Простая квантовая схема для атаки, демонстрирующей уязвимость квантовой криптографии с фазово-временным кодированием

Д.А.Кронберг

Математический институт им.В.А.Стеклова РАН

7 июня 2023 г.

План

- Квантовая криптография, достоинства и недостатки
- Протокол с фазово-временным кодированием и атака на него
- Полемика с разработчиками и квантовая схема
- Почему появилась уязвимость и предложения по подходу к обоснованию стойкости протоколов КРК

Квантовая криптография: основы

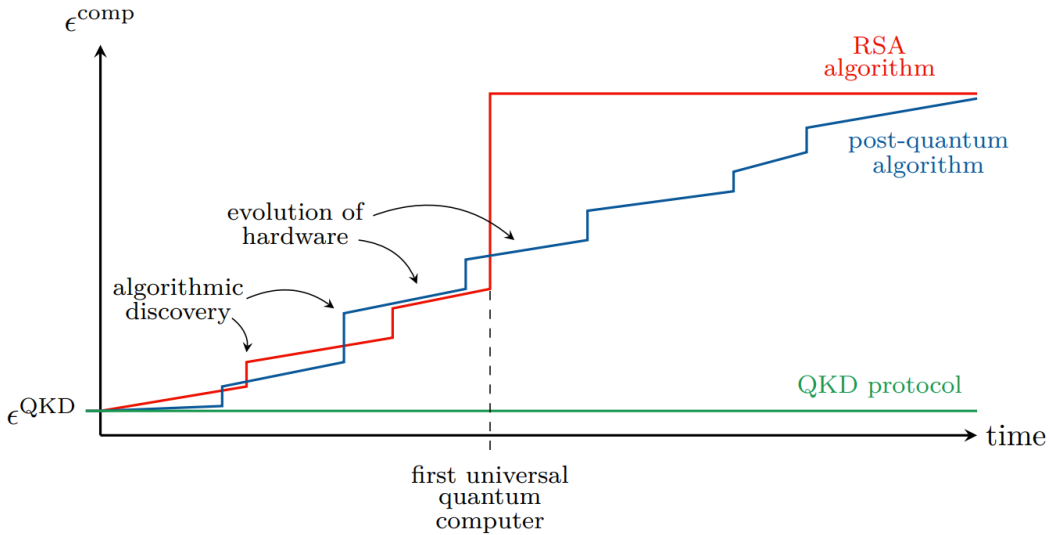
- Основная задача — распределение секретного ключа между удаленными пользователями (Алисой и Бобом) без вычислительных и технологических предположений о перехватчике (Еве)
- Ключевое преимущество квантовой криптографии — возможность математического доказательства стойкости. Это главный теоретический элемент протокола квантового распределения ключей
- Стойкость в квантовой криптографии основана на том, что любое измерение — это взаимодействие с системой, и невозможно измерить передаваемые состояния, не внеся в них помехи
- Доказательство стойкости — это формула для длины секретного ключа в зависимости от наблюдаемых параметров и утверждение, что если ключ после коррекции ошибок сжимать в соответствии с этой формулой, то он будет обладать нужными характеристиками (близок к идеальному ключу)
- Протокол BB84 был предложен в 1984 году, первое признанное доказательство стойкости появилось лишь в 2000 году

Квантовая криптография: чувствительность к аппаратуре

- Доказательство стойкости в квантовой криптографии возможно, когда зафиксирована математическая модель оборудования участников
- Доступное оборудование не обязательно в точности соответствует математическому описанию, и это может создать ряд уязвимостей
- Также противник может целенаправленно наносить урон оборудованию, чтобы заставить его работать не так, как планируют легитимные пользователи
- Недостаток квантовой криптографии — чувствительность к аппаратуре
- Классические криптографические протоколы также чувствительны к аппаратуре
- Возможен подход к квантовой криптографии, в котором стойкость не зависит от детекторов как самой уязвимой части аппаратуры (Measurement device independent, MDI QKD), а также без зависимости от аппаратуры вовсе (DI QKD). В этом случае протоколы сами по себе гарантируют, что оборудование работает должным образом
- Можно выделить атаки на состояния (против которых возможно доказательство стойкости) и атаки на оборудование, которые в стандартном подходе решаются сертификацией оборудования

Квантовая криптография: вечная стойкость

Вечная стойкость распределяемых ключей (everlasting security): если противник нашел новые способы использовать несовершенство оборудования, это не способно уменьшить стойкость уже распределенных ключей



