t = \left[\frac{1}{2} n_e^3 r_z \frac{V_\pi}{d} + \frac{3}{2} n_e^2 r_z \frac{V_\pi}{d} \omega_0 \frac{dn_e(\omega)}{d\omega}\bigg|_{\omega_0}\right] \frac{l_0}{c}, \tag{D2}$$

where $V_\pi = \frac{\chi_0 d}{n_e^3 r_z l_0}$ is the half-wave voltage that provides a phase modulation $\{\pi\}$ [39], l_0 is the length of the crystal, and c is the speed of light.

For a typical LiNbO$_3$ crystal working in the telecom wavelength $\chi_0 \sim 1550$ nm, it is well known that the relation between n_e and λ_0 is given by [41]

$$n_e^2 = 1 + \frac{2.980\lambda_0^2}{\lambda_0^2 - 0.020} + \frac{0.598\lambda_0^2}{\lambda_0^2 - 0.067} + \frac{8.954\lambda_0^2}{\lambda_0^2 - 416.08}. \tag{D3}$$

Notice that in a waveguide-based PM, one has to use the effective index, i.e., n_{eff}, to include the waveguide effect. We remark, however, that for LiNbO$_3$ material, n_{eff} and n_e are almost the same [42]. Hence, by plugging Eq. (D3) into Eq. (D2), we have $\Delta t \approx 4\times10^{-6}$ ns. In a QKD implementation, the optical pulse is typically around 1-ns width [7–9] or 0.1 ns [10–12], and thus $\Delta t \ll 0.1$ ns. Assuming that the optical pulse is Gaussian, Δt corresponds to a fidelity of $F(\rho^0,\rho^\pi) \approx 1$–$10^{-8}$ between $\{0\}$ and $\{\pi\}$. Therefore, timing remains (almost) the same for different phase modulations.

Spectral mode. First, in a standard one-way system, Alice can locally synchronize the devices so that the optical pulse passes through Alice's PM in the middle of the electrical modulation signal (flat response). Hence the optical pulse experiences a correct modulation *without* spectral change [43,44]. In a two-way system, Alice can monitor the timing information between the signal and reference pulse to guarantee the correct modulation and defend against side-channel attacks [43,44]. Second, to guarantee a single spectral mode from the output of a laser, one can use a standard wavelength filter. For instance, a recent QKD experiment used an off-the-shelf wavelength filter with a FWHM of $\Delta\nu = 15$ GHz for a different purpose [12]. In this case, given a Gaussian pulse with FWHM $\Delta t = 0.1$ ns in the time domain [12], it is quite close to the lower bound of time-bandwidth product [39], i.e., $\Delta t \times \Delta\nu \geqslant \frac{2ln2}{\pi}$. Wavelength filters with narrow bandwidth have already been widely available on the market [45]. Hence a single spectral mode can be guaranteed with high accuracy by using a wavelength filter.

2. Spatial mode

For a standard single-mode fiber (SMF), the core diameter is around 10 μm. Theory and experiments have already confirmed that a SMF in the telecom wavelength rejects all high-order modes and conducts only one fundamental transverse mode [46]. The cutoff wavelength of a standard SMF is about 1260 nm.[2] Using the software of BeamPROP, we have also performed a numerical simulation with a standard multimode fiber propagating into a SMF. The results show that after only about 1 mm, SMF rejects almost all high-order modes. The high-order modes decay exponentially; thus after about 10 mm, there is no high-order component left (less than 10^{-10} proportion). Notice that the input of a standard commercial PM usually has a certain length of pigtail fiber (about 1 m) [40]. Therefore the single-mode assumption on spatial mode can be easily guaranteed in practice.

3. Polarization mode

The input of a commercial PM is normally a pigtail of polarization-maintaining fiber [40], which can ensure that the input polarization is perfectly aligned with the principal axis of PM. Experimentally, before this polarization-maintaining fiber, one can use a fiber polarization beam splitter (PBS) to reject other polarization modes. A standard PBS has about a 30-dB extinction ratio. In the following, we discuss the error due to this finite extinction ratio (30 dB). Ideally, if the PBS has an infinite extinction ratio, the input state is perfectly aligned with the principal axis (z axis in Fig. 4) and Alice modulates the four BB84 states as

$$|\phi_j\rangle = \frac{1}{\sqrt{2}}(e^{ij\frac{\pi}{2}}|S_z\rangle + |R_z\rangle),$$

where $j \in \{0,1,2,3\}$ denotes the four BB84 states and $|S_z\rangle$ ($|R_z\rangle$) denotes the signal (reference) pulse with polarization along the z axis. However, due to the finite extinction ratio of PBS, the signal and reference pulse are expressed as

$$|S\rangle = \alpha|S_y\rangle + \beta|S_z\rangle, \quad |R\rangle = \alpha|R_y\rangle + \beta|R_z\rangle,$$

where $|S_y\rangle$ denotes the polarization component along the y axis. For a 30-dB extinction ratio, $\alpha^2 \approx 0.001$. Thus Alice's imperfect modulations can be described by

$$|\phi'_j\rangle = \tfrac{1}{\sqrt{2}}(\alpha e^{ij\frac{\pi}{6}}|S_y\rangle + \beta e^{ij\frac{\pi}{2}}|S_z\rangle + \alpha|R_y\rangle + \beta|R_z\rangle), \quad \text{(D4)}$$

where we assume that the relative modulation magnitude ratio between the polarization aligned with the principal axis (z axis) and the orthogonal polarization (y axis in Fig. 4) is 1:3 [39,43]. Using three new bases $\{|e_1\rangle, |e_2\rangle, |e_3\rangle\}$, Eq. (D4) can be written as (similar to [32])

$$\begin{aligned}|\phi'_j\rangle = \frac{1}{\sqrt{2}}\big[&\alpha\beta\big(e^{ij\frac{\pi}{6}} - e^{ij\frac{\pi}{2}}\big)|e_1\rangle \\ &+\big(\alpha^2 e^{ij\frac{\pi}{6}} + \beta^2 e^{ij\frac{\pi}{2}}\big)|e_2\rangle + |e_3\rangle\big]. \quad \text{(D5)}\end{aligned}$$

Hence the four imperfect states are spanned to three dimensions in Hilbert space, i.e., the information encoded by Alice is not only in the time-phase mode but also in the polarization mode. However, for a 30-dB extinction ratio, we find that it is almost impossible for Eve to attack the system, because the fidelity between $|\phi_j\rangle$ and $|\phi'_j\rangle$, $F(\rho^{|\phi_j\rangle},\rho^{|\phi'_j\rangle}) = \mathrm{tr}\,(\sqrt{\sqrt{\rho^{|\phi_j\rangle}}\rho^{|\phi'_j\rangle}\sqrt{\rho^{|\phi_j\rangle}}})$, is about $1 - 10^{-7}$ for $j \in \{0,1,2,3\}$. This shows that the imperfect states are highly close to the perfect BB84 states. Most importantly, one can derive a refined security proof to include this small imperfection into the secure key rate formula, which will be a subject of future investigation.

[2] See, for instance, Corning's SMF28; http://www.corning.com/docs/opticalfiber.

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002); V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, *ibid.* **81**, 1301 (2009); H.-K. Lo, M. Curty, and K. Tamaki, Nat. Photon. **8**, 595 (2014).

[2] D. Mayers, J. Assoc. Comput. Mach. **48**, 351 (2001); H.-K. Lo and H. Chau, Science **283**, 2050 (1999); P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).

[3] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **4**, 325 (2004).

[4] B. Huttner, N. Imoto, N. Gisin, and T. Mor, Phys. Rev. A **51**, 1863 (1995); G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000).

[5] 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).

[6] Y. Zhao, B. Qi, X. Ma, H.-K. Lo, and L. Qian, Phys. Rev. Lett. **96**, 070502 (2006).

[7] D. Rosenberg, J. W. Harrington, P. R. Rice, P. A. Hiskett, C. G. Peterson, R. J. Hughes, A. E. Lita, S. W. Nam, and J. E. Nordholt, Phys. Rev. Lett. **98**, 010503 (2007).

[8] T. Schmitt-Manderbach, H. Weier, M. Fürst, R. Ursin, F. Tiefenbacher, T. Scheidl, J. Perdigues, Z. Sodnik, C. Kurtsiefer, J. G. Rarity *et al.*, Phys. Rev. Lett. **98**, 010504 (2007).

[9] C. Z. Peng, J. Zhang, D. Yang, W. B. Gao, H. X. Ma, H. Yin, H. P. Zeng, T. Yang, X. B. Wang, and J. W. Pan, Phys. Rev. Lett. **98**, 010505 (2007).

