Seguridad de los esquemas prácticos de criptografía cuántica contrafáctica

Carlos Navas Merlo
Universidad de Valladolid

Juan Carlos Garcia-Escartin (Tutor)
Universidad de Valladolid, Dpto. Teoría de la Señal e Ing. Telemática, Paseo Belén nº 15, 47011 Valladolid, Spain
(Dated: July 24, 2020)

En el protocolo cuántico contrafáctico N09 de distribución de claves propuesto por Noh [T.-G. Noh, Phys.Rev.Lett. 103, 230501 (2009)] se permite el intercambio de claves con los fotones que han podido ser medidos por Bob pero no han salido al canal cuántico. Se han publicado diversas demostraciones sobre su seguridad frente a ataques cuando en el protocolo participa un único fotón, o para ataques generales en el caso del uso de estados coherentes. No obstante, en todas ellas se supone que no se puede obtener información sobre la medida que está haciendo Bob. Sin embargo, se ha demostrado que existen técnicas eficaces para obtener tal información aprovechando las características de los detectores. En este artículo se modela y se demuestra la efectividad de un ataque que combina cegado y estados falsos junto con las contramedidas para evitarlo.

I. INTRODUCCIÓN

La formulación de la mecánica cuántica ha abierto un amplio abanico de nuevas líneas de investigación y desarrollo tecnológico fruto de la aparición de nuevos algoritmos otrora inalcanzables con los sistemas clásicos [1].

La criptografía cuántica se basa principalmente en la propiedad de la medida cuántica [2]. Toda medida del sistema cuántico producirá un cambio irreversible en él, pudiéndose así detectar un intento de espiar el canal. Esta propiedad ha permitido el desarrollo protocolos de distribución cuántica de claves (QKD) en los que los extremos usan un canal publico inseguro para generar una clave secreta con ayuda de un canal público autenticado [3–5]. Por ejemplo, dos de los protocolos de distribución cuántica de claves más importantes, el BB84 [3] y el E-91 [4] basan su seguridad en este principio.

Un grupo importante de los protocolos QKD son los protocolos basados el concepto de medida libre de interacción (*interaction-free measurements*) propuesto por Elitzur y Vaidman [6]. En su artículo, Elitzur y Vaidman. exponen un experimento donde utilizan un interferómetro de tipo Mach-Zehnder y una fuente de un único fotón. En primer lugar, este fotón alcanza el primer divisor de haz con coeficiente de transmisión 1/2. A continuación, las funciones de onda transmitida y reflejada del fotón se reflejan en espejos orientados de forma que éstas interfieran en un segundo divisor de haz. El interferómetro se puede organizar de tal forma que cuando haya interferencia destructiva no se detecte nada en un detector D2, pero sí en un detector D1. En el caso de producirse una destrucción de la interferencia constructiva mediante la inserción de un obstáculo en uno de los brazos, como por ejemplo una bomba que se activase al interaccionar con ella con un único fotón, habría un 25% de probabilidades de que el fotón fuera detectado por el detector D2 sin haber interactuado con la bomba. Es decir, la bomba podría ser detectada sin detonarla debido a la presencia del fotón.

Si bien los protocolos QKD han demostrado un aumento en la seguridad respecto a sus equivalentes clásicos [7], siguen estando sujetos a posibles hackeos cuánticos como por ejemplo, ataques *man-in-the-middle* [8], ataques caballo de Troya [9], etc. Este artículo se centrará en el grupo de protocolos QKD de medida libre de interacción, concretamente en el protocolo propuesto por Noh [10], llamado de ahora en adelante N09 en este artículo.

El protocolo N09 está basado en el efecto contrafáctico cuántico. El nombre contrafáctico se debe a que la clave cribada, es decir, la clave que obtienen Alice y Bob tras el autentificado por un canal publico, se obtiene con los fotones que han podido ser medidos por Bob pero se han quedado en el sitio de Alice. Debido a esto último, este protocolo es muy atractivo desde el punto de vista de la seguridad.

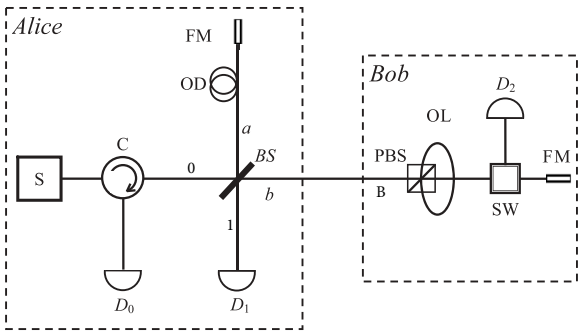

Figura 1: Esquema del sistema de distribución cuántica de claves del protocolo N09. D0, D1, D2 se corresponden con tres detectores, C es un circulador, OD y OL representan dos lineas de retardo, SW se corresponde con un switch y FM con espejos de Faraday. Los espejos de Faraday tienen la función de prevenir efectos de dispersión en el canal cambiando el estado de la polarización. Este cambio de polarización será ignorado en el ataque por simplicidad.