R.Renner, R.Wolf. Quantum advantage in cryptography. AIAA Journal, 61(5), 1895-1910, (2023)

Квантовая криптография: преимущества и недостатки

Преимущества

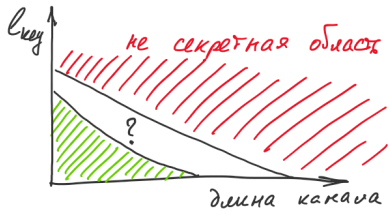
- Возможность математического доказательства стойкости
- Вечная стойкость распределяемых ключей

Недостатки

- Хрупкость и недостаточная изученность оборудования
- Сложность доказательства стойкости и возможность ошибок в нем

Верхние и нижние оценки криптографической стойкости

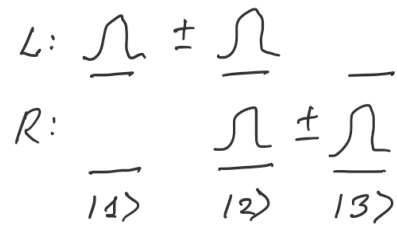
- Практически доступные линии связи имеют затухание, и перехватчик может использовать его для построения более эффективных атак
- С учетом затухания в линии связи доказательство стойкости представляет собой нижнюю оценку скорости генерации секретного ключа
- Явно построенная атака — верхняя оценка скорости генерации ключа
  - Возможен зазор между верхними и нижними оценками, который говорит о возможности улучшения какой-либо из оценок
  - Построить явную атаку — более сложная задача, чем найти пробелы в доказательстве стойкости
  - Корректность работы квантового компьютера легко проверить, с квантовой криптографией сложнее: проверка доказательства стойкости и построение атак может быть сложной задачей



Протокол с фазово-временным кодированием

Протокол с фазово-временным кодированием использует три временных окна $\{|1\rangle, |2\rangle, |3\rangle\}$, что при отправке однофотонных импульсов отвечает трёхмерному пространству

$$|0_L\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle), \qquad |0_R\rangle = \frac{1}{\sqrt{2}}(|2\rangle + |3\rangle),$$
$$|1_L\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle), \qquad |1_R\rangle = \frac{1}{\sqrt{2}}(|2\rangle - |3\rangle). \tag{1}$$



- На приемной стороне проводится измерение в левом или правом базисе, эти измерения описываются соответствующими наблюдаемыми (разложениями единицы)

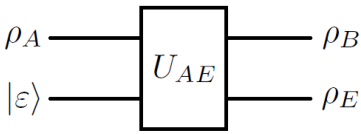
$$M_L = \{|0_L\rangle\langle 0_L|, \quad |1_L\rangle\langle 1_L|, \quad |3\rangle\langle 3|\},$$
$$M_R = \{|1\rangle\langle 1|, \quad |0_R\rangle\langle 0_R|, \quad |1_R\rangle\langle 1_R|\}. \tag{2}$$

- Неформально стойкость протокола следует из того, что если перехватчик проведет такое же измерение, но не угадает базис, он будет отправлять контрольные отсчеты: $|3\rangle$ в левом базисе и $|1\rangle$ в правом

S. N. Molotkov, “Tight finite-key analysis for two-parametric quantum key distribution”, Laser Phys. Lett., 16:3 (2019), 035203.

С. Н. Молотков, «О стойкости систем квантовой криптографии с фазово-временным кодированием к атакам активного зондирования», ЖЭТФ, 158:6(12) (2020), 1011–1031.

Протокол с фазово-временным кодированием



Обоснование стойкости протокола строится на рассмотрении унитарного преобразования перехватчика над исходным состонием и вспомогательной системой (анциллой), с представлением Стайнспринга:

$$\begin{aligned}\Phi_{A\to B}[\rho] &= \mathrm{Tr}_E\left[U_{AE}(\rho_A\otimes|\varepsilon\rangle\langle\varepsilon|_E)U_{AE}^\dagger\right],\\ \Phi_{A\to E}[\rho] &= \mathrm{Tr}_B\left[U_{AE}(\rho_A\otimes|\varepsilon\rangle\langle\varepsilon|_E)U_{AE}^\dagger\right].\end{aligned} \tag{3}$$

- Проведена оценка информации, извлекаемой из состояний перехватчика ρ_E, в зависимости от наблюдаемых при измерении ρ_B параметров $\{Q,\zeta\}$, что ведет к формуле длины секретного ключа

$$l_{\mathrm{key}}^{SP}(\zeta,Q) = 1-\mathrm{h}_2(Q)-\mathrm{h}_2\left(\frac{\zeta}{1-\zeta}\right). \tag{4}$$