[10] A. R. Dixon, Z. L. Yuan, J. F. Dynes, A. W. Sharpe, and A. J. Shields, Opt. Express **16**, 18790 (2008).

[11] M. Lucamarini, K. A. Patel, J. F. Dynes, B. Fröhlich, A. W. Sharpe, A. R. Dixon, Z. L. Yuan, R. V. Penty, and A. J. Shields, Opt. Express **21**, 24550 (2013).

[12] K. A. Patel, J. F. Dynes, M. Lucamarini, I. Choi, A. W. Sharpe, Z. L. Yuan, R. V. Penty, and A. J. Shields, Appl. Phys. Lett. **104**, 051123 (2014).

[13] S. Nauerth, F. Moll, M. Rau, C. Fuchs, J. Horwath, S. Frick, and H. Weinfurter, Nat. Photon. **7**, 382 (2013); J.-Y. Wang *et al.*, *ibid.* **7**, 387 (2013); B. Frohlich *et al.*, Nature (London) **501**, 69 (2013); A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, Phys. Rev. Lett. **111**, 130501 (2013); Y. Liu *et al.*, *ibid.* **111**, 130502 (2013); T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. C. doAmaral, G. P. Temporao, and J. P. von der Weid, Phys. Rev. A **88**, 052303 (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.*, *ibid.* **113**, 190501 (2014).

[14] Ø. Marøy, L. Lydersen, and J. Skaar, Phys. Rev. A **82**, 032337 (2010); E. Woodhead and S. Pironio, *ibid.* **87**, 032315 (2013).

[15] M. Dušek, M. Jahma, and N. Lütkenhaus, Phys. Rev. A **62**, 022306 (2000).

[16] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012); M. Curty, F. Xu, W. Cui, C. C. W. Lim, K. Tamaki, and H.-K. Lo, Nat. Commun. **5**, 3732 (2014); F. Xu, M. Curty, B. Qi, and H.-K. Lo, IEEE JSTQE **21**, 6601111 (2015).

[17] G. Berlín *et al.*, Nat. Commun. **2**, 561 (2011); A. Pappa *et al.*, *ibid.* **5**, 3717 (2014); V. Dunjko, E. Kashefi, and A. Leverrier, Phys. Rev. Lett. **108**, 200502 (2012).

[18] K. Tamaki, M. Curty, G. Kato, H.-K. Lo, and K. Azuma, Phys. Rev. A **90**, 052314 (2014).

[19] S. M. Barnett, B. Huttner, and S. Phoenix, J. Mod. Opt. **40**, 2501 (1993).

[20] T. Sasaki, Y. Yamamoto, and M. Koashi, Nature (London) **509**, 475 (2014); Z. Zhang, X. Yuan, Z. Cao, and X. Ma, arXiv:1505.02481.

[21] J.-Y. Guan, Z. Cao, Y. Liu, G. L. Shen-Tu, J. S. Pelc, M. M. Fejer, C. Z. Peng, X. Ma, Q. Zhang, and J. W. Pan, Phys. Rev. Lett. **114**, 180502 (2015).

[22] M. Hayashi and T. Surumaru, New J. Phys. **14**, 093014 (2012).

[23] M. Tomamichel, C. C.-W. Lim, N. Gisin, and R. Renner, Nat. Commun. **3**, 634 (2012).

[24] C. C.-W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, Phys. Rev. A **89**, 022307 (2014).

[25] M. Hayashi and R. Nakayama, New J. Phys. **16**, 063009 (2014).

[26] M. Ben-Or, M. Horodecki, D.-W. Leung, D. Mayers, and J. Oppenheim, in *Theory of Cryptography* (Springer, New York, 2005), pp. 386–406; R. Renner and R. König, in *Theory of Cryptography* (Springer, New York, 2005), pp. 407–425; R. Renner, Ph.D. thesis, ETH Zurich, 2005.

[27] J. C. Boileau, K. Tamaki, J. Batuwantudawe, R. Laflamme, and J. M. Renes, Phys. Rev. Lett. **94**, 040503 (2005); C.-H. F. Fung and H. K. Lo, Phys. Rev. A **74**, 042342 (2006).

[28] M. N. Wegman and J. L. Carter, J. Comput. Syst. Sci. **22**, 265 (1981).

[29] R. Renner, Ph.D. thesis, ETH Zurich, 2005).

[30] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Phys. Rev. A **72**, 012326 (2005)

[31] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).

[32] S.-H. Sun, M.-S. Jiang, and L.-M. Liang, Phys. Rev. A **83**, 062331 (2011).

[33] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, New J. Phys. **4**, 41 (2002).

[34] ID Quantique: http://www.idquantique.com/

[35] F. Xu, B. Qi, X. Ma, H. Xu, H. Zheng, and H.-K. Lo, Opt. Express **20**, 12366 (2012).

[36] K. Tamaki, H. K. Lo, C.-H. F. Fung, and B. Qi, Phys. Rev. A **85**, 042307 (2012).

[37] T. Honjo, K. Inoue, and H. Takahashi, Opt. Lett. **23**, 2797 (2004).

[38] A. Mizutani, M. Curty, C. C. W. Lim, N. Imoto, and K. Tamaki, arXiv:1504.08151.

39] A. Yariv and P. Yeh, *Photonics: Optical Electronics in Modern Communications* (Oxford University Press, Oxford, UK, 2007).

[40] For instance, EO-space: http://www.eospace.com; JDSU: http://www.jdsu.com

[41] D. E. Zelmon, D. L. Small, and D. Jundt, J. Opt. Soc. Am. B **14**, 3319 (1997).

[42] T. Suhara and M. Fujimura, *Waveguide Nonlinear-Optic Devices*, Springer Series in Photonics Vol. 11 (Springer-Verlag, Berlin, 2003).

[43] F. Xu, B. Qi, and H.-K. Lo, New J. Phys. **12**, 113026 (2012).

[44] M.-S. Jiang, S.-H. Sun, C.-Y. Li, and L.-M. Liang, J. Mod. Opt. **61**, 147 (2014).

[45] http://www.afwtechnologies.com.au; http://www.yenista.com

[46] D. K. Mynbaev and L. L. Scheiner, *Fiber-Optic Communications Technology* (Prentice Hall, Englewood Cliffs, NJ, 2001).

Appendix F

Experimental quantum fingerprinting with weak coherent pulses

ARTICLE

Received 18 Mar 2015 | Accepted 26 Sep 2015 | Published 30 Oct 2015 DOI: 10.1038/ncomms9735 OPEN

Experimental quantum fingerprinting with weak coherent pulses

Feihu Xu[1,2,3,⋆,†], Juan Miguel Arrazola[4,5,⋆], Kejin Wei[1,2,3,6], Wenyuan Wang[1,2,3,7], Pablo Palacios-Avila[4,5,8], Chen Feng[9], Shihan Sajeed[4,10], Norbert Lütkenhaus[4,5] & Hoi-Kwong Lo[1,2,3]

Quantum communication holds the promise of creating disruptive technologies that will play an essential role in future communication networks. For example, the study of quantum communication complexity has shown that quantum communication allows exponential reductions in the information that must be transmitted to solve distributed computational tasks. Recently, protocols that realize this advantage using optical implementations have been proposed. Here we report a proof-of-concept experimental demonstration of a quantum fingerprinting system that is capable of transmitting less information than the best-known classical protocol. Our implementation is based on a modified version of a commercial quantum key distribution system using off-the-shelf optical components over telecom wavelengths, and is practical for messages as large as 100 Mbits, even in the presence of experimental imperfections. Our results provide a first step in the development of experimental quantum communication complexity.

[1] Center for Quantum Information and Quantum Control, University of Toronto, Toronto, Ontario, Canada M5S 3H6. [2] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G4. [3] Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7. [4] Institute for Quantum Computing, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. [5] Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. [6] School of Science and State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China. [7] Department of Physics, University of Hong Kong, Pokfulam Road, Hong Kong, China. [8] Faculty of Science, National University of Engineering, Lima 15333, Peru. [9] School of Engineering, University of British Columbia, Kelowna, British Columbia, Canada V1V 1V7. [10] Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. ⋆ These authors contributed equally to this work. † Present address: Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. Correspondence and requests for materials should be addressed to N.L. (email: lutkenhaus.office@uwaterloo.ca) or to H.-K.L. (email: hklo@comm.utoronto.ca).

What technological advantages can be achieved by directly harnessing the quantum-mechanical properties of physical systems? In the context of communications, it is known that quantum mechanics enables several remarkable improvements, such as cryptographic protocols that are classically impossible[1–3], enhanced metrology schemes[4] and reductions in the communication required between distributed computing devices[5–13]. And yet, despite our advanced understanding of what these quantum advantages are, demonstrating them in a practical setting continues to be an outstanding and central challenge. Important progress has been made in this direction[14–22], but many cases of quantum improvements have never been realized experimentally.

