Laser Damage Attacks Against Optical Attenuators in QKD Systems

Ruoping Li

January 8, 2018

The following work is submitted as partial fulfillment of the requirements of the Physics 437A Research Project course at the University of Waterloo.

Abstract

Despite the claim of unconditional, informational-theoretic security of quantum key distribution (QKD) systems, there exists multiple attack vectors taking advantage of equipment imperfection which can compromise the security of a QKD network. As optical components can suffer from damage at high optical powers, our experiments show that laser damage near 37 dBm (5 W) from an erbium-doped fiber amplifier (EDFA) can permanently affect typical variable optical attenuators used in existing QKD systems in such a way that the underlying security of the quantum key exchange is compromised. Out of 6 samples, optical damage produces a permanent drop in attenuation of up to 9 dB, which multiplies the mean photon number μ inside the quantum channel by a factor of up to 2.8. This attack vector effectively breaks the security assumption of the quantum states being exchanged via single photons subject to the no-cloning theorem, and renders the exchanged secret key susceptible to the photon-number splitting (PNS) attack.

Contents

List of Figures

List of Tables

1 Introduction

Quantum key distribution (QKD) is often claimed to be the successor to traditional key distribution algorithms in classical cryptography, promising information-theoretic security with perfect forward secrecy [1, 2]. However, the unconditional security of QKD can be compromised from implementation problems and equipment imperfections [3–5]. In the Bennett-Brassard (BB84) protocol, the polarization states of single photons is used to carry the secret key, and any eavesdropper performing a measurement will induce bit errors which can then reveal the existence of said eavesdropper [2, 6]. In many commercial QKD systems, a variable optical attenuator (VOA) is used to attenuate the optical link to single photon levels [7]. If there is a method in which the attenuator can be modified from the fiber network side to allow a higher than expected mean photon number, it is then possible for an eavesdropper to extract information about the secret key without raising any detectable alert through a photon number splitting (PNS) attack [8, 9]. We hypothesize that this is possible using a high-powered fiber-coupled laser to optically damage the VOA, which, depending on the exact operating principle of the VOA, can result in a permanent drop in attenuation.

Note to referee: Please keep in mind that certain experiment sections contain potentially sensitive material under non-disclosure agreement. Please do not redistribute this report without prior consent.

2 Historical

Quantum key distribution (QKD) is a subset of quantum cryptography, which is itself a sub-field of the larger field of cryptography. QKD is a procedure for the distribution of a secret key (a shared, random piece of information with high entropy) which is inherent distinct from analogous algorithms in classical cryptography such as the Diffie-Hellman key exchange. The difference lies in that the security proof of quantum cryptography does not rely on conjectures in mathematics and computational complexity, but rather the laws of quantum mechanics [6, 8]. However, despite this claim, the security premise of QKD is only valid in an ideal world unrestricted by physical hardware or engineering limitations. In a real situation, there exist subtle implementation flaws and equipment imperfections which can be manipulated by an attacker to extract the secret key, either partially or in its entirety [8]. This process can effectively be called "quantum hacking", drawing the parallel with traditional cybersecurity.

2.1 Forward secrecy

Forward secrecy means that the future compromise of a given cryptosystem does not imply leakage of previously exchanged information. Some schemes in classical cryptography such as SSL and TLS do in fact offer forward secrecy [10]. In a real-life situation, nothing prevents a powerful nation state to gather encrypted data and simply wait until sufficient computational power is available to break the underlying cryptography, possibly through quantum computing or a solution to the P=NP computational complexity problem [11].

2.2 Unconditional security

Unconditional security implies that no arbitrary condition is being assigned to the potential cryptographic attacker, and that the cryptographic security is rooted in information theory instead of computational security [12]. A commonly used condition today is that the attacker is assumed to be economically restricted, setting an upper limit on the attacker's computational power. As well, when talking about key sizes, the statement is often made that "it would take over a million years (or longer than the age of the universe) for an attacker to break 256-bit AES". Such a statement still imposes a condition to the attacker, and implies that the attacker's knowledge of mathematics, algorithms and implementation of computing devices is similar to ours at the current time. While valid in the real world, this security model can sometimes be difficult to accept for a rigorous mathematician or cryptographer.

2.3 One-time pad

If used correctly, the one-time pad (OTP) cryptosystem offers perfect forward secrecy and unconditional security. Historically, it has been briefly used by British and Soviet spies near the first half of the 20th century [13]. The OTP system requires an encryption key being the same size as the plaintext which the user wishes to encrypt, and an encoding scheme (usually XOR). Each bit in the plaintext P would then be combined with a bit in the key K, giving the ciphertext C.

$$C = XOR(P, K) = P \oplus K \tag{1}$$

An intuitive explanation for the unconditional security of OTP is that the plaintext and ciphertext space are equivalent in size, and the transformation from plaintext to ciphertext is one-to-one (as in the case for the XOR OTP). For example, for a binary plaintext "101010" combined with a key "100101", the encoded ciphertext would be 001111. An attacker trying to decrypt the key by trial and error, knowing that the key is the same size as the plaintext, would not be able to distinguish between different key-plaintext pairs. For example, plaintext "001000" and key "000111" would be exactly as valid as plaintext "111000" and key "110111", all producing the same ciphertext "001111".

There are however a few subtleties in the OTP system which are required for perfect unconditional security, such as the key having a high entropy, or randomness [14]. As well, historically, the requirement that the key be exactly the same length as the plaintext, and the non-reusability of the OTP key has been a challenge to its implementation. QKD however, offers a potential solution to the logistic complexity behind the key-sharing process.

2.4 BB84 protocol

The BB84 protocol, eponymously proposed by Charles Bennett and Gilles Brassard in 1984, is a widely known protocol for quantum key distribution. Although various other QKD protocols exist (E91, B92, MDI-QKD, CVQKD, etc.), for the purposes of our experiment, considering the BB84 protocol is sufficient.