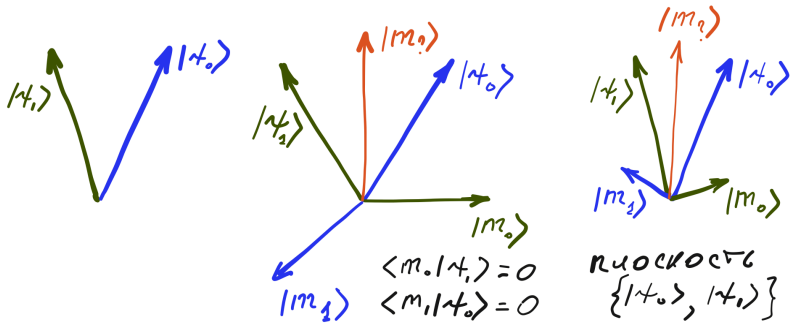
- Она проистекает из формулы Деветака-Винтера: $l_{\mathrm{key}} = I(A:B)-I(A:E)$
- Информация противника оценивается сверху через $\mathrm{h}_2\left(\frac{\zeta}{1-\zeta}\right)$, в то время как информация легитимных пользователей

$$I(A:B) = 1-\mathrm{h}_2(Q)$$

соответствует бинарному симметричному каналу связи с ошибкой Q.

Атака на протокол: увеличение размерности

- Ошибка доказательства стойкости в том, что не учтены атаки с затуханием, которые математически соответствуют более высокой размерности пространства выхода.
- Это можно увидеть на примере безошибочного различения состояний (USD)



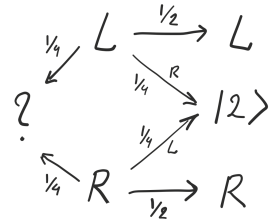
- При действии квантового канала $\rho \to \Phi[\rho]$ величина Холево состояний, отвечающая максимально доступной классической информации, монотонна
- При постселективных преобразованиях монотонность не обязательно выполняется, и информация может быть выше границы Холево

N.R.Kenbaev, D.A.Kronberg, Quantum postselective measurements: Sufficient condition for overcoming the Holevo bound and the role of max-relative entropy. Physical Review A, 105(1), 012609 (2022)

Атака на протокол: различение базисов

Рассмотрим квантовую наблюдаемую (разложение единицы), которая различает левые и правые базисы протокола:

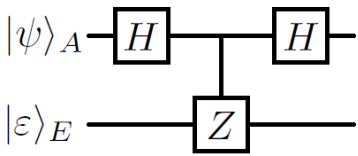
$$M_L = \frac{1}{2}\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix},\ M_R = \frac{1}{2}\begin{pmatrix}0&0&0\\0&1&0\\0&0&1\end{pmatrix},\ M_? = \frac{1}{2}\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}$$



$$p(L|L) = \mathrm{Tr} M_L \rho_L = \frac{1}{2}, \quad p(R|L) = \frac{1}{4}, \quad p(?|L) = \frac{1}{4}.$$

- В случае успеха и верного результата соответствующий квантовый канал $\rho_L \to \frac{\sqrt{M_L}\rho_L\sqrt{M_L}}{\mathrm{Tr} M_L \rho_L}$ оставляет состояния левого и правого базиса без изменения; при ошибочном результате состояния переходят в $|2\rangle$; с вероятностью $1/4$ возможна неудача (результат «?»)
- В итоге, если возможна блокировка $1/4$ части посылок, перехватчик может различить базисы с вероятностью правильного ответа $2/3$, что значительно выше вероятности $1/2$ при простом угадывании
- Существует наблюдаемая, которая различает левый и правый базисы без ошибки (но с возможностью неудачи), однако такая наблюдаемая портит состояния внутри базиса

Атака на протокол: различение состояний



Пусть Ева применяет простую квантовую схему из условного фазового сдвига и преобразования Адамара. Анцилла Евы при параметре γ находится в состоянии

$$|\varepsilon\rangle_E = \cos\gamma|1\rangle_E + \sin\gamma|2\rangle_E, \quad \gamma \in [0, \frac{\pi}{4}] \tag{5}$$