An important example of a quantum advantage occurs in the field of communication complexity: the study of the minimum amount of information that must be transmitted to solve distributed computational tasks[5–8]. It has been proven that for several problems, quantum mechanics allows exponential reductions in communication compared with the classical case[7,9–13]. These results, besides being of great fundamental interest[6,7,23], have important practical applications for the design of communication systems, very-large-scale integration circuit design and data structures[24].

There are two types of communication complexity problems. The first one is to minimize the amount of information that must be transmitted to solve a task, and the second one is to minimize the error probability to solve a task with a fixed amount of transmitted information. These two problems are really two sides of the same coin, since any given protocol requires a certain amount of transmitted information to reach a given error probability. However, conceptually and experimentally, they belong to different regimes. To date, only a few proof-of-principle implementations of quantum communication complexity protocols have been reported[25–27]. For instance, ref. 27 was the first experiment that demonstrated an advantage of quantum over classical communication for the second problem, even without entanglement. However, all such experiments have faced daunting scalability issues, limiting their results to a quantum advantage for the second problem only, with the transmitted information restricted to single qubits. Up until now, a quantum advantage for the first problem, a reduction in the transmitted information compared with the classical case—which is the central issue in quantum communication complexity[7]—has not yet been demonstrated.

Quantum fingerprinting is arguably the most appealing protocol in quantum communication complexity, as it constitutes a natural problem for which quantum mechanics permits an exponential reduction in the transmitted information[9,28,29]. In this problem, Alice and Bob are each given an n-bit string, which we label x and y, respectively. In the simultaneous message passing model[5], they must each send a message to a third party, the referee, whose task is to decide whether the inputs x and y are equal or not with an error probability of at most ϵ. Alice and Bob do not have access to shared randomness and there is only one-way communication to the referee. It has been proven that any classical protocol for this simultaneous message passing problem must transmit at least $\Omega(\sqrt{n})$ bits of information to the referee for a desired error probability[30,31]. On the other hand, using quantum communication, Alice and Bob only need to transmit $O(\log_2 n)$ qubits of information to solve the problem with the same error probability. Therefore, for the specific goal of reducing the transmitted information, quantum communication provides an exponential improvement over the classical case[9].

Refs. 25,26 have reported heroic attempts at the implementation of quantum fingerprinting. Nonetheless, as noted already in ref. 25, a serious drawback of these approaches is that their fingerprint states must be highly entangled. As a result, even for low input sizes, the experimental requirements greatly exceed that which is possible to achieve with current technology. For this reason, the implementations of refs 25,26 are restricted to one single-qubit transmission and within a few metres, without a practical possibility of scaling them to demonstrate a reduction in the transmitted information.

In this work, we present a proof-of-concept experimental demonstration of a quantum fingerprinting system over a 5-km standard fibre operating at telecom wavelengths. The protocol is practical for input sizes as large as 100 Mbits. Crucially, our system is capable of transmitting less information than the best-known classical protocol for the fingerprinting problem. Our system is based on the quantum fingerprinting protocol with weak coherent states of ref. 29. Although this protocol is already practical, we overcome various challenges to its experimental implementation. First, we develop an efficient error-correction algorithm that allows us to substantially relax the requirements on the experimental devices and reduce the running time of the protocol. Second, we use an improved decision rule for the referee compared with the one used in ref. 29. Finally, we perform detailed simulations of the protocol that allows us to identify the appropriate parameters for performing the experiment. This enables us to run the protocol using commercial off-the-shelf components. Indeed, we implemented the protocol by using a commercial plug and play system originally designed for quantum key distribution (QKD)[32], to which we added several important modifications. We also characterized the system and showed that, within our theoretical model of the experiment, its performance is consistent with achieving the desired error probability. Finally, we experimentally tested the system for input sizes of up to 100 Mbits and obtained data that are consistent with the protocol transmitting less information than the best-known classical protocol.

Results

Coherent-state quantum fingerprinting protocol. In the quantum fingerprinting protocol of ref. 29, portrayed in Fig. 1, Alice first applies an error-correcting code (ECC) $E:\{0,1\}^n \rightarrow \{0,1\}^m$ to her input x of n bits. This results in a codeword $E(x)$ of $m = \frac{n}{R}$ bits, which she uses to prepare a sequence of m coherent states,

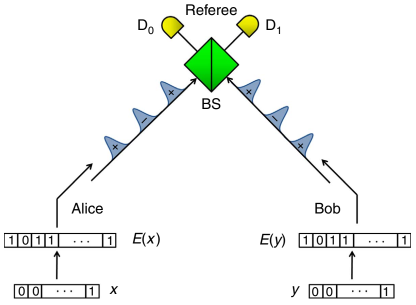


Figure 1 | A schematic illustration of the quantum fingerprinting protocol. Alice and Bob receive inputs x and y, respectively, which they feed to an ECC to produce the codewords $E(x)$ and $E(y)$. Using these codewords, they modulate the phases of a sequence of coherent pulses that they send to the referee. The incoming signals interfere at a beam splitter (BS) and photons are detected in the output using single-photon detectors D_0 and D_1. In an ideal implementation, detector D_1 fires only when the inputs to Alice and Bob are different.

where $R = \frac{n}{m} < 1$ is the rate of the code. This sequence of coherent states is given by the state

$$|\alpha, x\rangle = \bigotimes_{i=1}^{m} \left| (-1)^{E(x)_i} \frac{\alpha}{\sqrt{m}} \right\rangle_i . \tag{1}$$

Here $E(x)_i$ is the ith bit of the codeword and α is a complex amplitude. Notice that all the coherent states have the same amplitude, but their individual phases depend on the particular codeword, which in turn is determined by the input x. The total mean photon number in the entire sequence is $\mu := |\alpha|^2$, which in general depends on the length of the codewords m.

In our protocol, the encoded fingerprinting states are coherent states, instead of single-photon states as required in previous schemes[25]. Hence, a perfect two-photon interference is not required[33]. All we need is a measurement by the referee that allows her to verify whether the relative phases of the incoming pulses are equal or different. A way of achieving this consists of a phase interferometer in which the individual pulses enter a balanced beam splitter, and whenever there is a click in the output detectors, it is unambiguously revealed whether their phases are the same or not[34].

Indeed, in our scheme, Bob does the same as Alice for his input y, and they both send their sequence of states to the referee, who interferes the individual states in a balanced beam splitter. The referee checks for clicks at the outputs of the phase interferometer using single-photon detectors, which we label 'D_0' and 'D_1'. In the ideal case, a click in detector D_1 will never happen if the phases of the incoming states are equal, that is, if $E(x)_i \oplus E(y)_i = 0$. However, it is possible for a click in detector D_1 to occur if the phases are different, that is, if $E(x)_i \oplus E(y)_i = 1$. Thus, if $x \neq y$, we expect a number of clicks in D_1 that is proportional to the total mean number of photons and the Hamming distance between the codewords. This allows the referee to distinguish between equal and different inputs by simply checking for clicks in detector D_1.

In ref. 29, it was proven that the quantum information Q that can be transmitted by sending the states of equation (1) satisfies

$$Q = O(\mu \log_2 n). \tag{2}$$

For fixed μ, this corresponds to an exponential improvement over the classical case, where $\Omega(\sqrt{n})$ bits of information must be transmitted[30,31]. It is precisely in terms of this reduction in the transmitted information that the quantum protocol provides an advantage over the classical case.

The states of equation (1) can be thought of as a coherent-state version of the encoding of an m-dimensional state into the state of a single photon across m modes, as discussed in depth in ref. 35. Essentially, by fixing the total mean photon number to a constant, we are restricting ourselves to an exponentially small subspace of the larger Hilbert space associated with the optical modes, which in turn restricts the capability of these systems to transmit information. Thus, to achieve the central goal of a reduction in the transmitted information, our protocol must use a number of modes that is linear in the input size n, with the benefit that the total mean photon number μ is independent of input size and therefore very small.

Finally, we remark that a quantum protocol without entanglement or two-photon interference was demonstrated previously in ref. 27. The demonstration in ref. 27 utilized polarization qubits to tackle the communication complexity problem of maximizing the probability of solving the modulo-4 sum problem[8] with a restricted amount of transmitted information. In principle, both ref. 27 and our current paper can use coherent pulses and a phase interferometer. However, from a physical point of view, since the aims of our work and ref. 27 were different, the underlying physics also has some differences. Our protocol employed states of large dimension to encode more classical information, while ref. 27 used coherence properties of qubits, which were fixed to a two-dimensional system without interactions among the states. To use large dimensionality, we utilize time bins with phase encoding and perform an interaction of the states with a phase interferometer.