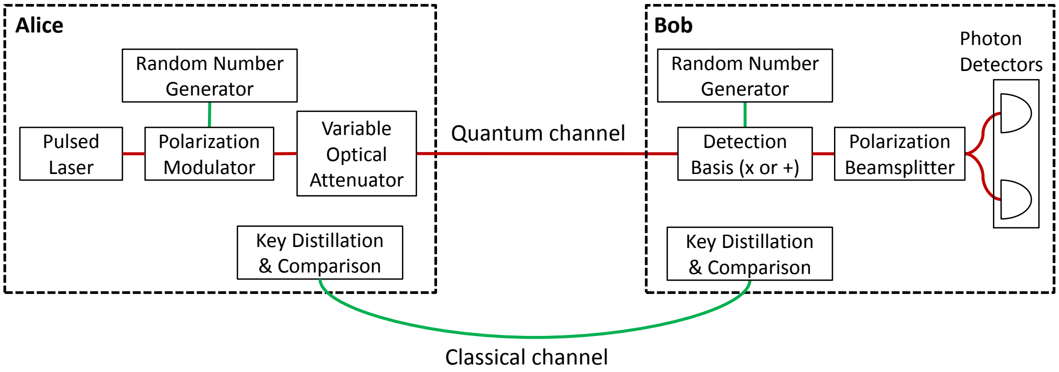


Figure 1: Simplified diagram of a QKD system employing the BB84 protocol. The red lines denote the quantum link through fiber optics, and the green lines denote classical channels (not necessarily secure).

Following the tradition used in various textbook examples, let Alice and Bob be the two parties involved in the BB84 key exchange, and let the potential eavesdropper be named Eve. An optical link is first established between Alice and Bob, which can be either via free-space optical communication or via existing fiber optic networks. The setup is shown in Fig. 1, and the procedure can be described as follows:

1. Alice sets her own variable optical attenuator such that the outgoing quantum link to Bob is attenuated to an expected mean photon number $\mu < 1$ (to single-photon levels) [6].
2. Alice generates a first random number R_1, which will be used for selecting the polarization direction of her outgoing signal to Bob.
3. Alice generates a second random number R_2, which will be used for modulating her outgoing signal between the $+45°/-45°$ and $0°/90°$ polarization basis (or equally, linear and left/right circular basis).
4. Alice selects the polarization direction of her outgoing signal with every bit of R_1, each time using the basis given by the corresponding bit in R_2 (0 being the $+45°/-45°$ basis and 1 being the $0°/90°$ basis), forming a set of four possible polarizations.
5. Bob, having previously generated his own random number R_3, receives the stream of single photons from Alice and measures them according to

the basis given by each bit in R_3, using a corresponding basis (+45°/−45° or 0°/90°). Let the measurement obtained by Bob be M.

6. Bob sends his random basis choice given by R_3 to Alice through the classical channel, which reveals which measurement basis he used.

7. Alice and Bob selects the single photons which were generated and measured using the same basis (expected fraction of 1/2), with the result shared through the classical channel. This process is equivalent to performing the bitwise operation $XNOR(R_2, R_3)$.

8. Bob takes the result from given by Alice in step 7, corresponding to selecting only the bits from M which were both generated and measured with the same basis, and calls this result the sifted key K_s. This sifted key should be identical for Alice and Bob if no other measurement (via a potential eavesdropper Eve, or any fiber line loss) has taken place.

9. Based on an estimate of the information leaked to Eve (due to multiple photons from Poisson statistics, or potential device imperfections), Alice and Bob both perform a *privacy amplification* scheme to reduce the key size to obtain the final secret key K_f which is shared by both parties.

10. Bob and Alice then encrypt a message using the one-time pad scheme with a portion of the final secret key K_f and verify whether it is the same. A bijective search can be used to narrow down and remove the portion of K_f which contain errors.

Point 1 assures that for the majority of bits being transferred, a given quantum state is sent using a single photon and cannot be separated out by the eavesdropper Eve and measured independently. If this optical attenuation component can be altered and its attenuation decreased, either permanently or temporarily, it is then theoretically possible for an eavesdropper to separate the extra photon and perform a measurement on it, which does not change the communication between Alice and Bob.

The privacy amplification process is used to eliminate the information which is physically possible to eavesdrop from a secret key [15]. It uses an arbitrary compression algorithm (see hash function) to reduce the keyspace size of the transmitted key. For example, a given key of length n bits is shared between Alice and Bob, in which k insecure bits exist. These insecure bits are obtained from the error fraction of the shared bits, which may be due to equipment imperfection, channel losses or a potential eavesdropper. The minimum privacy amplification process needs to compress the keyspace size to $n - k$ bits. It can be shown that the information gained by an eavesdropper Eve is less than $1/ln(2)$ bits of the privacy-amplified keyspace $n-k$ [16–18].

2.5 Quantum hacking

There exists several vulnerabilities against quantum key distribution systems which have been shown by previous publications. For the BB84 protocol, the photon number splitting (PNS) attack relies on the eavesdropper Eve splitting away polarization states which are transmitted with multiple photons (which necessarily exist for a lossy transmission medium with a laser source attenuated to a mean photon number $\mu < 1$) [8, 9, 19]. This method will allow Eve to measure a given fraction of the secret key, which will also however raise error when her measurement basis differs from Alice [19].

Normally, the maximum fraction of the secret key which can be eavesdropped is taken into account at the privacy-amplification stage [15]. If however, the actual mean photon number μ is higher than the expected μ used in privacy amplification (as in the case of a damaged attenuator with its attenuation permanently decreased), Eve will then be able to obtain a nonzero fraction of the secret key.