- Схема действует на состояния следующим образом:

$$\begin{aligned} |0_L\rangle_A \otimes |\varepsilon\rangle_E &\to |0_L\rangle_A \otimes (\cos\gamma|1\rangle_E + \sin\gamma|2\rangle_E), \\ |1_L\rangle_A \otimes |\varepsilon\rangle_E &\to |1_L\rangle_A \otimes (\cos\gamma|1\rangle_E - \sin\gamma|2\rangle_E), \\ |2\rangle_A \otimes |\varepsilon\rangle_E &\to \sin\gamma|1\rangle_A \otimes |2\rangle_E + \cos\gamma|2\rangle_A \otimes |1\rangle_E, \end{aligned} \tag{6}$$

- Это копирование состояний в память Евы, где они становятся частично различимыми. При этом состояние $|2\rangle$ частично переходит в левое временное окно $|1\rangle$, что приводит к контрольным временным отсчетам (ненулевой параметр ζ).

Скорость генерации ключа

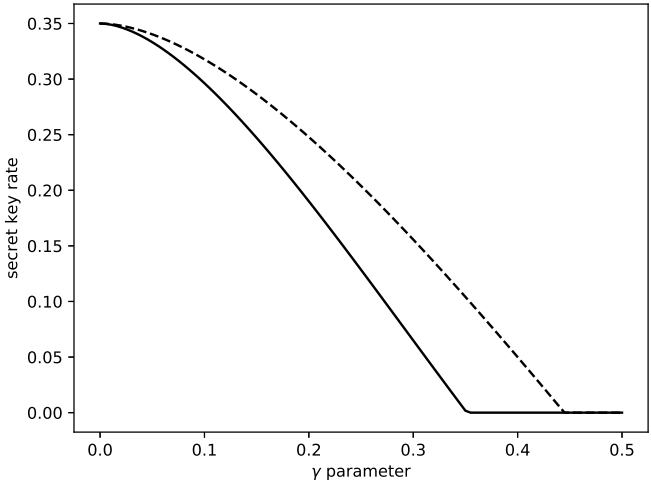
- При неверном определении базиса построенная атака приводит к отсчету в контрольном временном окне $|1\rangle$ с вероятностью $\sin^2\gamma$, также оно может привести к информационному отсчету с вероятностью $\cos^2\gamma$, и в этом случае битовая ошибка будет возникать в половине посылок
- Вероятность неверного определения базиса Евой равна $\frac{1}{3}$, что ведет согласно формуле разработчиков (4) к

$$l_{\text{key}}(\gamma) = 1 - \mathrm{h}_2\left(\frac{1}{6}\right) - \mathrm{h}_2\left(\frac{\sin^2\gamma}{3-\sin^2\gamma}\right). \tag{7}$$

- С точки зрения Евы, с вероятностью $\frac{2}{3}$ базис был угадан верно, и информация Евы дается величиной Холево от чистых состояний $\cos\gamma|1\rangle \pm \sin\gamma|2\rangle$ (она равна $h\left(\frac{1-\cos 2\gamma}{2}\right) = h(\sin^2\gamma)$), откуда имеем для разности информации легитимных пользователей и перехватчика

$$l_{\text{key}}^{\text{attack}}(\gamma) = 1 - \mathrm{h}_2\left(\frac{1}{6}\right) - \frac{2}{3}\mathrm{h}_2\left(\sin^2\gamma\right). \tag{8}$$

Скорость генерации ключа



Скорость генерации секретного ключа в зависимости от параметра атаки γ согласно построенной атаке (сплошная линия), а также согласно доказательству стойкости протокола (штриховая линия)

Область на графике, соответствующая $\gamma \in [0.357, 0.452]$, — это значения γ, при которых противнику известен весь ключ. В то же время согласно формуле скорости генерации ключа, ключ ненулевой, то есть легитимные пользователи считают, что сгенерировали полностью секретный ключ.