El esquema completo del protocolo N09 se muestra en la figura 1. En primer lugar, Alice produce un pulso óptico corto que contiene únicamente un fotón. A continuación, prepara el estado del fotón $|\psi\rangle$ con polarización horizontal $|H\rangle$ o vertical $|V\rangle$, elegidos al azar. Tras el paso por el divisor de haz el pulso se divide en dos caminos, a y b. En este momento, el estado inicial $|\psi\rangle$ evoluciona a

$$\left|\phi_{0(1)}\right\rangle = \sqrt{T}\left|0\right\rangle_a\left|\psi\right\rangle_b + i\sqrt{R}\left|\psi\right\rangle_a\left|0\right\rangle_b \tag{1}$$

donde R y $T = 1 - R$ son el coeficiente de reflexión y transmisión del BS respectivamente. $|0\rangle$ denota el estado vacío, y los subíndices $a(b)$ indican el estado transmitido por el camino $a(b)$. Únicamente el pulso transmitido a la salida b llega a Bob a través del canal cuántico público. Por otro lado, Bob puede elegir que polarización redirige hacia su detector y cual refleja de vuelta al canal. Cuando Alice y Bob eligen la misma polarización el sistema cuántico evoluciona a

$$\begin{aligned}\left|\phi_s\right\rangle &= \sqrt{T}\left|0\right\rangle_a\left|\psi\right\rangle_b + i\sqrt{R}\left(\sqrt{T}\left|\psi\right\rangle_1\left|0\right\rangle_0 + i\sqrt{R}\left|0\right\rangle_1\left|\psi\right\rangle_0\right)\\ &= \sqrt{T}\left|0\right\rangle_a\left|\psi\right\rangle_b + i\sqrt{RT}\left|\psi\right\rangle_1\left|0\right\rangle_0 - R\left|0\right\rangle_1\left|\psi\right\rangle_0\end{aligned} \tag{2}$$

de esta expresión se deduce que hay tres posibilidades:

1. Con una probabilidad $T/2$ el fotón se detecta en el detector D2 en el sitio de Bob.
2. Con una probabilidad $R^2/2$ el fotón se detecta en el detector D0.
3. Con una probabilidad $RT/2$ el fotón se detecta en el detector D1.

El factor $1/2$ se introduce debido a que hay un 50% de probabilidades de que Alice y Bob elijan la misma polarización.

En el caso de que el fotón no se redirigiera hacia el detector de Bob, lo que significaría que Bob y Alice han elegido diferente polarización, se reflejará en el espejo de Bob de vuelta al canal donde se le introducirá un desplazamiento de fase π. Esto último, producirá una interferencia en el BS de Alice entre el pulso que estaba en el camino a y el que vuelve por el camino b. El estado final será

$$\begin{aligned}\left|\phi_s\right\rangle &= -\sqrt{T}\left(\sqrt{T}\left|0\right\rangle_1\left|\psi\right\rangle_0 + i\sqrt{R}\left|\psi\right\rangle_1\left|0\right\rangle_0\right)\\ &\quad + i\sqrt{R}\left(\sqrt{T}\left|\psi\right\rangle_1\left|0\right\rangle_0 + i\sqrt{R}\left|0\right\rangle_1\left|\psi\right\rangle_0\right)\\ &= -\left|0\right\rangle_1\left|\psi\right\rangle_0\end{aligned} \tag{3}$$

detectándose con total seguridad en D0 con un factor $1/2$ debido a que hay un 50% de probabilidades de que Alice y Bob elijan diferente polarización. Por último, para obtener la clave final Alice escoge únicamente los eventos donde se ha producido una detección en D1 y los hará públicos. En este caso ambos lados conocen la polarización consiguiendo un bit en común. Alice y Bob logran una clave idéntica sin que el fotón salga de Alice. Con esta última interpretación se debe tener precaución, puesto que se está interpretando de forma clásica un efecto puramente cuántico dado que, una vez el fotón atraviesa el BS de Alice se obtienen dos estados cuánticos entrelazados.

Tras la publicación del protocolo N09 se han hecho diversos análisis de seguridad ante diversos ataques, entre los que destacan el ataque de tipo interceptación y reenvío [11] y de tipo caballo de Troya [12]. Sin embargo, existen algunas limitaciones en estas demostraciones, como la que supone que los láseres emiten un único fotón, cuando en todos los experimentos se usan estados coherentes[13–15]. La demostración de seguridad del protocolo N09 cuando intervienen estados coherentes se da en el artículo de Yin et al.[16] donde se proporciona una cota de clave segura para un ataque general. Por ejemplo, que para unos valores típicos de transmisión del canal y efectividad de los detectores se podría obtener una clave segura hasta 50 km de canal público con fibra óptica estándar.

No obstante, la demostración supone que no se puede obtener la polarización con la que Bob mide. Sin embargo, se han demostrado ataques con estados invisibles que permiten obtener tal información [17, 18].

En este artículo se aprovecha otra característica de los detectores que permiten su cegado y forzar una detección como se demuestra en el artículo de V.Makarov [19]. El artículo citado explica como es posible enviar luz con la que se consiga cegar o forzar un click en los detectores. Esto último dependerá de si mantenemos la potencia del haz de luz en un determinado rango o si en una ventana de tiempo de microsegundos detenemos el cegado. Estas ventanas de tiempo se deben a que los detectores en infrarojo tienen una gran cantidad de cuentas en la oscuridad y para mejorar su eficiencia en activan en periodos cortos de tiempo.

El uso de esta estrategia ha demostrado gran efectividad en ataques de estados falsos en otros protocolos QKD importantes como el BB84 [20].

Es por ello, que se considera necesario un análisis de seguridad frente a ataques que hagan uso de estas estrategias.