3 Experimental techniques

The Quantum Hacking group aims to show that there exist hardware vulnerabilities in QKD systems which can be exploited by a potential attacker. As explained in the BB84 section above, the variable optical attenuator (VOA), which makes sure that the outgoing signal is transmitted by single photons, is a key component assuring the physical security of the device. If it is possible to reduce the attenuation of the VOA from the network-side, either permanently or temporarily, an attacker can then remotely change the critical attenuation setting in step 1 above allowing for a higher than expected fraction of bits to be sent via multiple photons, which can then be eavesdropped via a photon-number splitting (PNS) attack [8, 9].

3.1 Fiber fuse protection

The "fiber fuse" is a curious phenomenon which occurs when an exceptionally high power is sent through single-mode optical fibers. Being accurately named, the visual appearance of the ”fiber fuse” is that a segment of the fiber lights up akin to a pyrotechnic fuse, which then travels down the fiber towards the high powered optical source [20, 21]. The physical principle of such a phenomenon is that a small defect in either the fiber itself, or more commonly along a splice or joint, will have a slight impedance change for the incoming light. This small change in impedance then leads to localized light absorption and thermal heating of the optical fiber [20, 22].

When a critical power density is reached at the location of the imperfection, the glass of the fiber undergoes fusion, which dramatically increases the local impedance change even more. This leads to an immediate heating of the area upstream of the imperfection, which also undergoes fusion. The cycle repeats again and again, with the location undergoing fusion travelling upstream along the fiber towards the source of high-powered light. The visual effect is that of a bright spot of light travelling along the fiber at a speed on the order of 1 m/s [23].

The fiber fuse process is a destructive process. This means that our high-powered laser source needs a protective mechanism to prevent the fiber fuse from travelling into the amplifier. A previous undergraduate engineering project has been to design a small interlock device with two photodiodes placed near the fiber while optically isolated from the room. As soon as one of the photodiodes detects light from the fiber fuse, a shutdown signal is sent to the laser source. Previously, this safety mechanism has been successfully tested on our experimental setup.

3.2 Optical setup

Figure 2. shows the optical setup used for testing our attenuators under optical damage attack. The entire experiment is done at the 1550 nm wavelength commonly used in telecommunications. The eavesdropper Eve sends out a high power laser from an erbium-doped fiber amplifier (EDFA) and attempts to disrupt Alice’s variable attenuation in her QKD system. Typically, the variable optical attenuator (VOA) is the last element in Alice’s QKD setup. To measure the resulting change in attenuation and whether the attack is successful, Eve checks the power of the Alice’s laser using power meter C (Thorlabs PM200 with S154C).

The high powered laser is an erbium-doped fiber amplifier manufactured

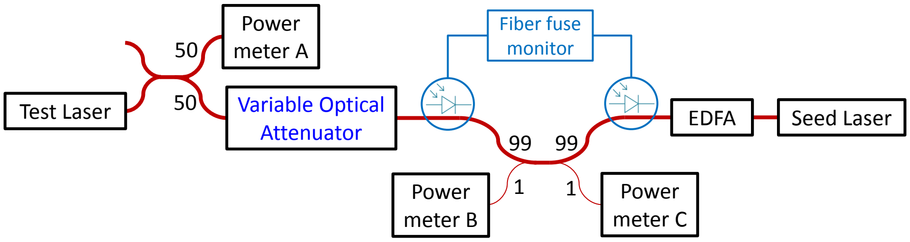


Figure 2: Diagram of the optical setup used for attenuator testing, illustrating the fiber fuse monitor, as well as the systems representing Alice and attacker Eve in the QKD scheme.

by QGLex Inc. paired with a 1550 nm seed laser at 200 mW providing the original optical input to be amplified. The maximum output power of this setup is 39dBm, which is approximately 8W.

The test laser is a low power single-mode 1550 nm fiber-pigtail laser diode (Gooch & Housego AA1406), and is used to provide an initial source of light for measuring attenuation of the VOA. Power meter B (Thorlabs PM200 with S146C detector) is used to monitor the EDFA power. Power meter C, connected after the test laser and the VOA, serves to check the attenuation of the VOA before and after optical damage. Power meter A (Joinwit JW3208) is used to monitor the power of the test laser. The fiber fuse monitor is placed before and after the four-terminal 99:1 fiber coupler, and is connected to the EDFA interlock to immediately shut off the EDFA in case a fiber fuse occurs.

A digital single-lens reflex (DSLR) camera and a separate thermal imaging camera using a microbolometer array (Flir E60) is placed over the VOA to observe the physical response to the high power laser.

The optical setup for our experiment is designed such that no component is being reconnected between the high powered laser "attack" phase and the measurement phase. Thus as long as the optical channel remains undisturbed across a single test sequence, a change in the measurement ratio between power meter A and power meter C can be attributed to the VOA. This is however, true only is none of the fiber couplers or power meter probes have changed before and after the testing scheme, which needs to be confirmed during the experiment.

The attenuation in dB of the VOA, as well as its measurement uncertainty, can be calculated by the following equations:

$$A(P_A, P_C) = 10log_{10}(P_A) - 10log_{10}(P_C) - 10log_{10}(100) \tag{2}$$

$$\Delta A(P_A, P_C) = \sqrt{\left(\frac{10\Delta P_A}{P_A ln(10)}\right)^2 + \left(\frac{10\Delta P_C}{P_C ln(10)}\right)^2} \tag{3}$$

where P_A and P_C are the powers (in mW) measured from power meter A and C, respectively. The factor in the logarithm comes from the 99:1 fiber coupler, which further attenuates the measurement P_C by a factor of 100. The measurement uncertainty formula is derived from partial derivatives.