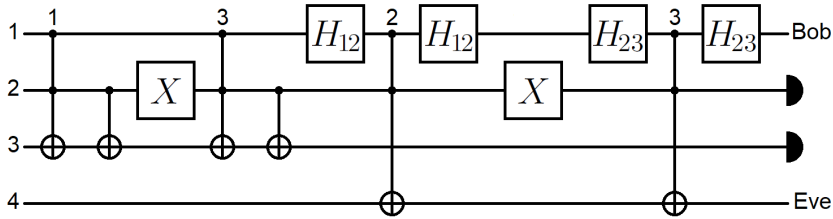
Построение атаки: выводы

- Уязвимость связана с ошибкой в доказательстве стойкости, которая заключается в рассмотрении в качестве самой общей стратегии перехватчика унитарного преобразования, действующего на паре трехмерных пространств Алисы (натянутого на векторы $\{|1\rangle, |2\rangle, |3\rangle\}$) и Евы
- При наличии затухания в линии связи размерность пространства на приемной стороне увеличивается на единицу, и дополнительное измерение соответствует потере сигнала
- Был построен пример конкретного унитарного преобразования, которое более эффективно для перехватчика, чем унитарное преобразование на паре трехмерных пространств, рассмотренное в работах по доказательству стойкости
- Эта же ошибка встречается в ряде других работ с обоснованием стойкости протоколов квантовой криптографии, в частности, протокола на геометрически однородных состояниях и более ранних версий протокола с фазово-временным кодированием
- При использовании ослабленного лазерного излучения вместо однофотонных посылок протокол теряет стойкость уже начиная со сколь угодно малой длины линии связи, виной тому неверное использование GLLP-подхода оценки скорости генерации ключа, разработанного для протокола BB84

Реакция разработчиков протокола

- Статья про уязвимость протокола была отклонена на основании рецензии, где говорится о высокой сложности данной атаки: «Любую умозрительную атаку непосредственно на квантовые состояния нужно всерьез принимать в расчет только в том случае, если такая атака физически реализуема. Предлагаемая атака (физическая реализация операторов A_L, A_R, $A_?$) требует неразрушающего измерения состояний $|0_L, 1_L\rangle\langle 0_L, 1_L|$ в левом базисе (оператор A_L) и для состояний $|0_R, 1_R\rangle\langle 0_R, 1_R|$ в правом базисе (оператор A_R). $<$ ... $>$ такое неразрушающее измерение практически нереально.»
- Исходные статьи при этом не содержали ссылок на сложность тех или иных атак: утверждалась стойкость против *самой общей* атаки: «Наиболее общая атака подслушивателя на однофотонные состояния сводится к унитарной атаке, которая строится явно. $<$...$>$ Наиболее общей атакой на однофотонные состояния является унитарная атака (9)–(11)»
- Если же говорить о сложности атак, то аргумент рецензии был таким: «атака похожа на PNS-атаку. Мы в другой статье приводили пример PNS, это была сложная реализация, значит, PNS-атака сложна в реализации. Значит, ни PNS-атаку, ни эту атаку не надо рассматривать. Значит, работу публиковать не следует.»
- Однако, все утверждения в пункте выше неверные, и сходство атаки с PNS не означает сложности в её реализации

Квантовая схема предложенной атаки



- Предложенная схема требует квантового компьютера с тремя дополнительными кубитами, и шести нетривиальных квантовых операций
- Проектор, про который разработчиками утверждалась сложность, реализуется с помощью модификации CNOT («контролируемого НЕ»): если система в состоянии $|3\rangle$, анцилла должна измениться (для A_L), иначе это тождественное преобразование

$$A_L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_? = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad CNOT_3 : \begin{array}{l} |3\rangle|0\rangle \to |3\rangle|1\rangle, \\ |3\rangle|1\rangle \to |3\rangle|0\rangle. \end{array}$$

- Другие операции отвечают вероятностному применению данного преобразования, поскольку перед проектором стоит коэффициент $\frac{1}{2}$
- Нижние оценки сложности операции от разработчиков неверны

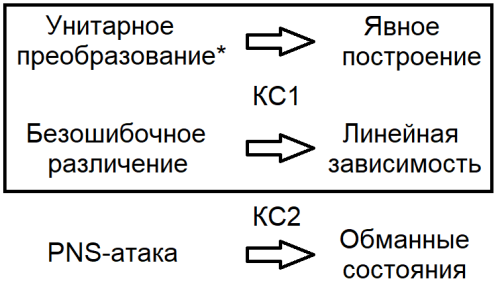
Подход к обоснованию стойкости квантовой криптографии

Классическая криптография

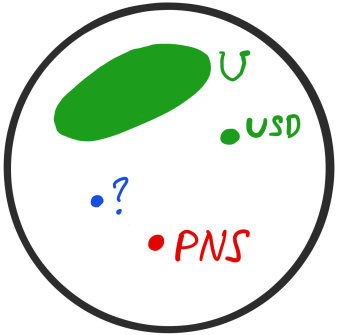
Подход к обоснованию стойкости квантовой криптографии