En este articulo se demuestra la efectividad de un ataque para clave infinita que aprovecha las características de los esquemas prácticos comentadas anteriormente. En la sección 2 se presentan las características experimentales del sistema que justifican el ataque. En la sección 3 se presenta un ataque efectivo contra el protocolo ideal de un único fotón y contra el de estados coherentes. No obstante, este presentará unos requisitos muy restrictivos. En la sección 4 se introduce un ataque más general, seguido de su modelado para los diferentes esquemas experimentales propuestos hasta la fecha. En primer lugar, se trata el esquema donde todos los detectores son simples, es decir, detectores de click o no click sin diferenciar la polarización. En el siguiente apartado, se modela el ataque cuando todos

los detectores permiten distinguir las polarizaciones (detectores especializados) en detectar una determinada polarización. Esto se consigue introduciendo los detectores $D0H|D0V$, $D1H|D1V$, $D2H|D2V$ y un divisor de haz polarizado delante de los mismos que redirija los fotones según su polarización. Por último, se estudia el esquema donde Bob tiene el detector D2 especializado y Alice D0 o D1 especializado. En la sección 5 se realiza un estudio numérico de cada uno de los esquemas para mostrar su efectividad. Por último, en la sección 6 se establecen las conclusiones junto con las contramedidas que se deben seguir para evitar este ataque.

II. CARACTERÍSTICAS DE LOS ESQUEMAS PRÁCTICOS DEL PROTOCOLO

En los esquemas experimentales propuestos hasta la fecha se usan láseres atenuados como fuente de fotones [13–15], debido al precio o al bajo rendimiento de las fuentes de un solo fotón [21]. Es por esto, por lo que en todos lo esquemas experimentales del protocolo se trabaja con estados coherentes atenuados y no con un único fotón.

Los estados coherentes no producirán un entrelazamiento entre el estado en el sitio de Alice y el del canal al atravesar el divisor de haz, pudiéndose así medir en el canal sin que el estado en el sitio de Alice se vea afectado.

Por otro lado, para la transmisión de información se usará normalmente fibra óptica. Por ello se trabaja con láseres con una longitud de onda de 1550nm, pues es en esta frecuencia donde menos pérdidas se producen [22].

Debido a la longitud de onda de los láseres con los que se trabaja se utilizan detectores con eficiencias del 10%, ya que de aumentar la eficiencia se incrementaría demasiado las cuentas de oscuridad, llegando a ser para detectores con un 20% de eficiencia, seis veces mayor a las que se dan en los detectores con un 10% de eficiencia [23]. No obstante, existen detectores con diferentes tecnologías que pueden alcanzar eficiencias del 70% al 90% [23], siendo su uso únicamente académico, debido a su elevado precio y dificultad de construcción.

Por último, en los ataques propuestos en este artículo se está sujeto a los principios de Kerckhoffs, y en concreto, al segundo, donde se estable que el sistema criptográfico no debe suponerse secreto y no debe ser un problema que caiga en manos del enemigo. Esto se traducirá en que Eva conocerá todas las características del sistema que están usando Alice y Bob.

III. ATAQUE DE CEGADO Y REDUCCIÓN DE LA ATENUACIÓN

A continuación, se presenta el esquema de un primer ataque efectivo contra el protocolo ideal de un único fotón y contra el de estados coherentes. No obstante, este ataque solo será realizable cuando Eva puede reducir las pérdidas del canal a la mitad.

Protocolo de ataque:

Eva introducirá un divisor de haz polarizado (PBS) que permitirá la transmisión de una determinada polarización manteniendo a su vez la polarización reflejada en un circuito adjunto. Junto al estado transmitido, Eva enviará luz cegadora con la polarización opuesta usando el mismo divisor. Aquí se podrán dar los siguientes resultados:

- La polarización que transmite el PBS de Eva es la misma que han elegido Alice y Bob. En este caso, se producirá un click en el detector D2 de Bob. A su vez, Eva conocerá la polarización que ha elegido Bob debido a que se reflejará la luz cegadora que ésta le envió.

- La polarización que transmite el PBS de Eva es la misma que ha elegido Alice, pero diferente a la que ha elegido Bob. En este caso, el estado será reflejado para permitir la interferencia esperada en el divisor de haz de Alice. Eva conocerá la polarización que ha elegido Bob debido a que no se reflejará la luz cegadora que ésta le envió.

- La polarización que transmite el PBS de Eva es diferente a la que han elegido Alice y Bob. En este caso Bob perderá una posible detección. Eva conocerá la polarización que ha elegido Bob debido a que no se reflejará la luz cegadora que ésta le envió.

- La polarización que transmite el PBS de Eva es diferente a la que ha elegido Alice y la misma que ha elegido Bob. En este caso Eva reenviará a Alice el estado reflejado que mantiene en el circuito adjunto. Eva conocerá la polarización que ha elegido Bob debido a que se reflejará la luz cegadora que ésta le envió.

Para compensar la mitad de detecciones que ha sufrido Bob, Eva tendrá que duplicar la transmisión del canal y atenuar los estados que reenvía a Alice. Consiguiendo así, que se mantengan las estadísticas esperadas en todos los detectores.

Como se puede observar, este ataque es efectivo. No obstante, los requisitos para que sea realizable son algo restrictivos debido a que aunque se podría reducir un canal con perdidas de 0.2dB/Km a 0.1dB/Km [22], el ataque sería tecnológicamente complicado. Por ello, se presenta a continuación un ataque más general donde se reducen significativamente estos requisitos.