3.3 MEMS variable optical attenuator (VOA)

The variable optical attenuators (VOA) we have tested use a movable MEMS mirror to control the amount of coupled light between the input and outputs. MEMS stands for Micro Electro-Mechanical Systems, and are often manufactured using a nanofabrication process similar to traditional microelectronics.

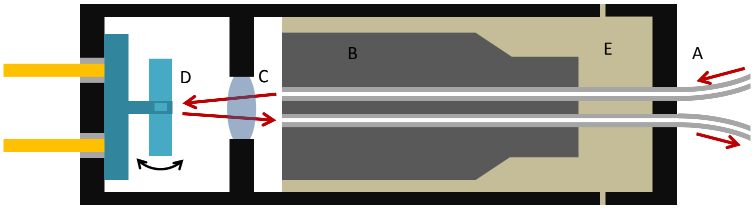


Figure 3: Simplified schematic of the MEMS variable optical attenuator with labeled parts (not to scale). (A) Incoming and outgoing single-mode fibers; (B) Glass ferrule; (C) Collimation lens; (D) Voltage adjustable MEMS mirror on torsion mount; (E) Adhesive filler.

As shown in Fig. 3, both the optical fiber carrying the incoming and outgoing signals are encased in a glass ferrule for mechanical stability. The beam of light first exits the incoming fiber and is directed towards a single large collimation lens. The light is then reflected off the MEMS mirror assembly and directed towards the second, outgoing fiber. The reflected light is then coupled to the outgoing fiber through the same collimation lens. The position of the MEMS mirror is controlled electrostatically via an applied voltage, which changes the amount of light coupled to the outgoing fiber [24]. This process achieves adjustable attenuation between the input and output fibers in the range of approximately 0dB to >50dB.

A total of 8 variable optical attenuators (VOA) have been provided by the third-party participant for our experiment. These attenuators are all of the MEMS type, with 4 being from one manufacturer and the other 4 from a separate manufacturer, labelled type A or B, respectively. These attenuators are arranged in pairs on a PCB (with attenuators from the same manufacturer sharing the same PCB), making 4 PCBs in total for the 8 attenuators to be tested. The VOA's are are attached using a soft silicone glue to the PCB assembly, providing a thermally accurate representation of the actual QKD system. Due to non-disclosure agreements associated with the project, any manufacturer names and identifying serial number of the attenuators are not shown in this report.

3.4 Experimental procedure

It is a necessity for the seed laser to be turned on before any other component, since the EDFA is only designed for a fixed input power range. The 1550nm seed laser for the EDFA is controlled using a Thorlabs CLD1015 laser driver, and is first turned on to a current of 200mA with thermoelectric temperature control set to 25°C. For our laser, this corresponds to a measured power of 22.98 (±0.01) dBm or approximately 200 mW.

After the seed laser is set to the correct power, the EDFA is switched on along with the fiber fuse detector. To verify the accuracy of the stated output power from the included control software, we first physically measure its power, using power meter B connected directly to the EDFA without the 99:1 fiber coupler in between. The results are shown in the following figure. Using least-squares curve fitting, the best fit degree-3 polynomial is given by:

$$f(x) \approx -0.0009637x^3 + 0.1051x^2 - 2.787x + 45.464 \tag{4}$$

This is the formula used to determine the actual power from the EDFA for future measurements. The difference between the setpoint and the measured output power is due to the EDFA having a minimum amplification gain, explaining the larger difference in the low ranges of output power. Note that this method effectively ignores backscattering and any nonlinearities which can occur for a more complex experimental setup. However for our purposes, only a reasonably accurate control of the optical power sent to the attenuator is sufficient.

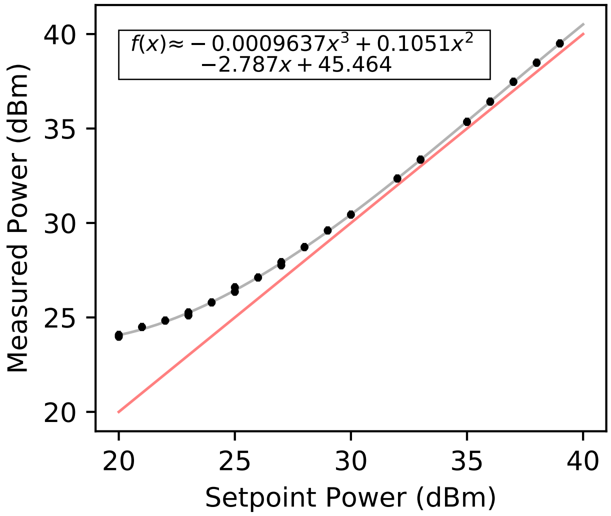


Figure 4: Measured power curve for the EDFA output with a seed laser power of 200mW, compared to the software setpoint. Red line denotes the ideal EDFA power curve.

The EDFA power output is then shut off after the preliminary power curve measurement. The test laser is turned on and its current set such that power meter A measures approximately 7dBm (5mW). This value will be used throughout the experiment to calculate the attenuation of the VOA. Afterward, the procedure described below is done for each tested attenuator:

1. Establish the attenuation vs. voltage curve for the VOA being tested. This process is done with the EDFA turned off.

2. A testing voltage is determined based on two factors. The first being that the VOA needs to be set to a sufficiently high attenuation to prevent damage to the test laser. The second factor is that the attenuation needs to fall reasonably within the value currently used in the targeted QKD systems. This operation range is given by the participating third-parties companies.

3. The EDFA is set to a starting value of about 30 dBm (1W). This starting power is determined to be below the damage threshold both from manufacturer specification and previous experiments.

4. The video camera and infrared camera are both started, to record physical and thermal behaviour of the attenuator under optical stress.

5. The EDFA is turned on and maintained at the desired set point, for at least 60 seconds, until the maximum temperature measured by the thermal camera and the measured value at power meters C both reach a stable plateau.