Protocol in the presence of experimental imperfections. In the presence of experimental imperfections such as detector dark counts and optical misalignment, detector D_1 may fire even when the inputs are equal. Therefore, it does not suffice to check for clicks in this detector—we must introduce a different decision rule for the referee. The decision rule proposed in ref. 29, which is based on the fraction of clicks that occur in detector D_1, is extremely sensitive to experimental imperfections. Instead, in this work we construct a better decision threshold based only on the total number of clicks observed in detector D_1.

Let $D_{1,\mathrm{E}}$ and $D_{1,\mathrm{D}}$ be random variables corresponding to the number of clicks in detector D_1 for the case of equal and worst-case different inputs, respectively. It can be shown that these distributions can be well approximated by binomial distributions $D_{1,\mathrm{E}} \sim Bin(m,p_\mathrm{E})$ and $D_{1,\mathrm{D}} \sim Bin(m,p_\mathrm{D})$, where m is the number of modes and p_E, p_D are the probabilities of observing a click in each mode for the case of equal and worst-case inputs, respectively. These probabilities are given by ref. 29:

$$p_\mathrm{E} = \left(1 - e^{-\frac{2(1-\nu)\mu}{m}}\right) + p_\mathrm{dark} \tag{3}$$

$$p_\mathrm{D} = \delta\left(1 - e^{-\frac{2\nu\mu}{m}}\right) + (1-\delta)\left(1 - e^{-\frac{2(1-\nu)\mu}{m}}\right) + p_\mathrm{dark}. \tag{4}$$

Here ν is the interference visibility—which quantifies the contrast of the interferometer—and p_dark, the dark count probability, is the probability that a detector will fire even when no incident photons from the signals are present. As before, μ is the total mean photon number in the signals and δ is the minimum distance of the ECC, which is defined as the smallest relative Hamming distance between any two distinct codewords.

The referee sets a threshold value $D_{1,\mathrm{th}}$ such that, if the number of clicks is smaller than or equal to $D_{1,\mathrm{th}}$, he will conclude that the inputs are equal. Otherwise, he concludes that they are different. Note that, unlike the ideal case, in the presence of imperfections, an error can occur even when the inputs are equal. In our protocol, the value of $D_{1,\mathrm{th}}$ is chosen in such a way that an error is equally likely to occur in both cases, so that the probability of error is given by

$$\Pr(\mathrm{error}) = \Pr(D_{1,\mathrm{E}} > D_{1,\mathrm{th}}) = \Pr(D_{1,\mathrm{D}} \leq D_{1,\mathrm{th}}), \tag{5}$$

which can be calculated directly from the distributions of $D_{1,\mathrm{E}}$ and $D_{1,\mathrm{D}}$. This is illustrated in Fig. 2. In general, for each input size n, the total mean photon number μ is uniquely determined by finding the value of μ such that $\Pr(\mathrm{error}) \leq \epsilon$, where ϵ is the desired error probability of the protocol.

Note that this model is expected to be correct as long as the parameters quantifying the experimental imperfections as well as the mean photon number μ are all constant during the run of the protocol. In practice this is not necessarily the case, so our model should be understood as an approximation of the actual performance of the system.

Finally, we note that in any implementation of the protocol there will be some loss captured by the combined effect of limited detector efficiency and channel loss. We quantify this with the single parameter $\eta < 1$. As shown in ref. 29, the effect of loss can be compensated by adjusting the total mean photon number accordingly: $\mu \rightarrow \mu/\eta$. Thus, the protocol is robust to loss.

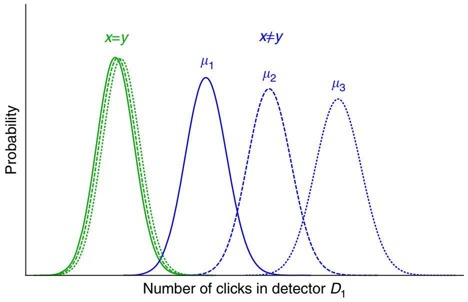


Figure 2 | An illustration of the probability distributions for the total number of clicks observed in detector D_1 for equal inputs and worst-case different inputs. The distributions are shown for three different total mean photons numbers: μ_1 (solid), μ_2 (dashed) and μ_3 (dotted), where $\mu_1 < \mu_2 < \mu_3$. For illustration, values of $\mu_1 = 617$, $\mu_2 = 693$ and $\mu_3 = 776$ were chosen for the figure. The distributions for equal inputs (green) are dominated by dark counts, so they are largely unaffected by the changes in μ. On the other hand, for the worst-case different inputs (blue), the mean value of the distributions depends strongly on μ. Therefore, the error in distinguishing both distributions can be controlled by choosing μ appropriately.

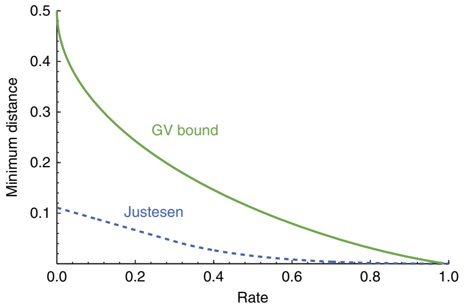


Figure 3 | The Gilbert–Varshamov bound compared with the distance-rate relationship achieved by Justesen codes. For various rates, a code satisfying the Gilbert-Varshamov bound—like the one achieved in this paper—provides a minimum distance that is more than three times the value for Justesen codes, which were used in previous works[9,29].

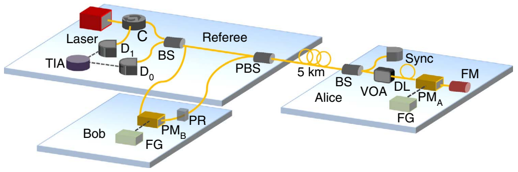


Figure 4 | Experimental set-up for quantum fingerprinting. The laser source at the referee's set-up emits photon pulses at about 1,551 nm, which are separated at a 50:50 beam splitter (BS) into two pulses, the signal pulse and the reference pulse. The signal pulse passes through Bob's phase modulator (PM) and then through a polarization rotator (PR), which rotates the pulses' polarization by 90°. The pulses are then recombined at a polarization beam splitter (PBS) where they exit through the same port and travel to Alice through the 5-km fibre. After passing through Alice's BS, the reference (forward) pulse is split into two pulses, where one is used as a synchronization (Sync) and the other one continues travelling. Similarly, the signal (backward) pulse is split into two. Then, Alice uses her PM to set the phase of the signal pulse only, according to her codeword $E(x)$. Once the reference and the signal pulses are reflected back by the Faraday mirror (FM), she attenuates them to the desired photon level by using the VOA. When the two pulses return in the direction of the referee, because of Alice's FM, the reference pulse will travel through Bob, who uses his PM to modulate the pulse according to his codeword $E(y)$. Both Alice and Bob use two external function generators (FG) to control the PMs. Finally, the two pulses arrive simultaneously at the BS, where they interfere and are detected by two detectors D_0 and D_1. The detection events are recorded by a time interval analyser (TIA).

Error-correcting code. In quantum fingerprinting, an ECC is used to amplify the Hamming distance between the inputs of Alice and Bob. Even if these inputs are originally very close to each other—for example, if they differ in a single position—after applying the ECC, the resulting codewords will have a much larger Hamming distance. In the worst-case scenario, this distance is given by the minimum distance of the code. Note that an important difference between a standard classical error-correction implementation and our current implementation is that in our implementation, Alice and Bob only need to perform encoding, but not the decoding of the ECCs. For this reason, we are concerned only with the computational complexity in encoding. This greatly simplifies our requirements.

The quantum fingerprinting protocol of ref. 29 used Justesen codes as an example to illustrate the properties of the protocol. However, these codes are not optimal for quantum fingerprinting. Here we construct a more efficient ECC that significantly relaxes the requirements on the experimental devices and leads to a faster implementation of the protocol. We make use of a subclass of random linear codes (RLCs)[36] whose generator matrices are Toeplitz matrices. Our ECC can asymptotically approach the Gilbert–Varshamov bound[37,38]. For various rates, it provides a minimum distance that is more than three times the value for Justesen codes. This is clearly illustrated in Fig. 3. The implementation details of our ECC are shown in Methods.

Experimental set-up. We demonstrate our proof-of-concept quantum fingerprinting protocol using a plug and play scheme[39], initially designed for QKD. The advantage of the plug and play system with respect to other viable systems is that it offers a particularly robust and stable implementation. This allows us to perform reliable experiments with highly attenuated coherent states for long time durations. We implement the protocol on top of two commercial systems, namely ID-500 and Clavis2, manufactured by ID Quantique.