IV. ATAQUE DE CEGADO Y ESTADOS FALSOS GENERAL

A. Protocolo con todos los detectores simples

Para este ataque supondremos que Eva puede mejorar la transmisión del canal. Esta suposición es razonable puesto que se está considerando un aumento del 10% al 20% para los casos específicos en los que Alice elige una transmitancia de su divisor de haz muy alta a cambio de disminuir drásticamente la eficiencia del protocolo.

En primer lugar, Eva tendrá que elegir entre medir a la salida de Alice o no.

Eva elige no medir

En este caso Eva tendrá que seguir una de estas dos estrategias:

- **Cegar el detector D2 de Bob** con luz polarizada a 45° de tal forma que la polarización que Bob no está midiendo se refleje a Eva. Esto supondrá la pérdida de detecciones en el detector de Bob.
- **Seguir la estrategia que se expuso en el ataque anterior.** En este caso solo, será necesario el aumento de la transmisión del canal un 10% o 20% para compensar las pérdida de detecciones de Bob.

Eva elige medir a la salida de Alice

En este caso se podrán dar dos situaciones:

- **Eva no detecta nada.** En este caso la forma de proceder de Eva será:

  1. Comprueba que polarización está midiendo Bob, enviando luz polarizada a 45° de tal forma que la polarización con la que no está midiendo se refleje a Eva.
  2. Introducirá en Alice el estado coherente con la misma amplitud que esperaría y con una polarización opuesta a la que ha detectado en Bob. Además, Eva decidirá entre cegar o no todos los detectores de Alice para mantener las estadísticas esperadas. En el caso de que no cegarlos se podrán dar dos situaciones:

  – Alice y Bob eligen la misma polarización. En este caso, al introducir Eva un estado con la polarización opuesta a la que tiene Bob y por lo tanto, opuesta a la que tiene Alice, se producirá un mayor número de detecciones en D0 y D1.

  – Alice y Bob eligen diferente polarización. En este caso, al introducir Eva un estado con la polarización opuesta a la que tiene Bob y por lo tanto, la misma que tiene Alice, se seguirá produciendo la interferencia constructiva esperada (D0).

- **Eva detecta los fotones del canal.** En este caso Eva actuará dependiendo de la polarización en la que Bob está midiendo:

  – La polarización del estado medido es la misma que está midiendo Bob. En este caso Eva asegurará un click en el detector D2 de Bob.

  – La polarización de los fotones detectados es diferente a la que Bob está midiendo. Eva reintroduce el estado en Alice con la amplitud esperada para que haya interferencia constructiva (D0).

Obtención de las las estadísticas que esperan Alice y Bob

Alice introduce en el sistema un estado coherente $|\alpha\rangle$ que en la representación de estados número con la misma polarización será de la forma

$$|\alpha, H(V)\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{(n!)^{1/2}} |n, H(V)\rangle \tag{4}$$

siendo $\langle n\rangle = |\alpha|^2$ el número medio de fotones. A la salida del divisor de haz el estado evoluciona a [24]

$$|\alpha\rangle_0 |0\rangle_1 \rightarrow |r\alpha, H(V)\rangle_a |t\alpha, H(V)\rangle_b \tag{5}$$

con $T = t^*t$ y $R = r^*r$, donde r y t satisfacen las relaciones $r^*r + t^*t = 1$ y $r^*t + rt^* = 0$.

Por otro lado, cuando Bob elige la misma polarización que Alice, el sistema cuántico evoluciona a

$$|\sigma r\alpha\rangle_a |0\rangle_b \left|\sqrt{\sigma}t\alpha\right\rangle_B \rightarrow |\sigma r^* r\alpha\rangle_0 |\sigma t^* r\alpha\rangle_1 \left|\sqrt{\sigma}t\alpha\right\rangle_B \tag{6}$$

donde por simplicidad se ha omitido la polarización. En esta ecuación, σ representa la transmisión del canal.

De la ecuación (6) se deduce que las probabilidades de detección de uno o más fotones en los diferentes detectores son

$$P_{D0}(n > 0) = \frac{1}{2}(1 - e^{-\eta_0 \sigma^2 |r^* r\alpha|^2}) \tag{7}$$

$$P_{D1}(n > 0) = \frac{1}{2}(1 - e^{-\eta_1 \sigma^2 |r^* t\alpha|^2}) \tag{8}$$

$$P_{D2}(n > 0) = \frac{1}{2}(1 - e^{-\eta_2 \sigma |t\alpha|^2}) \tag{9}$$

Donde el 1/2 se introduce debido a que hay un 50% de probabilidades de que Alice y Bob elijan la misma polarización. Las eficiencias de los detectores D0, D1 y D2 se han representado por η_0, η_1 y η_2 respectivamente.