6. The EDFA is turned off with the video and IR cameras remaining on, recording the cooldown process. In the meantime, power meter C is set to log the measured power.

7. Once the measured value at power meter C stabilizes, the video and IR cameras are shut off. The measured optical power is recorded to calculate the attenuation after optical damage at a given power.

8. If the attenuation after optical damage is similar to the reference attenuation, the EDFA power setpoint is incremented by 0.5–1dBm depending on how close it is to the expected damage threshold (going to step 3). However if the attenuations before and after are significantly different, the experimental scheme is stopped (proceeding to next item).

9. The attenuation vs. voltage curve is measured again (with EDFA turned off) and compared with the curve taken in step 1, before any optical damage.

Through the first experimental trials, we have noticed that there is a temporary increase in attenuation while the VOA is still in thermal disequilibrium right after the laser has been turned off. To confirm whether any change in attenuation after optical damage is permanent and not due to thermal reasons, step 9 above is measured several minutes after the EDFA is shut off, when the observed attenuation stabilizes and the case temperature returns back to ambient as seen from the thermal camera. To further confirm the change in attenuation, an additional measurement is done 2–5 days afterwards.

4 Results and discussion

From our set of 8 attenuators, 6 are tested (3 from each manufacturer), with the two remaining serving as reference samples for future use. To recapitulate, our experiment aims to verify whether there can be a drop in attenuation following optical damage. We can define a successfully "hacked" sample as one having a permanent drop in attenuation post-damage, for any given range of input voltages within specifications.

#	Manufacturer	Testing Voltage (V)	Attenuation Before (dB)	Attenuation After (dB)	Δ (dB)	Threshold of Damage (dBm)
1	A	12.0	33.05	35.44	+2.39	36.9
2	A	12.0	33.88	32.95	−0.93	37.4
3	A	11.5	32.81	64.28	+31.47	37.9
4	B	14.0	38.79	32.32	−6.47	35.9
5	B	14.5	≈ 68*	58.82	≈ -9.2	36.4
6	B	13.5	31.21	22.29	−8.92	34.8

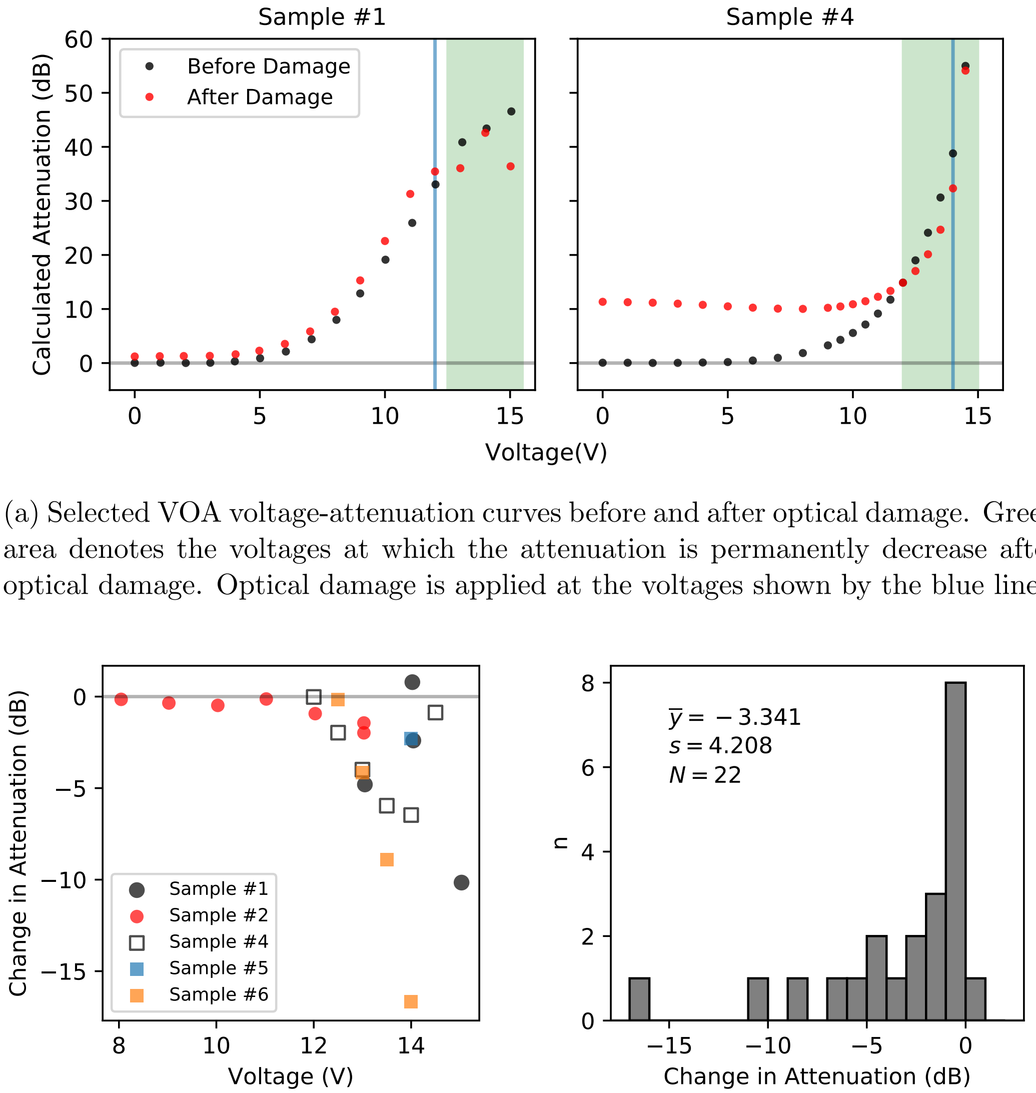
*Measurement near the minimum power range of the power meter

Table 1: Results after optical damage for MEMS VOA samples. Testing voltage refers to the parameter used in step 2 of the experimental procedure. The Δ column refers to the change in attenuation observed at the testing voltage.