In our set-up, which is shown in Fig. 4, the referee starts by sending two strong pulses at about 1,551 nm to Alice over a 5-km fibre. Once the two pulses reach Alice, she uses the reference pulse as a synchronization signal to activate her phase modulator, which she employs to set the phase of the signal pulse according to her codeword $E(x)$. Both pulses are reflected back by a Faraday mirror, which rotates the pulses' polarization by 90, and she attenuates them to the desired photon level using the variable optical attenuator (VOA). Once the pulses return back, due to the Faraday mirror, the pulses take opposite paths, such that the reference pulse now passes through Bob and its phase is modulated by Bob's phase modulator according to $E(y)$. Finally, the two pulses interfere at the referee's beam splitter and the detection events are registered using two high-quality single-photon detectors D_0 and D_1. It is important to note that the returning signal pulse modulated by Alice travels directly to the referee, while the returning reference pulse passing through Bob does not contain any information about Alice's codeword.

This guarantees that there is no communication between Alice and Bob.

Since the operating conditions of our protocol are significantly different from those of standard QKD, using a commercial QKD equipment for our implementation requires several important modifications to the system. First, two single-photon detectors—ID220 (manufactured by ID Quantique)—with low dark count rates were installed. Second, we performed several calibration and synchronization processes to enable the system work at an ultra-low mean photon number level, which is about four orders of magnitude lower than those typically used for QKD. Finally, we implemented two external function generators (Agilent 88250A) loaded with the codewords to control Alice's and Bob's phase modulator. The details of our modifications are presented in Methods. We observed high interference visibility of about (99 ± 0.5)% after careful calibration.

Experimental results. We perform the proof-of-concept quantum fingerprinting experiment over a standard telecom fibre of 5 km between Alice and the referee. The overall loss between the output of Alice's VOA and the input of the referee's detector D_1—which includes the losses of quantum channel, polarization beam splitter, beam splitter and the circulator—is about 3 dB (2.36 dB) for ID-500 (Clavis2). The channel between Bob and the referee is about a few metres, and its overall loss including Bob's channel, the beam splitter and the circulator is about 1.5 dB (1 dB). We summarize all system parameters in Table 1. On the basis of these parameters, for a given input size n, we use our model of the protocol to optimize the photon number μ to achieve a desired error probability ϵ.

Because there is loss in the channels and the detectors are not perfectly efficient, Alice and Bob must use higher mean photon numbers compared with the case with no channel loss and with perfect detectors. As implied by equation (2), this also leads to an increase in the transmitted information, which we take into account in our calculations of the transmitted information. In particular, if Alice and Bob experience different amounts of loss, they must choose a different mean photon number when preparing their signals, ensuring that the amplitude of their pulses is equal when they interfere in the referee's beam splitter.

In the experiment, the detection events registered on D_0 and D_1 in conjunction with the known experimental conditions in the system can be used to characterize the photon numbers sent out by Alice and Bob, the dark count probability and the visibility of the interferometer. From the characterization of these parameters, we find that there is a good agreement with our model of the system. The main source of uncertainty is due to an imperfect matching between the observed mean photon numbers and those pre-calibrated from the VOA. This uncertainty is determined by the fluctuations of several devices, such as laser power, VOA and detector efficiency. The detailed values of this uncertainty are shown in Methods.

The quantum fingerprinting protocol is tested over several values of the input size n. For each n, we record the detection counts on D_1 for two types of input data: equal inputs $E(x) = E(y)$, and the worst-case different inputs, that is, those for which the codewords $E(x) \neq E(y)$ have a distance equal to the minimum distance. For our experiment, we minimize the transmitted information by choosing an optimal value of $\delta = 0.22$ for the minimum distance. From the threshold value $D_{1,\mathrm{th}}$ that is pre-calculated from our model, the referee can distinguish between equal and different inputs. The upper bound Q on the quantum information Alice and Bob is calculated from their respective mean photon numbers μ_A and μ_B, as well as the codeword length m.

Table 1 | Parameters measured in the implementations.

η_{AR}	η_{BR}	η_{det}	p_{dark}	v
3 dB (2.36 dB)	1.5 dB (1 dB)	20.0%	$(3.5 \pm 0.2) \times 10^{-6}$	(99 ± 0.5)%

The overall loss between the output of Alice's VOA and the input to the referee's detectors is given by the parameter η_{AR}. Similarly, η_{BR} defines the overall loss between the output of Bob's phase modulator and the referee's detectors. Both η_{AR} and η_{BR} are carefully characterized in ID-500 (Clavis2). The other parameters are the detector's quantum efficiency η_{det}, dark count rate per pulse p_{dark} for each detector, and system visibility v, which are nearly the same for ID-500 and Clavis2.

In Fig. 5, we show the transmitted information as a function of the input size n for an error probability of $\epsilon = 5 \times 10^{-5}$. An error of 5×10^{-5} was chosen because it was the lowest error probability that was achieved by all runs of the experiment. The error probability was calculated from our theoretical model of the experiment. Within experimental uncertainty, the worst-case values of the mean photon number, visibility and dark count probability were used to reconstruct the probability distributions of clicks in detector D_1. These distributions, in turn, were used to calculate the error probability from equation (5). Since our theoretical model is only an approximation, the error probability should also be understood as approximate. The blue area in Fig. 5 indicates the region where the best-known classical protocol of ref. 30 transmits less information than our quantum protocol. For this target error probability, the classical protocol requires the transmission of $16\sqrt{n}$ bits. The red points show our experimental results, where the data point for the largest n is obtained from ID-500 and the other three data points are obtained from Clavis2. Note that Clavis2 and ID-500 have almost the same optics and functionality. We use the same measurement and processing method for the data obtained from these two systems, and show the experimental results together in one figure instead of two. The error bars come from the uncertainty in the estimation of the mean photon number μ. For large n, our experimental results are strictly better than those of the classical protocol for a wide range of practical values of the input size.

To obtain further insight into our results, we define the quantum advantage γ as the ratio between the transmitted

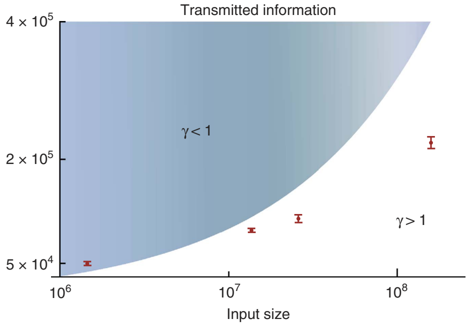


Figure 5 | Log-linear plot of the total transmitted information by Alice and Bob in our protocol. The blue area indicates the region where the classical protocol transmits less information than our protocol, while the red points show our experimental results. The quantum advantage γ corresponds to the ratio between the transmitted classical information and the upper bound on the transmitted quantum information. The error bars correspond to one standard deviation. For large n, our results are strictly better than the best-known classical protocol for a range of practical values of the input size.

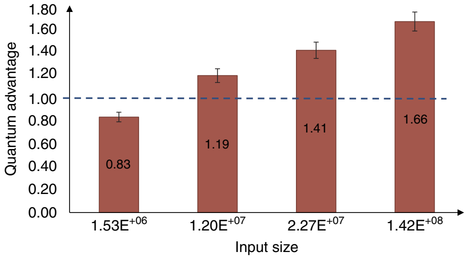


Figure 6 | Quantum advantage γ between the transmitted classical information and the upper bound on the transmitted quantum information. The uncertainty refers to one standard deviation, which mainly comes from the error in estimating the mean photon number per pulse. For the three large input sizes, the ratio is well above 1. The quantum advantage was as large as $\gamma = 1.66$, which implies that the transmitted information in the classical protocol was 66% larger than in the quantum case.

classical information C of the best-known classical protocol[30] and the upper bound Q on the transmitted quantum information:

$$\gamma = \frac{C}{Q}. \tag{6}$$

A value $\gamma > 1$ for a given error probability ϵ implies that less information is transmitted in the quantum case than in the classical one. This allows us to use the quantum advantage as a figure of merit to assess the performance of our quantum fingerprinting implementation. In Fig. 6, we show the experimental results for γ as a function of different input sizes. For the three largest input sizes, the ratio is well above 1, and the classical protocol transmitted as much as 66% more information than the quantum protocol. For the smallest input size, no quantum improvement was obtained.