Por otro lado, cuando Alice y Bob eligen una polarización diferente, el estado que entra en sitio de Bob es reflejado de vuelta al canal y en Alice tenemos la evolución

$$\begin{aligned} |\sigma r\alpha\rangle_a |\sigma t\alpha\rangle_b &\to |\sigma(r^* r\alpha + t^* t\alpha)\rangle_0 |0\rangle_1 \\ &= |\sigma\alpha\rangle_0 |0\rangle_1 \end{aligned} \tag{10}$$

De la ecuación (10) se deduce que la probabilidad de detección de uno o más fotones en el detector D0 es

$$P_{D0}(n>0) = \frac{1}{2}(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \tag{11}$$

Modelado del ataque

Supongamos que Eva mide y no detecta fotones en el canal. En este caso, Eva introduce un estado con la misma amplitud que espera Alice y polarización opuesta a la que está midiendo Bob. Luego el estado en el sitio de Alice cuando Alice y Bob eligen la misma polarización será

$$|\sigma r\alpha\rangle_a |\sigma t\alpha_\perp\rangle_b \to |\sigma r^* r\alpha\rangle_0 |\sigma t^* t\alpha_\perp\rangle_0 |\sigma t^* r\alpha\rangle_1 |\sigma r^* t\alpha_\perp\rangle_1 \tag{12}$$

donde $\alpha_\perp$ simboliza el mismo estado que α pero con una polarización ortogonal. Por otro lado, cuando Alice y Bob eligen diferente polarización se tiene

$$|\sigma r\alpha\rangle_a |\sigma t\alpha\rangle_b \to |\sigma\alpha\rangle_0 |0\rangle_1 \tag{13}$$

A continuación, si llamamos $x \in [0,1]$ a la proporción de pulsos que Eva mide y $z \in [0,1]$ a la proporción de pulsos que ciega los detectores de Alice, la probabilidad de detección de fotones en los detectores de Alice cuando Alice y Bob eligen la misma polarización vendrá dada por

$$\begin{aligned} P_{D1(1)}(n>0) = \frac{1}{2}(1-z)xe^{-\eta_E|t\alpha|^2} \\ \cdot (1 - e^{-\eta_1\sigma^2(|r^* t\alpha|^2 + |t^* r\alpha_\perp|^2)}) \\ + \frac{1}{2}(1-x)(1 - e^{-\eta_1\sigma^2|r^* t\alpha|^2}) \end{aligned} \tag{14}$$

$$\begin{aligned} P_{D0(1)}(n>0) = \frac{1}{2}(1-z)xe^{-\eta_E|t\alpha|^2} \\ \cdot (1 - e^{-\eta_0\sigma^2(|r^* r\alpha|^2 + |t^* t\alpha_\perp|^2)}) \\ + \frac{1}{2}(1-x)(1 - e^{-\eta_0\sigma^2|r^* r\alpha|^2}) \end{aligned} \tag{15}$$

donde η_E es la eficiencia del detector de Eva.

Cuando Alice y Bob eligen diferente polarización se tiene para el detector D0 que

$$\begin{aligned} P_{D0(1)}(n>0) = \frac{1}{2}(1-z)xe^{-\eta_E|t\alpha|^2}(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \\ + \frac{1}{2}(1-x)(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \\ + \frac{1}{2}x(1 - e^{-\eta_E|t\alpha|^2})(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \end{aligned} \tag{16}$$

Por último, si llamamos $y \in [0,1]$ a la proporción de pulsos que Eva ciega a Bob la probabilidad de detección de fotones del detector de Bob vendrá dada por

$$\begin{aligned} P_{D2(1)}(n>0) = \frac{1}{4}(1-y)(1-x)(1 - e^{-\eta_2\sigma'|t\alpha|^2}) \\ + x\frac{1}{2}(1 - e^{-\eta_E|t\alpha|^2}) \end{aligned} \tag{17}$$

donde σ' representa la transmisión aumentada del canal.

B. Protocolo con todos los detectores especializados

A diferencia del caso anterior, Eva tendrá que cegar siempre que mida y no detecte fotones. No obstante, Eva podrá cegar los detectores con la polarización que ella elija enviando luz cegadora con la polarización deseada. Por ello, Eva cegará los detectores de Alice que tienen polarización ortogonal a la que está midiendo Bob. Ésto tendrá como consecuencia la pérdida de todos las detecciones del detector D0 cuando se da interferencia constructiva, es decir, cuando Alice y Bob eligen diferente polarización.

Cuando Alice y Bob eligen la misma polarización, las ecuaciones para las probabilidades de detección serán

$$\begin{aligned} P_{D1(2)}(n>0) = \frac{1}{2}xe^{-\eta_E|t\alpha|^2}(1 - e^{-\eta_1\sigma^2|r^* t\alpha|^2}) \\ + \frac{1}{2}(1-x)(1 - e^{-\eta_1\sigma^2|r^* t\alpha|^2}) \end{aligned} \tag{18}$$

$$\begin{aligned} P_{D0(2)}(n>0) = \frac{1}{2}xe^{-\eta_E|t\alpha|^2}(1 - e^{-\eta_0\sigma^2|r^* r\alpha|^2}) \\ + \frac{1}{2}(1-x)(1 - e^{-\eta_0\sigma^2|r^* r\alpha|^2}) \end{aligned} \tag{19}$$

Y cuando Alice y Bob eligen diferente polarización, la ecuación que da la probabilidad de detección de más de un fotón en el detector D0 será

$$\begin{aligned} P_{D0(2)}(n>0) = \frac{1}{2}(1-x)(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \\ + \frac{1}{2}x(1 - e^{-\eta_E|t\alpha|^2})(1 - e^{-\eta_0\sigma^2|\alpha|^2}) \end{aligned} \tag{20}$$

Las estadísticas para Bob vendrán dadas por la ecuación (17).