At the testing voltage, 4 out of the 6 attenuators tested exhibited a permanent drop in attenuation. Furthermore, for 5 out of 6 attenuators, there exists some voltage range in which the attenuation is decreased post-damage, as shown by the voltage-attenuation curves and histogram in Fig. 5, with an average drop in attenuation of 3.3 dB and a maximum of 16.6 dB in sample 6. We can deem the optical damage attack to be successful for these range of voltages.

Attenuator #3 exhibited a near-total failure, where the attenuation after optical damage is dramatically increased from its normal value. Effectively, this is similar to a component failing "open" in electronics. This sample represents a case of total component failure, which is the undesired outcome when performing the hypothesized optical damage attack.

The situation of attenuator #5 is peculiar. While first measuring the attenuation-voltage curve by incrementing the applied voltage, the attenuation value appears to become latched after 14.5V (before activating the high optical power). Subsequent voltage adjustments down to even 0V did not change this measured attenuation. The 14.5V voltage however, does appear to be in the working range for the other attenuators from manufacturer B. Since the applied voltage is close to the maximum voltage specified, it is likely that the latching observed at this voltage is from to inherent variability in the



(a) Selected VOA voltage-attenuation curves before and after optical damage. Green area denotes the voltages at which the attenuation is permanently decrease after optical damage. Optical damage is applied at the voltages shown by the blue lines.

(b) Change in attenuation observed after damage, starting from first voltage at which $Att_{after} < Att_{before}$

(c) Histogram of the measured change in attenuation as shown in (b)

Figure 5: VOA voltage-attenuation curves before and after successful optical damage attack, showing the permanent drop in attenuation for 5 out of 6 samples. Error bars are smaller than the marker size.

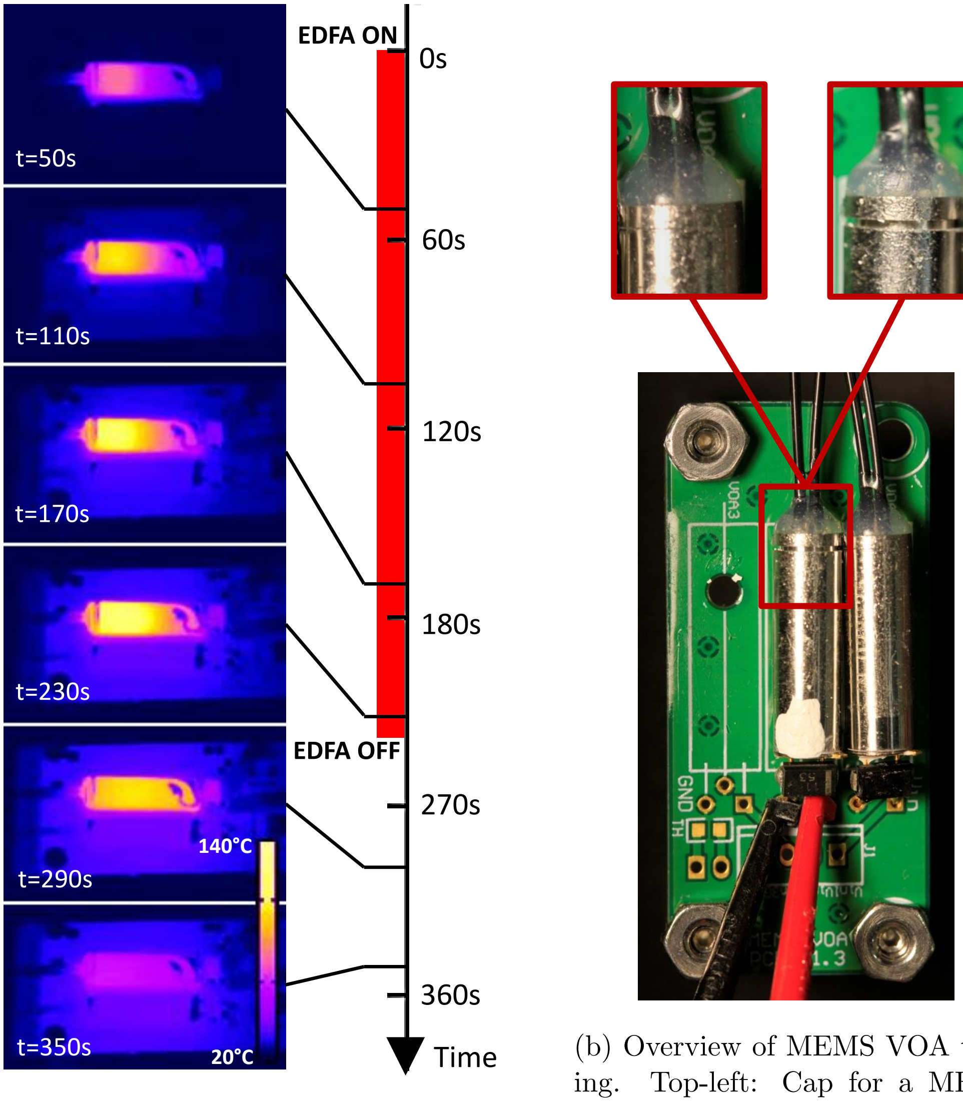
working voltage range between components. However despite this unexpected malfunction, a permanent decrease in attenuation is observed for this sample.