Discussion

On the basis of the protocol of ref. 29, we have experimentally demonstrated a proof-of-concept quantum fingerprinting system that is capable of transmitting less information than the best-known classical protocol for this problem. Our experimental test of this system indicates that its operation is consistent with our model of the devices and hence also with achieving the desired error probability. Moreover, we have operated our system in a parameter regime in which the information transmitted in the protocol is up to 66% lower than the best-known classical protocol. This constitutes the first time that a quantum fingerprinting protocol has been carried out that is capable of achieving this reduction in the transmitted information.

It is an appealing and useful property of this quantum fingerprinting protocol that we can achieve a quantum advantage without the need for entanglement, single-photon sources or squeezing. Where does the improvement come from? As discussed extensively in ref. 35, the states of equation (1) that are used in our protocol are a coherent-state version of an encoding of m-dimensional quantum states into states of a single photon across m modes. Through this encoding, exponentially more 'sufficiently distinguishable' quantum states can be fitted into an $O(\log_2 m)$-qubit Hilbert space as opposed to orthogonal classical states. In our protocol, instead of $O(\log_2 m)$ qubits, the same amount of quantum information can be encoded into a sequence of coherent states.

One can understand the quantum advantage as arising from the non-orthogonality of weak coherent states and the quantum-mechanical properties of single-photon detectors. In the protocol, the weak coherent states have a very low mean photon number. This means that the two possible states that are sent in each mode, $|+\frac{a}{\sqrt{m}}\rangle$ and $|-\frac{a}{\sqrt{m}}\rangle$, are highly non-orthogonal and fundamentally difficult to distinguish. Therefore, very little information can be learnt by looking at each pulse. This is essentially the reason why the transmitted information is very low—exponentially less than in the classical case. On the other hand, after the coherent states interfere in the beam splitter, a click in the single-photon detector unambiguously provides valuable information to the referee: she now knows whether the phases of the coherent states are equal or not. This unambiguous information is only possible because the detectors respond quantum mechanically to the incoming light field.

The main goal of our experiment is to demonstrate a reduction in the transmitted information compared with the best-known classical protocol. However, from a practical perspective, one might be interested in additional quantities, such as energy expenditures or running time, beyond the abstract transmitted information. In our protocol, the running time is quadratically larger than in the classical case, provided we ignore the running time required for the ECC, which is the dominant one. Therefore, if running time during communication is a priority, our protocol has a disadvantage: the quantum protocol may become infeasible for a very large input size of time bins, limited by the repetition rate of the laser source. Nonetheless, if minimizing energy expenditures is a priority, our protocol offers a significant advantage. In particular, the number of photons used is more than quadratically smaller than in a classical protocol using photonic bits, where $O(\sqrt{n})$ photons are needed compared with $O(1)$ photons in the quantum case.

Finally, in this work, we have tested our model of the system and used that test to make an indirect assessment of the error probability based on our theoretical model. Future implementations should improve on this by treating the system as a black box, using the data directly to make statistical inferences about the error probability, without relying on an approximate model of the system. Overall, it is remarkable that quantum fingerprinting can be realized while revealing only a very small amount of information to the referee—a feature of the protocol that may have important applications to fields such as cryptography[40] and information complexity[41], where this extremely small leakage of information plays a fundamental role. Our results constitute a significant first step in the development of experimental quantum communication complexity, which may also be extended to other protocols with a proven exponential advantage over the classical case[10,11,35,42].

Methods

Error-correcting code. In quantum fingerprinting protocol, an ECC with a high rate and a large minimum distance is desired, since a higher rate leads to lower transmitted information and larger tolerance for dark counts, while a larger minimum distance leads to smaller error probability for fixed mean photon number. Fundamentally, there is an inherent trade-off between the rate and distance of ECCs. In particular, the Gilbert–Varshamov bound states that there exists some binary linear code whose rate R and minimum distance δ satisfy the relation

$$R \geq 1 - H_2(\delta), \tag{7}$$

where $H_2(\cdot)$ is the binary entropy function. Using a binary linear code that approaches this bound would constitute a significant improvement over the codes used in previous protocols.

It is well known in coding theory that RLCs can asymptotically approach the Gilbert–Varshamov bound with encoding complexity $O(n^2)$ (ref. 43). However, in quantum fingerprinting, the input size n is typically very large (for example, $n = 10^8$), thus making the encoding time prohibitively high. To reduce this encoding complexity, we make use of a subclass of RLCs whose generator matrices are Toeplitz matrices. A Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant. An $n \times m$ Toeplitz matrix is completely determined by the $n + m - 1$ elements on its first row and column. This structure implies that only $O(n \log n)$ time for encoding is required for this subclass of RLCs[36]. In addition, these codes also asymptotically approach the Gilbert–Varshamov bound. By using this family of codes, we are able to reduce the encoding times by several orders of magnitude, making them suitable for practical applications.

The exponential separation between quantum and classical communication complexity for the equality function only holds if Alice and Bob do not have access to shared randomness that is generated in each run of the protocol[30]. However, even though the generator matrices of our RLCs are randomly constructed, once they have been created they remain fixed for all future instances of the protocol. This ensures that no new randomness is generated in each run of the protocol, as required to satisfy the conditions of the exponential separation. In particular, Alice and Bob can store the generator matrices in memory and use them to encode their inputs in exactly the same way as if they had been generated deterministically.

For our experiment, an encoder programme written in C++ was built and tested, demonstrating the feasibility of this subclass of RLCs. The free Fast-Fourier Transform library FFTW was used to accelerate multiplications with Toeplitz matrices[44] and the random numbers to construct the matrices were generated from a quantum random number generator[45]. The results from an optimized encoder are shown in Table 2. As we can see, our encoder is highly practical, can be run on any common lab personal computer (PC) and finishes the encoding in an acceptable time frame for input sizes as large as $n = 3 \times 10^8$. Faster encoding times could be obtained by using dedicated hardware.

Experimental details. We performed several modifications on top of the plug and play system, to implement the quantum fingerprinting protocol. First, two single-photon detectors with low dark count rates were installed. Indeed, as can be deduced from equations (3) and (4), lower dark count rates permit the operation of the system at lower mean photon numbers, which lead to a reduction in the transmitted information. Fortunately, our error-correction codes improve the tolerance of the protocol to dark counts, which permits us to use commercial detectors. We employ two commercial free-running InGaAs avalanche photodiodes—ID220. The dark count rate per 1 ns detection gate is about $(3.5 \pm 0.2) \times 10^{-6}$ and the corresponding quantum efficiency is about 20%. The detections are recorded by a high-precision time interval analyser (PicoQuant HydraHarp 400). The system was run at a repetition rate of 5 MHz with the detector dead time set at 10 μs. This means that after a click occurred, the following 50 pulses are blocked before the detector is active again. This is not a problem in our experiment because the mean photon number in each pulse is extremely low, therefore, the expected number of undetected photons as a result of this effect is negligible compared with other sources of error.

In addition, new functionalities and control signals were added to the system. On one hand, we use the VOA inside Alice to reduce the mean photon number per pulse down to suitable numbers. These values—in the order of 10^{-5} per pulse—were in fact four orders of magnitude lower than those typically used for QKD. Hence, several calibration processes of the system are required, which imposes particular care in the synchronization of the phase modulation and attenuation signals. On the other hand, commercial QKD systems like Clavis2 have an internal random number generator to set the phase modulations, which does not allow us to modulate the phases according to the pre-generated codewords. We solve this difficulty by using two external function generators (Agilent 88250A) loaded with the codewords to control Alice's and Bob's phase modulator. This requires precise synchronization and calibration procedures to guarantee correct phase modulations.

Table 2 | The performance of the encoder for different input sizes, using a computer with a quad-core i7-4770 @3.4 GHz CPU and 16 GB RAM.

n (bit)	*m* (bit)	Time (s)	Memory (Mbit)
10^6	5×10^6	6	52
10^7	5×10^7	106	733
3×10^7	1.5×10^8	181	1,654
3×10^8	1.5×10^9	4,831	10,000

Running times are acceptable for experimental applications for input sizes as large as $n = 3 \times 10^8$.

In the proof-of-concept implementation on ID-500, the random numbers controlling the phase modulations are accessible to users. We use our codewords to replace those random numbers directly. However, after testing for an input data size of $n = 1.42 \times 10^8$ on ID-500, an unexpected hardware problem made ID-500 unavailable for further experiments. To further test the feasibility of our protocol for different input sizes, we switched to Clavis2 for measurements. In the implementation on Clavis2, since each function generator has a small memory, for simplicity we load a frame of about 430 random numbers to each function generator and reuse these random numbers. This allows us to create binary sequences with the desired distance δ that can be used to test the performance of the system. All the above modifications led to the development of a practical system that is capable of performing quantum fingerprinting.