C. Protocolo con el detector D2 de Bob especializado y el detector D0 o D1 de Alice especializado

Al igual que el caso anterior, Eva tendrá que cegar siempre que mida y no detecte fotones. Sin embargo, al tener Alice un detector no especializado, el cegado hará que pierda todas las detecciones en ese detector. Las estadísticas para Bob vendrán dadas por la ecuación (17). Para este esquema del protocolo se dan dos casos, detallados a continuación.

1. *Detector D2 y D0 especializados*

En este caso cuando Alice y Bob eligen la misma polarización se tendrá que

$$P_{D1(3)}(n>0) = \frac{1}{2}(1-x)(1-e^{-\eta_1\sigma^2|r^*t\alpha|^2}) \tag{21}$$

$$\begin{aligned} P_{D0(3)}(n>0) = {} & \frac{1}{2}xe^{-\eta_E|t\alpha|^2}(1-e^{-\eta_0\sigma^2|r^*r\alpha|^2}) \\ & + \frac{1}{2}(1-x)(1-e^{-\eta_0\sigma^2|r^*r\alpha|^2}) \end{aligned} \tag{22}$$

Y cuando Alice y Bob eligen diferente polarización, la ecuación que da la probabilidad de detección de más de un fotón en el detector D0 será

$$\begin{aligned} P_{D0(3)}(n>0) = {} & \frac{1}{2}(1-x)(1-e^{-\eta_0\sigma^2|\alpha|^2}) \\ & + \frac{1}{2}x(1-e^{-\eta_E|t\alpha|^2})(1-e^{-\eta_0\sigma^2|\alpha|^2}) \end{aligned} \tag{23}$$

2. *Detector D2 y D1 especializados*

En este caso, cuando Alice y Bob eligen la misma polarización se tendrá que

$$\begin{aligned} P_{D1(4)}(n>0) = {} & \frac{1}{2}xe^{-\eta_E|t\alpha|^2}(1-e^{-\eta_1\sigma^2|r^*t\alpha|^2}) \\ & + \frac{1}{2}(1-x)(1-e^{-\eta_1\sigma^2|r^*t\alpha|^2}) \end{aligned} \tag{24}$$

$$P_{D0(4)}(n>0) = \frac{1}{2}(1-x)(1-e^{-\eta_0\sigma^2|r^*r\alpha|^2}) \tag{25}$$

Y cuando Alice y Bob eligen diferente polarización, la ecuación que da la probabilidad de detección de más de un fotón en el detector D0 será

$$\begin{aligned} P_{D0(4)}(n>0) = {} & \frac{1}{2}(1-x)(1-e^{-\eta_0\sigma^2|\alpha|^2}) \\ & + \frac{1}{2}x(1-e^{-\eta_E|t\alpha|^2})(1-e^{-\eta_0\sigma^2|\alpha|^2}) \end{aligned} \tag{26}$$

V. ESTUDIO NUMÉRICO

En el ataque expuesto el objetivo de Eva será la optimización de los parámetros x , y y z, de tal forma que las estadísticas de los detectores sean los más próximas posibles a las ecuaciones (7), (8), (9) y (11).

Si suponemos unos valores típicos para la transmisión del canal de $\sigma = 0.1$, un número medio de fotones introducido por Alice de $\langle n\rangle = 0.1$, unas eficiencias para los detectores de $\eta_0 = \eta_1 = \eta_2 = \eta_E = 0.1$. Para este estudio, Eva mejorá la transmisión del canal un 20%, es decir, $\sigma' = 1.2\sigma$ cuando no mide a la salida de Alice y tampoco ciega a Bob. Por último, se da un ejemplo con los resultados cuando Eva tiene un detector con una eficiencia del 90%. La cuarta columna de todas las tablas corresponde a las estadísticas del detector D0 cuando Alice y Bob eligen diferente polarización.

A. Estudio para $R = 0.5$ y $T = 0.5$ del divisor de haz de Alice

Tabla I: Protocolo con detectores simples.

x = 0.0432; y = 0.01; z = 0.45			
$P_{D1(1)}/P_{D1}$	$P_{D0(1)}/P_{D0}$	$P_{D2(1)}/P_{D2}$	$P_{D0(1)}/P_{D0}$
1.004	1.004	0.999	0.980

Tabla II: Protocolo con todos los detectores especializados.

x = 0.04; y = 0.01			
$P_{D1(2)}/P_{D1}$	$P_{D0(2)}/P_{D0}$	$P_{D2(2)}/P_{D2}$	$P_{D0(2)}/P_{D0}$
0.999	0.999	0.969	0.960

Tabla III: Protocolo con detectores D2 y D0/D1 especializados.

x = 0.04; y = 0.01			
$P_{D1(3)}/P_{D1} = P_{D0(4)}/P_{D0}$	$P_{D0(3)}/P_{D0} = P_{D1(4)}/P_{D1}$	$P_{D2(2)}/P_{D2}$	$P_{D0(3)}/P_{D0} = P_{D0(4)}/P_{D0}$
0.960	0.999	0.969	0.960

B. Estudio para $R = 0.4$ y $T = 0.6$ del divisor de haz de Alice

Tabla IV: Protocolo con detectores simples.

x = 0.0432; y = 0.01; z = 0.61			
$P_{D1(1)}/P_{D1}$	$P_{D0(1)}/P_{D0}$	$P_{D2(1)}/P_{D2}$	$P_{D0(1)}/P_{D0}$
0.990	1.012	0.999	0.973