Классическая криптография

Квантовая криптография



Подход к обоснованию стойкости квантовой криптографии



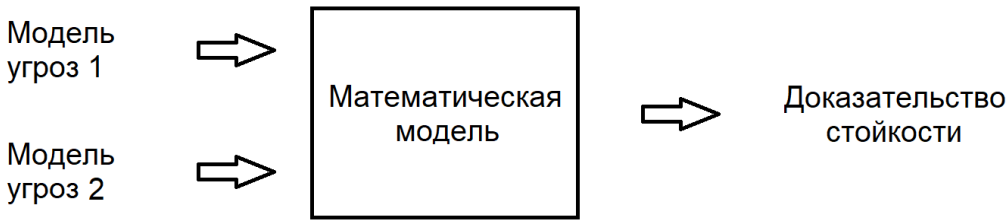
Подход с перечислением возможных атак перехватчика неудачен, так как нельзя перечислить всё континуальное множество возможных атак; любая категоризация будет условной

- Не атака унитарным преобразованием*: другое пространство
- Не атака безошибочным различением состояний (USD): есть ошибка, и сохраняется «квантовость» состояний
- Не PNS-атака: работает в однофотонном случае, нет операции подсчета количества фотонов

- Критика атаки «задним числом» возможна, но не должна сбивать с толку и оправдывать ошибки разработчиков, иначе выходит «мы доказали стойкость против всех атак[1]»
- От разработчиков поступил новый комментарий на статью, где говорится, что атака все равно похожа на PNS-атаку, поэтому не реализуема, теперь говорится о сходстве операции в многофотонном случае.

[1]Кроме тех, которые просмотрели

Предлагаемый подход к обоснованию стойкости



- Перечисление атак *на оборудование* уместно при формулировании математической модели, для которой остается актуальным требование строгого доказательства стойкости
- Такой подход сделает невозможным нахождение уязвимостей в смысле атак *на квантовые состояния*, которое произошло в данном случае
- Подход использует одно из основных преимуществ квантовой криптографии: возможность строгого доказательства стойкости против *всех* атак на состояния

История отношений с разработчиками

Исследование	Статьи	Опровержения
Протокол КРК на чистых геометрически однородных когерентных состояниях	Письма в ЖЭТФ, 95(6), 361–366 (2012) Письма в ЖЭТФ, 101 (8), 637–643 (2015) Письма в ЖЭТФ, 105 (9), 570–576 (2017)	LJM 41, 2332-2337 (2020) КЭ 51(10), 928-937 (2021)
Критика метода обманных состояний	ЖЭТФ, 155(4), 636-644 (2019) ЖЭТФ 156(8), 379-380 (2019)	УФН 191(1), 93-109 (2021) ЖЭТФ 161(5), 627-630 (2022)
Протокол КРК на геометрически однородных когерентных состояниях с фазовой рандомизацией	PRA 100(4), 042329 (2019)	PRA 104(2), 026401 (2021)
Протокол КРК с фазово-временным кодированием	LPL 16(3), 035203 (2019) ЖЭТФ 158(6), 1011-1031 (2020)	ТМФ 214(1), 140-152 (2023)

Качество рецензирования статей

ЖЭТФ, 2020, том 158, вып. 6 (12), стр. 1011–1031

О СТОЙКОСТИ СИСТЕМ КВАНТОВОЙ КРИПТОГРАФИИ С ФАЗОВО-ВРЕМЕННЫ́М КОДИРОВАНИЕМ К АТАКАМ АКТИВНОГО ЗОНДИРОВАНИЯ

С. Н. Молотков*

Институт физики твердого тела Российской академии наук
142432, Черноголовка, Московская обл., Россия

Академия криптографии Российской Федерации
121552, Москва, Россия

Центр квантовых технологий, Московский государственный университет им. М. В. Ломоносова
119899, Москва, Россия

Поступила в редакцию 11 мая 2020 г.,
после переработки 11 мая 2020 г.
Принята к публикации 3 июля 2020 г.

Статья длиной в 21 страницу была отрецензирована и переработана за 1 день

Выводы

- Проверка корректности работы систем квантового распределения ключей — нетривиальная задача
- Доказательство стойкости — ключевой теоретический элемент, и ошибки в этой части делают всю дорогостоящую разработку бесполезной: возможна явная атака, которая дает противнику информацию о ключе
- Подход разработчиков к доказательству стойкости через построение оптимального унитарного преобразования некорректен из-за ошибок в размерности пространства
- Подход к сертификации через «модель угроз нарушителя $\rightarrow$ обоснование стойкости» уместен для формулирования математической модели с учетом возможных атак на оборудование, но не для доказательства стойкости против атак на квантовые состояния

Работа выполнена при поддержке гранта РНФ (проект 19-11-00086)

Спасибо за внимание!