4.1 Possible damage mechanisms

Observing the behaviour of the MEMS VOA under high optical power, the front end of the VOA casing (end with input/output fibers) has a higher temperature, shown in Fig. 6a. This means that the majority of the optical power is dissipated near the front end of the VOA, corresponding to the extremity of the glass ferrule in Fig. 3. The cap holding the input and output fibers appears to bulge outwards near the threshold of damage (Fig. 6b), possibly pulling the fiber inside the VOA out of alignment with the collimation lens. In a catastrophic damage scenario near 38 dBm ($\approx 6.3W$), the cap detaches itself from the attenuator casing, with smoke emitted. Since the process of coupling a beam of light into a single-mode fiber is highly dependent on the relative positions of the involved optical elements [25], we hypothesize that the structural deformation under high temperatures is one of the possible causes responsible for the observed change in attenuation.

Following successful optical damage, one VOA is physically disassembled to check for traces of damage. The result is shown in Fig. 7.

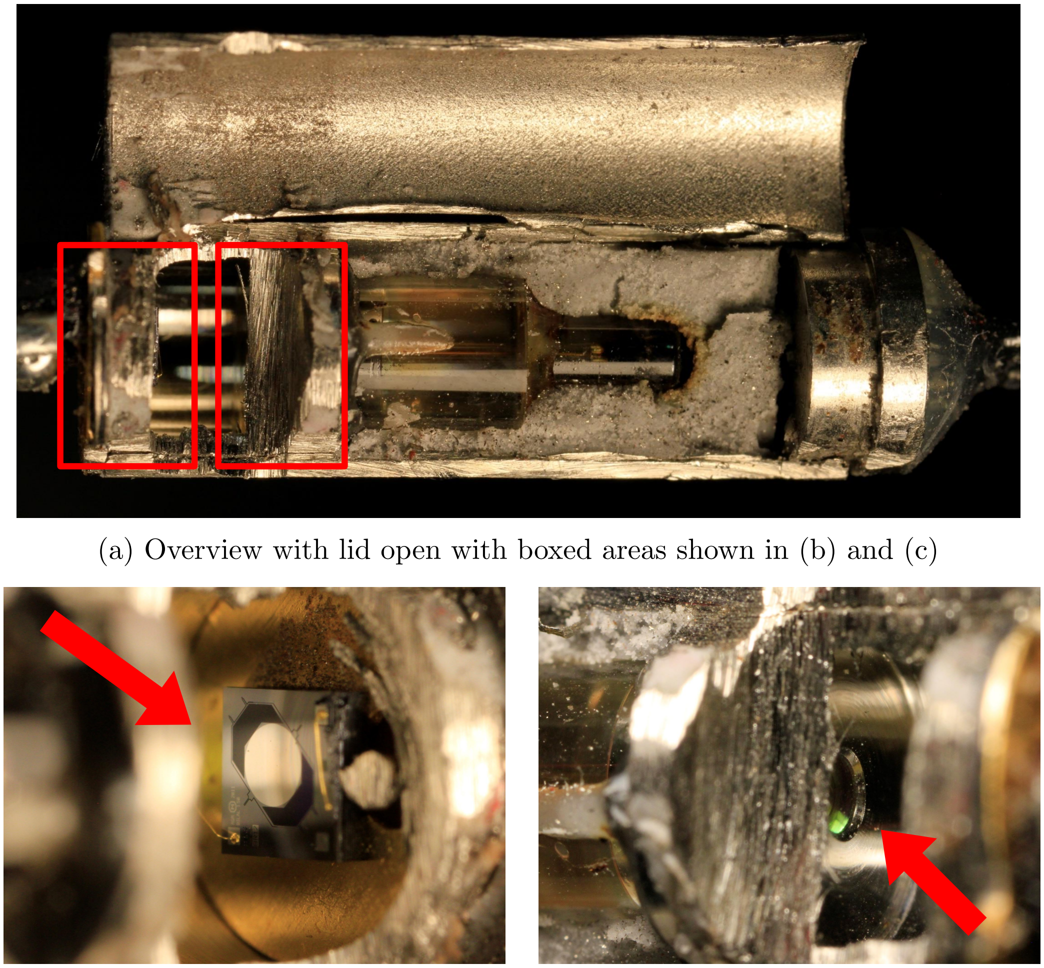
Another possible cause is that for typical MEMS materials used (Si, SiN, SiC, etc.) [26], the operating temperature can induce lattice strain and change its spring constant [27]. The ductility of polycrystalline Si is reported to increase at temperatures near 500°C [27]. Thus since the MEMS micromirror used in the VOA is fixed using a torsion mount, the amount of deflection induced by a given voltage can change with temperature. The voltage-attenuation curve will be different once the VOA heats up and exceeds its proper operating temperature range, which can either result in a drop in attenuation or an increase, depending on the exact material behaviour under high temperature. Our observations using the thermal camera show that the outer casing of the VOA reached 120°C at the damage threshold power. However, from the observations and physical disassembly, it appears that the area near the fiber end of the VOA is visibly more affected by thermal damage, which attributes some doubt to this mechanism of damage.



(a) Temperature profile of the VOA from manufacturer B near the threshold of damage. The high power laser was set to 34.5 dBm (2.8 W) and turned on between $t = 0s$ and $t = 250s$.

(b) Overview of MEMS VOA testing. Top-left: Cap for a MEMS VOA sample before damage. top-right: Displaced cap for a catastrophically damaged VOA.

Figure 6: Temperature profile and physical deformation of a typical MEMS VOA under optical damage test.



(a) Overview with lid open with boxed areas shown in (b) and (c)

(b) Details of MEMS mirror

(c) Details of collimation lens

Figure 7: MEMS VOA from manufacturer A after successful optical damage attack resulting in permanently reduced attenuation. Signs of optical and thermal damage are visible as dark burn marks near the tip of glass sheath.