Practical considerations. In communication complexity, it is assumed that the parties have unlimited computational power. However, from a practical perspective, it may not always be possible to ignore these computational requirements. In fact, even though the running time during communication of our experiment scales linearly with the input size, the total running time of the protocol is dominated by the time required to run the ECC—which is a crucial component of the protocol. For instance, at a repetition rate of 5 MHz, it takes 5 min to run the communication for an output size of $m = 1.5 \times 10^9$. On the other hand, even with the use of RLCs with quasi-linear encoding complexity, more than 1 h is needed to run the encoding algorithm, as seen in Table 2. Therefore, the practical advantages of quantum fingerprinting, in terms of reductions in resource expenditures, will likely be found in a reduction of the number of photons used. This is a major property that our protocol possesses. Indeed, for the largest input size that we tested, $n = 1.42 \times 10^8$, a total mean photon number of only $\mu \approx 7 \times 10^3$ was used. Moreover, because the protocol does not require time resolution in the detectors—the referee only cares about the number of clicks, not when they happen—in principle it is possible to run this protocol at very fast rates, limited only by the source repetition rate.

In our quantum fingerprinting protocol, the maximum reduction in the transmitted information depends crucially on the dark count probability and the overall loss in the system. Thus, our results can be directly improved by using detectors with higher efficiency and lower dark counts. This can lead to a quantum fingerprinting protocol that, with the use of available technology[46], transmits several orders of magnitudes less information than the best-known classical protocol for large input sizes. Even though there is no proof that the best-known classical protocol is optimal, a lower bound for the classical transmitted information was proven in ref. 30. This lower bound states that, for any classical protocol with error probability smaller than 0.01, Alice and Bob must send at least $\frac{\sqrt{n}}{20}$ bits of information. This is roughly two orders of magnitude smaller than the transmitted information of the best-known classical protocol. By using state-of-the-art detectors, it should be possible to demonstrate a quantum fingerprinting protocol capable of beating this classical lower bound. Achieving this would constitute a significant milestone for experimental quantum communication complexity.

Finally, in our implementation, a reference pulse is transmitted between the two participants for a share of synchronization and phase reference. In practice,

Table 3 | Detailed experimental results.

System	Clavis2	Clavis2	Clavis2	ID-500
n	1.53×10^6	1.20×10^7	2.27×10^7	1.42×10^8
μ_A	1,914 ± 68	3,295 ± 118	3,670 ± 131	7,120 ± 254
μ_B	1,398 ± 50	2,407 ± 86	2,681 ± 96	5,014 ± 179
$D_{1,E}$	22	277	830	1,939
$D_{1,D}$	131	318	954	2,224
$D_{1,th}$	49	302	902	2110
Q	47,689 ± 1,703	93,152 ± 3,326	108,129 ± 3,860	229,713 ± 8,201
γ	0.83 ± 0.02	1.19 ± 0.05	1.41 ± 0.05	1.66 ± 0.06
ϵ	$(1.6 \pm 0.9) \times 10^{-9}$	$(2.3 \pm 1.4) \times 10^{-7}$	$(6.6 \pm 3.7) \times 10^{-6}$	$(2.9 \pm 1.3) \times 10^{-5}$

The parameter μ_A is the mean photon number for Alice and μ_B is the mean photon number for Bob. For the clicks in detector D_1 we report the observed averages for the case of equal inputs $D_{1,E}$, different inputs $D_{1,D}$ and the threshold value used by the referee $D_{1,th}$. As before, Q is the upper bound on the quantum transmitted information, γ is the quantum advantage and ϵ the error probability of the protocol.

one can overcome this by using a system where each of Alice and Bob holds a frequency-locked laser source separately. A common phase reference can be established before the start of the protocol or the referee can employ phase-locking techniques to interfere the two pulses from Alice and Bob. Indeed, a potential method for such an implementation is to use the techniques that have been recently developed in the field of QKD[47–49]. This configuration, unlike the plug and play scheme, can also permit Bob to be situated at a large distance from the referee.

Error probability analysis. We prove that the Toeplitz matrix based RLCs also asymptotically approach the Gilbert–Varshamov bound. Let G be a random $n \times m$ Toeplitz matrix over $\mathbb{F}_2$. There are two failure events associated with G: the minimum distance δ being not as large as promised (which results in less-than-expected worst-case performance) and the matrix G being not full rank (which can cause two different inputs to be mapped to the same output, leading to a minimum distance of $\delta = 0$). We will show that, for any fixed rate $R < 1 - H_2(\delta)$, the probabilities of both failure events decreases exponentially with the output size m and can thus be neglected for sufficiently large m.

Theorem 1 ref. 50. Let $G \in \mathbb{F}_2^{n\times m}$ be a Toeplitz matrix chosen uniformly at random. Let $\delta_{\min}(G)$ be the minimum distance of the linear code with G as generator matrix. Then, for any $\delta \in (0,1/2)$,

$$\Pr(\delta_{\min}(G) \leq \delta) \leq 2^{-m(1-H_2(\delta)-R)}.$$

In particular, if $R = 1 - H_2(\delta) - \epsilon$, for some $\epsilon > 0$, then

$$\Pr(\delta_{\min}(G) \leq \delta) \leq 2^{-\epsilon m}.$$

The above theorem guarantees that, if we sacrifice an arbitrarily small quantity ϵ of the rate with respect to the Gilbert–Varshamov bound (that is, we set $R = 1 - H_2(\delta) - \epsilon$), the probability of obtaining an incorrect minimum distance decreases exponentially with the output size. For example, for a value of $m = 10^7$ and $\epsilon = 10^{-3}$, this probability is $< 10^{-104}$.

Theorem 2. Let $G \in \mathbb{F}_2^{n\times m}$ be a Toeplitz matrix chosen uniformly at random. Then,

$$\Pr(G \text{ is not full rank}) = 2^{-1}2^{-m(1-R)}.$$

Theorem 2 is an immediate consequence of Theorem 1 in ref. 51. Once again, this probability decreases exponentially with the output size m.

Detailed experimental results. In Table 3, we report the complete results of our experiment. The dominating source of uncertainty is the uncertainty in the total mean photon number of the signals. This uncertainty is due to the summation of the fluctuations of several devices, such as laser power, VOA and varying loss in the channel. For each input size n, we perform a calibration process to determine μ. In this process, with a proper value of VOA selected from our numerical optimization, the referee sends out around $10^7 \sim 10^8$ pulses to Alice and Bob. From the total detection counts on D_0 and D_1 and the pre-calibrated losses (Table 1), we estimate the μ. We repeat this calibration process a few rounds and obtain the mean value and the s.d. for μ. These results are shown in the second column of Table 3. For all tested cases, the uncertainty in mean photon number was below 4%.

Note that the mean photon numbers for Alice and Bob are unequal. This is because in the implementation, to guarantee a good interference visibility, we carefully control the attenuations such that the light from Alice and the light from Bob have the same amplitude when they interfere at the referee. Since the attenuations from Alice to the referee and from Bob to the referee are unequal (Table 1), we choose unequal mean photon numbers for Alice and Bob.

From our model of the protocol, we use the uncertainty in the mean photon number to directly calculate an uncertainty for the quantum transmitted information as well as for the error probability of the protocol. As it can be seen from Table 3, all error probabilities are compatible with the system operating below the target value of $\epsilon = 5 \times 10^{-5}$. In addition, we have included the average values observed for the number of clicks in detector D_1 for equal and different inputs, as well as the threshold values used by the referee.

Finally, we estimate the effect of detector dead times in our experiment as follows. For each input size, we can calculate the probability p that an individual pulse leads to a click in detector D_1. In our set-up, after a click occurs, the following 50 pulses are blocked by the detector and cannot be registered. The probability p' that a click occurs for these 50 pulses is given by $p' = 1 - (1-p)^{50} \approx 50p$. This number is very small whenever p is small, as is the case in our experiment. For instance, for an input size of $n = 1.42 \times 10^8$, the expected number of blocked clicks is $\sim 0.1\%$ of the total expected clicks. Therefore, this effect is negligible compared with fluctuations in the mean photon number, which is of the order of 4%.