Tabla V: Protocolo con todos los detectores especializados.

x = 0.04; y = 0.01			
$P_{D1(2)}/P_{D1}$	$P_{D0(2)}/P_{D0}$	$P_{D2(2)}/P_{D2}$	$P_{D0(2)}/P_{D0}$
0.999	0.999	0.969	0.960

Tabla VI: Protocolo con detectores D2 y D0/D1 especializados.

x = 0.04; y = 0.01			
$P_{D1(3)}/P_{D1} =$ $P_{D0(4)}/P_{D0}$	$P_{D0(3)}/P_{D0} =$ $P_{D1(4)}/P_{D1}$	$P_{D2(2)}/P_{D2}$	$P_{D0(3)}/P_{D0} =$ $P_{D0(4)}/P_{D0}$
0.960	0.999	0.969	0.960

C. Estudio para $R = 0.1$ y $T = 0.9$ del divisor de haz de Alice

Tabla VII: Protocolo con detectores simples.

x = 0.0432; y = 0.01; z = 0.986			
$P_{D1(1)}/P_{D1}$	$P_{D0(1)}/P_{D0}$	$P_{D2(1)}/P_{D2}$	$P_{D0(1)}/P_{D0}$
0.958	1.006	0.998	0.958

Tabla VIII: Protocolo con todos los detectores especializados.

x = 0.04; y = 0.01			
$P_{D1(2)}/P_{D1}$	$P_{D0(2)}/P_{D0}$	$P_{D2(2)}/P_{D2}$	$P_{D0(2)}/P_{D0}$
0.999	0.999	0.969	0.960

Tabla IX: Protocolo con detectores D2 y D0/D1 especializados.

x = 0.04; y = 0.01			
$P_{D1(3)}/P_{D1} =$ $P_{D0(4)}/P_{D0}$	$P_{D0(3)}/P_{D0} =$ $P_{D1(4)}/P_{D1}$	$P_{D2(2)}/P_{D2}$	$P_{D0(3)}/P_{D0} =$ $P_{D0(4)}/P_{D0}$
0.960	0.999	0.969	0.960

D. Estudio para $R = 0.1$ y $T = 0.9$ del divisor de haz de Alice y un 90% de eficiencia en el detector de Eva (η_E)

Tabla X: Protocolo con detectores simples.

x = 0.01; y = 0.727; z = 0.987; $\sigma' = \sigma$			
$P_{D1(1)}/P_{D1}$	$P_{D0(1)}/P_{D0}$	$P_{D2(1)}/P_{D2}$	$P_{D0(1)}/P_{D0}$
0.990	0.999	0.991	1.000

Tabla XI: Protocolo con todos los detectores especializados.

x = 0.01; y = 0.727; $\sigma' = \sigma$			
$P_{D1(2)}/P_{D1}$	$P_{D0(2)}/P_{D0}$	$P_{D2(2)}/P_{D2}$	$P_{D0(2)}/P_{D0}$
0.999	0.999	0.991	1.000

Tabla XII: Protocolo con detectores D2 y D0/D1 especializados.

x = 0.01; y = 0.727; $\sigma' = \sigma$			
$P_{D1(3)}/P_{D1} =$ $P_{D0(4)}/P_{D0}$	$P_{D0(3)}/P_{D0} =$ $P_{D1(4)}/P_{D1}$	$P_{D2(2)}/P_{D2}$	$P_{D0(3)}/P_{D0} =$ $P_{D0(4)}/P_{D0}$
0.990	0.999	0.991	1.000

Como se puede observar, en el estudio se consigue la igualdad en prácticamente todos los casos. La máxima diferencia entre los resultados esperados por Alice y los recibidos tras el ataque se da cuando esta relación es igual a 0.960, lo que supone una reducción en el número de cuentas por parte de Alice del 4% respecto a las esperadas. Estas pérdidas se podrían atribuir a una mala calibración del divisor de haz o pérdidas de fotones en el canal, pasando así desapercibidas en el régimen de clave infinita.

Por otro lado, como se muestra en las tablas del apartado D, cuando suponemos que Eva tiene un detector con una eficiencia del 90% se obtendría la igualdad en todos los esquemas experimentales sin la necesidad de mejorar la transmisión del canal.

VI. CONCLUSIÓN, CONTRAMEDIDAS Y LÍNEAS FUTURAS

En este artículo se ha demostrado que los protocolos contrafácticos también son inseguros frente a ataques de cegado, siendo el único requerimiento que Eva sea capaz de conocer el estado de Bob y reducir moderadamente las pérdidas del canal.

Este ataque se puede evitar con contramedidas que han sido demostradas efectivas para otros protocolos. Por ejemplo, usando detectores tipo "perro guardián" o *watchdog* [25], que consiste en un divisor de haz muy transmitivo y donde la parte reflejada se redirige a un detector clásico para saber si se esta introduciendo una potencia por encima del nivel cuántico. De esta forma, se afecta de forma muy leve a la eficiencia de intercambio de clave y se puede detener el protocolo si se detectara algo en el último detector.

Otro estudio necesario sería analizar las posibles limitaciones o ventajas del ataque propuesto en el régimen de clave finita, debido a que en ocasiones cambian las condiciones de seguridad [17, 26].

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. USA: Cambridge University Press, 2011.

[2] S. Wiesner, "Conjugate coding," *SIGACT News*, vol. 15, no. 1, p. 78–88, Jan. 1983.