4.2 Real-world attack scenario

In an actual QKD network, the laser damage attack can be effectively conducted by an eavesdropper Eve attempting to compromise a given Alice node. Eve would have an EDFA setup similar to the one used in our experiment, and would proceed by cutting the fiber line and splicing her EDFA at the other end of Alice, removing Bob from the line entirely. Eve will then turn on the EDFA and increment the optical power applied, while monitoring the signal power from Alice. Once Alice's signal power spontaneously rises due to optical damage of the VOA, the EDFA is turned off and removed from the fiber line, with the line connected back to Bob. This attack however, is not without risks, as attenuators can also catastrophically fail under a high optical power, which results in an interrupted fiber line and a damaged QKD system for Alice. The necessary denial-of-service from the attack is also likely to trigger an alert from Alice and Bob.

4.3 Possible countermeasures

To prevent the laser damage attack against optical attenuators, a method of detecting or limiting the maximum power through optical fibers is needed. An analogous electronics element accomplishing a similar function for electrical current is the fuse. An optical version of the current-limiting fuse has been proposed using a TeO_2 soft glass segment inserted inline of a standard fiber, which can prevent pulses higher than approximately 1W (on the range of 1s) to pass through [28]. A high powered laser from Eve would then trigger the optical fuse and cut off the fiber path. This scheme has the advantage of being a simple and robust device, compared to possible active monitors for optical power, but would however result in denial of service in case of an attack.

5 Concluding remarks

From our experiments, we have confirmed that it is possible for the variable optical attenuators (VOA) used in QKD to exhibit a permanent attenuation drop after optical damage. For 5 of our samples, the average attenuation drop is 3.3 dB with a maximum of 16.6 dB. This attenuation decrease can be induced at a distance from the fiber network side by an attacker Eve equipped with a high powered laser. Our experiments effectively shows that the variable optical attenuator (VOA) does indeed represent a vulnerability which can potentially be exploited by an external attacker having access to the fiber network.

As for future work, further tests can be made to distinguish whether the primary cause for the observed damage is directly from the effect of high temperatures. For example, one could heat the entire MEMS VOA without applying high powered laser, and measure the resulting attenuation. There is also the possibility of testing the laser damage attack in a live QKD system instead of individual attenuator components.

6 Acknowledgements

This work would not have been possible without the guidance and support of Anqi Huang and Professor Vadim Makarov from the Quantum Hacking Group at the University of Waterloo. We also thank previous undergraduate student Krtin Kumar for constructing the fiber fuse monitor, and also the third-party participants for providing their respective attenuator assemblies for our experiment.

References

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

[2] 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.

[3] 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.

[4] 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 J. Quantum Electron.*, 52(11):8000211, 2016.

[5] 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.

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

[7] G.A. Wellbrock, T.J. XIA, and D.Z. CHEN. Quantum key distribution system, July 16 2013. US Patent 8,488,790.

[8] Gilles Brassard, Norbert Lütkenhaus, Tal Mor, and Barry C. Sanders. Limitations on practical quantum cryptography. *Phys. Rev. Lett.*, 85:1330–1333, Aug 2000.

[9] Norbert Lütkenhaus. Security against individual attacks for realistic quantum key distribution. *Phys. Rev. A*, 61:052304, Apr 2000.

[10] L. S. Huang, S. Adhikarla, D. Boneh, and C. Jackson. An experimental study of tls forward secrecy deployments. *IEEE Internet Computing*, 18(6):43–51, Nov 2014.

[11] Lance Fortnow. The status of the p versus np problem. *Commun. ACM*, 52(9):78–86, September 2009.

[12] Stefan Wolf. *Unconditional Security in Cryptography*, pages 217–250. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[13] David Kahn. *The Codebreakers*. Macmillan, 1996.

[14] Zia Ahmed Ron Rivest. *Computer and Network Security Lecture 2*. MIT, 1997.

[15] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. *How to Reduce your Enemy's Information (extended abstract)*, pages 468–476. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[16] Charles H. Bennett, François Bessette, Gilles Brassard, Louis Salvail, and John Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5(1):3–28, Jan 1992.

[17] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

[18] Sharon Goldwater. *Quantum Cryptography and Privacy Amplification.* SRI International, 1996.

[19] Bruno Huttner, Nobuyuki Imoto, Nicolas Gisin, and Tsafrir Mor. Quantum cryptography with coherent states. *Phys. Rev. A*, 51:1863, 1995.

[20] R Kashyap and KJ Blow. Observation of catastrophic self-propelled self-focusing in optical fibres. *Electron. Lett.*, 24:47–49, 1988.

[21] D. P. Hand and P. St. J. Russell. Solitary thermal shock waves and optical damage in optical fibers: the fiber fuse. *Opt. Lett.*, 13(9):767–769, Sep 1988.

[22] Woosung Ha, Yoonseob Jeong, and Kyunghwan Oh. Fiber fuse effect in hollow optical fibers". *Opt. Lett.*, 36(9):1536–1538, May 2011.

[23] Raman Kashyap. The fiber fuse - from a curious effect to a critical issue: A 25th year retrospective". *Opt. Express*, 21(5):6422–6441, Mar 2013.

[24] A.A. Godil. Micro electro mechanical system using comb and parallel plate actuation, January 4 2011. US Patent 7,863,799.

[25] J. Martin. *Coupling Efficiency and Alignment Sensitivity of Single Mode Optical Fibers*. University of Central Florida, 1979.

[26] Loughborough University. *An Introduction to MEMS (Microelectromechanical Systems)*. PRIME Faraday Partnership, 2002.

[27] William N. Sharpe. Tensile testing of mems materials at high temperatures. In *Advances in Experimental Mechanics IV*, volume 3 of *Applied Mechanics and Materials*, pages 59–64. Trans Tech Publications, 9 2005.

[28] Shin ichi Todoroki and Satoru Inoue. Observation of blowing out in low loss passive optical fuse formed in silica glass optical fiber circuit. *Japanese Journal of Applied Physics*, 43(6A):L728, 2004.