References

1. Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conf. Comput. Syst. Signal Process.* 175–179 (IEEE (1984).
2. Ekert, A. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.* **67,** 661–663 (1991).
3. Gottesman, D. & Chuang, I. Quantum digital signatures. Preprint at http://arxiv.org/abs/quant-ph/0105032 (2001).
4. Giovannetti, V., Lloyd, S. & Maccone, L. Advances in quantum metrology. *Nat. Photon.* **5,** 222–229 (2011).
5. Yao, A. C.-C. in *Proceedings of the 11th Annual ACM Symposium on Theory of Computing* 209–213 (New York, NY, USA, 1979).
6. Brassard, G. Quantum communication complexity. *Found. Phys.* **33,** 1593–1616 (2003).
7. Buhrman, H., Cleve, R., Massar, S. & de Wolf, R. Nonlocal-ity and communication complexity. *Rev. Mod. Phys.* **82,** 665–698 (2010).
8. Buhrman, H., van Dam, W., H0yer, P. & Tapp, A. Multiparty quantum communication complexity. *Phys. Rev. A* **60,** 2737 (1999).
9. Buhrman, H., Cleve, R., Watrous, J. & de Wolf, R. Quantum fingerprinting. *Phys. Rev. Lett.* **87,** 167902 (2001).
10. Raz, R. in *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing* 358–367 (Atlanta, GA, USA, 1999).
11. Bar-Yossef, Z., Jayram, T. S. & Kerenidis, I. in *Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing* 128–137 (Chicago, IL, USA, 2004).
12. Gavinsky, D., Kempe, J., Kerenidis, I., Raz, R. & De Wolf, R. in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* 516–525 (San Diego, CA, USA, 2007).
13. Regev, O. & Klartag, B. in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* 31–40 (San Jose, CA, USA, 2011).
14. Becerra, F. *et al.* Experimental demonstration of a receiver beating the standard quantum limit for multiple nonorthogonal state discrimination. *Nat. Photon.* **7,** 147–152 (2013).
15. Xiang, G.-Y., Higgins, B. L., Berry, D., Wiseman, H. M. & Pryde, G. Entanglement-enhanced measurement of a completely unknown optical phase. *Nat. Photon.* **5,** 43–47 (2011).
16. Ng, N. H. Y., Joshi, S. K., Ming, C. C., Kurtsiefer, C. & Wehner, S. Experimental implementation of bit commitment in the noisy-storage model. *Nat. Commun.* **3,** 1326 (2012).
17. Clarke, P. J. *et al.* Experimental demonstration of quantum digital signatures using phase-encoded coherent states of light. *Nat. Commun.* **3,** 1174 (2012).
18. Lunghi, T. *et al.* Experimental bit commitment based on quantum communication and special relativity. *Phys. Rev. Lett.* **111,** 180504 (2013).
19. Liu, Y. *et al.* Experimental unconditionally secure bit commitment. *Phys. Rev. Lett.* **112,** 010504 (2014).
20. Collins, R. J. *et al.* Realization of quantum digital signatures without the requirement of quantum memory. *Phys. Rev. Lett.* **113,** 040502 (2014).
21. Berlin, G. *et al.* Experimental loss-tolerant quantum coin flipping. *Nat. Commun.* **2,** 561 (2011).
22. Pappa, A. *et al.* Experimental plug and play quantum coin flipping. *Nat. Commun.* **5,** 3717 (2014).
23. Steane, A. M. & van Dam, W. Physicists triumph at guess my number. *Phys. Today* **53,** 35–39 (2000).
24. Kushilevitz, E. & Nisan, N. *Communication Complexity* (Cambridge Univ. Press, 2006).
25. Horn, R. T., Babichev, S. A., Marzlin, K.-P., Lvovsky, A. I. & Sanders, B. C. Single-qubit optical quantum fingerprinting. *Phys. Rev. Lett.* **95,** 150502 (2005).
26. Du, J. *et al.* Experimental quantum multimeter and one-qubit fingerprinting. *Phys. Rev. A* **74,** 042319 (2006).
27. Trojek, P. *et al.* Experimental quantum communication complexity. *Phys. Rev. A* **72,** 050305 (2005).
28. Massar, S. Quantum fingerprinting with a single particle. *Phys. Rev. A.* **71,** 012310 (2005).
29. Arrazola, J. M. & Lutkenhaus, N. Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A* **89,** 062305 (2014).
30. Babai, L. & Kimmel, P. G. in *Proc. 12th Annu. IEEE Conf. Comput. Complexity* 239–246 (IEEE, IEE (1997).
31. Newman, I. & Szegedy, M. in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* 561–570 (Philadelphia, PA, USA, 1996).
32. IDQuantique, Geneva http://www.idquantique.com.
33. Hong, C., Ou, Z. & Mandel, L. Measurement of subpicosec-ond time intervals between two photons by interference. *Phys. Rev. Lett.* **59,** 2044 (1987).
34. Andersson, E., Curty, M. & Jex, I. Experimentally realizable quantum comparison of coherent states and its applications. *Phys. Rev. A.* **74,** 022304 (2006).
35. Arrazola, J. M. & Lutkenhaus, N. Quantum communication with coherent states and linear optics. *Phys. Rev. A* **90,** 042335 (2014).
36. Emiris, I. Z. & Pan, V. Y. in *Algorithms and Theory of Computation Handbook* 18–18 (Chapman & Hall/CRC, 2010).
37. Gilbert, E. N. A comparison of signalling alphabets. *Bell Syst. Tech. J.* **31,** 504–522 (1952).
38. Varshamov, R. Estimate of the number of signals in error correcting codes. *Dokl. Akad. Nauk SSSR* **117,** 739–741 (1957).
39. Stucki, D., Gisin, N., Guinnard, O., Ribordy, G. & Zbinden, H. Quantum key distribution over 67 km with a plug & play system. *New J. Phys.* **4,** 41 (2002).
40. Gavinsky, D. & Ito, T. Quantum fingerprints that keep secrets. *Quantum Inf. Comput.* **13,** 583–606 (2013).

41. Chakrabarti, A., Shi, Y., Wirth, A. & Yao, A. in *Proc. 42nd IEEE Symp. Found. Comput. Sci.* 270–278 (IEEE (2001).
42. Arrazola, J. M. & Lutkenhaus, N. in *9th Conference on the Theory of Quantum Computation, Communication and Cryptography* 36–47 (Singapore, 2014).
43. Barg, A. & Forney, G. Random codes: Minimum distances and error exponents. *IEEE Trans. Inf. Theory* **48,** 2568–2573 (2002).
44. Frigo, M. & Johnson., S. G. The design and implementation of fftw3. *Proc. IEEE* **93,** 216–231 (2005).
45. Xu, F. *et al.* Ultrafast quantum random number generation based on quantum phase fluctuations. *Opt. Express.* **20,** 12366–12377 (2012).
46. Marsili, F. *et al.* Detecting single infrared photons with 93% system efficiency. *Nat. Photon.* **7,** 210–214 (2013).
47. Rubenok, A., Slater, J. A., Chan, P., Lucio-Martinez, I. & Tittel, W. Real-world two-photon interference and proof-of-principle quantum key distribution immune to detector attacks. *Phys. Rev. Lett.* **111,** 130501 (2013).
48. Liu, Y. *et al.* Experimental measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **111,** 130502 (2013).
49. Tang, Z. *et al.* Experimental demonstration of polarization encoding measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **112,** 190503 (2014).
50. Guruswami, V., Rudra, A. & Sudan, M. *Essential Coding Theory* (Univ. of Buffalo (2014).
51. Daykin, D. Distribution of bordered persymmetric matrices in a finite field. *J. Reine Angew. Math.(Crelles J.)* **203,** 47–54 (1960).

Acknowledgements

This work was supported in part by ID Quantique, CFI, OIT, NSERC RTI, the NSERC SPG FREQUENCY, the NSERC Discovery Program, the CRC Program, Connaught Innovation fund and Industry Canada. We thank A. Ignjatovic, Pedro Palacios-Avila, Z. Tang and Francisco de Zela for valuable discussions. Particularly, F. Xu and K. Wei thank V. Makarov for his hospitality during their visit to IQC. We acknowledge the support from the Mike and Ophelia Lazaridis Fellowship, the Office of Naval Research (ONR), the Air Force Office of Scientific Research (AFOSR), the National Natural Science Foundation of China (Grant No. 61178010), China Scholarship Council (No. 201406470051), the University of Hong Kong Physics Department Summer Overseas Research Program and the Science Faculty Overseas Research Fellowship, CONCYTEC-Peru, USEQIP, the IQC Summer URA, the NSERC Postdoctoral Fellowships Program and CryptoWorks21.

Author contributions

F.X., K.W. and S.S. conducted the experiment; J.A., P.P.-A. and N.L. performed the theoretical analysis; F.X., K.W. and H.-K.L. analysed the experimental data; W.W. and C.F. developed the error-correcting code; F.X. and J.A. conceived the idea; N.L. and H.-K.L. supervised the project; F.X., J.A., H.-K.L. and N.L. wrote the manuscript, with input from all the authors.

Additional information

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Xu, F. *et al.* Experimental quantum fingerprinting with weak coherent pulses. *Nat. Commun.* 6:8735 doi: 10.1038/ncomms9735 (2015).