[3] C. H. Bennett and G. Brassard, "Quantum public key distribution system," *IBM Technical Disclosure Bulletin*, vol. 28, no. 7, pp. 3153–3163, 1985.

[4] C. H. Bennett, G. Brassard, and N. D. Mermin, "Quantum cryptography without Bell's theorem," *Phys. Rev. Lett.*, vol. 68, pp. 557–559, Feb 1992.

[5] A. K. Ekert, "Quantum cryptography based on bell's theorem," *Phys. Rev. Lett.*, vol. 67, pp. 661–663, Aug 1991.

[6] A. C. Elitzur and L. Vaidman, "Quantum mechanical interaction-free measurements," *Foundations of Physics*, vol. 23, no. 7, pp. 987–997, 1993.

[7] M. Bhatt, A. Aneja, S. Tripathi, and G. Nagar, "Classical cryptography v/s quantum cryptography a comparative study," *International Journal of Electronics and Computer Science Engineering*, vol. 1, 01 2012.

[8] Y.-Y. Fei, X.-D. Meng, M. Gao, H. Wang, and Z. Ma, "Quantum man-in-the-middle attack on the calibration process of quantum key distribution," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[9] S. E. Vinay and P. Kok, "Extended analysis of the trojan-horse attack in quantum key distribution," *Physical Review A*, vol. 97, no. 4, p. 042335, 2018.

[10] T.-G. Noh *et al.*, "Counterfactual quantum cryptography," *Physical review letters*, vol. 103, no. 23, p. 230501, 2009.

[11] S. Zhang, J. Wang, and C.-J. Tang, "Security proof of counterfactual quantum cryptography against general intercept-resend attacks and its vulnerability," *Chinese Physics B*, vol. 21, no. 6, p. 060303, jun 2012.

[12] Z. Yin, L. Hongwei, C. Wei, H.-W. Li, and G. Guangcan, "Security of counterfactual quantum cryptography," *Physical Review. A*, vol. 82, 10 2010.

[13] M. Ren, G. Wu, E. Wu, and H. Zeng, "Experimental demonstration of counterfactual quantum key distribution," *Laser Physics*, vol. 21, no. 4, pp. 755–760, 2011.

[14] G. Brida, A. Cavanna, I. Degiovanni, M. Genovese, and P. Traina, "Experimental realization of counterfactual quantum cryptography," *Laser Physics Letters*, vol. 9, no. 3, pp. 247–252, jan 2012.

[15] Y. Liu, L. Ju, X.-L. Liang, S.-B. Tang, G.-L. S. Tu, L. Zhou, C.-Z. Peng, K. Chen, T.-Y. Chen, Z.-B. Chen, and J.-W. Pan, "Experimental demonstration of counterfactual quantum communication," *Phys. Rev. Lett.*, vol. 109, p. 030501, Jul 2012.

[16] Z. Q. Yin, H. W. Li, Y. Yao, C. M. Zhang, S. Wang, W. Chen, G. C. Guo, and Z. F. Han, "Counterfactual quantum cryptography based on weak coherent states," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 86, no. 2, pp. 1–8, 2012.

[17] Z.-H. Li, L. Wang, J. Xu, Y. Yang, M. Al-Amri, and M. S. Zubairy, "Counterfactual trojan horse attack," *Phys. Rev. A*, vol. 101, p. 022336, Feb 2020.

[18] X. Yang, K. Wei, H. Ma, S. Sun, Y. Du, and L. Wu, "Trojan horse attacks on counterfactual quantum key distribution," *Physics Letters A*, vol. 380, no. 18-19, pp. 1589–1592, 2016.

[19] V. Makarov, "Controlling passively quenched single photon detectors by bright light," *New Journal of Physics*, vol. 11, no. 6, p. 065003, 2009.

[20] 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 Journal of Physics*, vol. 13, no. 1, p. 013043, 2011.

[21] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, "Invited review article: Single-photon sources and detectors," *Review of Scientific Instruments*, vol. 82, no. 7, p. 071101, 2011.

[22] S. J and M. Jamro, *Optical Fiber Communications—Principles and Practice*, 01 2009.

[23] Y. Kang, H. Lu, Y.-H. Lo, D. Bethune, and W. Risk, "Dark count probability and quantum efficiency of avalanche photodiodes for single-photon detection," *Applied Physics Letters*, vol. 83, pp. 2955–2957, 10 2003.

[24] R. Loudon, *The Quantum Theory of Light*. Oxford: Clarendon Press, 1973.

[25] V. Chistiakov, A. Huang, V. Egorov, and V. Makarov, "Controlling single-photon detector id210 with bright light," *Opt. Express*, vol. 27, no. 22, pp. 32 253–32 262, Oct ts , doi = 10.1364/OE.27.032253, abstract = We experimentally demonstrate that a single-photon detector ID210 commercially available from ID Quantique is vulnerable to blinding and can be fully controlled by bright illumination. In quantum key distribution, this vulnerability can be exploited by an eavesdropper to perform a faked-state attack giving her full knowledge of the key without being noticed. We consider the attack on standard BB84 protocol and a subcarrier-wave scheme and outline a possible countermeasure.,.

[26] W. Wang, K. Tamaki, and M. Curty, "Finite-key security analysis for quantum key distribution with leaky sources," *New Journal of Physics*, vol. 20, no. 8, p. 083027, 2018.