Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier

An Analysis of the Completion Time of the BB84 Protocol

SOUNAK KAR[1] AND JEAN-YVES LE BOUDEC[2], (Fellow, IEEE)
[1]QuTech, TU Delft, Netherlands (email: s.kar-1@tudelft.nl)
[2]EPFL, Switzerland (email: jean-yves.leboudec@epfl.ch)

ABSTRACT In the BB84 quantum key distribution (QKD) protocol, the sender and receiver estimate the quantum bit error rate (QBER) over a certain fraction of the transmitted qubits to detect eavesdropping, while using the remaining qubits to generate a private key. Here, we consider an entanglement-based implementation of the BB84 protocol in which the sender and receiver are connected through a single repeater node. Under the present hardware infrastructure, elementary link generation, and consequently end-to-end link generation, is a failure-prone process. Therefore, performance analyses of such protocols typically focus on the completion time, that is, the time until successful completion, rather than the duration of a single attempt. Due to decoherence, the success of an attempt generally depends on the duration of the individual phases of that attempt, since quantum states must wait in memory while the success or failure of a generation phase is communicated to the relevant parties. In this work, we perform a performance analysis of the completion time of the BB84 protocol, assuming a simplified noise model and certain distributional forms for the generation and communication phases of end-to-end link generation over a single quantum repeater. We provide a method for computing the moment generating function (MGF) of the completion time and subsequently derive an estimate of its distribution, as well as a bound on the tail probability. This result allows us to characterize the tail behavior of the completion time in terms of the parameters governing the elementary phases of entanglement generation, without requiring repeated executions of the protocol. We also provide an efficient simulation scheme to generate the protocol completion time, which relies on expressing the completion time in terms of the end-to-end link generation time until success. We numerically compare our approach with a full-scale simulation and observe good agreement between them.

INDEX TERMS Quantum key distribution, simulation.

I. INTRODUCTION

Quantum networks have the potential to support applications beyond the capabilities of classical communication networks. A key requirement for realizing such networks is the reliable distribution of entanglement between spatially separated nodes. In optical fiber, photon transmission decreases exponentially with distance as a result of attenuation, which severely constrains long-distance entanglement generation based on direct transmission [1]. At the same time, the no-cloning theorem [2] precludes producing identical copies of an unknown quantum state and therefore rules out the signal-amplification methods routinely employed in classical communication. Quantum repeaters [3]–[6] address these limitations by introducing intermediate nodes through which entanglement can be established across substantially longer distances.

In a repeater-assisted network, entanglement is first established over elementary links connecting the end nodes to the repeater, which are subsequently connected through entanglement swapping to generate an end-to-end link between the end nodes. Under the present hardware infrastructure, elementary-link generation, and consequently end-to-end link generation, are failure-prone processes. Thus, multiple attempts may be required before an end-to-end link is successfully generated. Furthermore, the duration of the individual phases of end-to-end link generation determines the degree of decoherence of the involved qubits, since a successfully generated quantum state may need to remain in memory while the corresponding state on the other elementary link is being generated and while the outcomes of the relevant generation phases are communicated. These repeated and duration-dependent attempts make the completion time, i.e., the time until successful completion, a natural performance metric for such systems. The authors in [7]–

[9] have analyzed the completion time of certain quantum communication modules in a discrete-time setting.

In this work, we analyze the completion time of the entanglement-based implementation of the BB84 Quantum Key Distribution (QKD) protocol [10] in a continuous-time setting. Specifically, we consider the communicating parties, Alice and Bob, to be connected through a single first-generation quantum repeater [5], i.e., a repeater that does not rely on quantum error correction. This represents the smallest nontrivial repeater-assisted network instance and captures the interaction between probabilistic elementary-link generation, storage of entanglement in quantum memories, decoherence, and probabilistic entanglement swapping.

An essential step of the BB84 protocol is the estimation of the quantum bit error rate (QBER), which acts as its principal security test [11]. Errors contributing to the QBER may arise from physical imperfections and decoherence or from adversarial intervention, and the protocol aborts when the estimated QBER exceeds a prescribed threshold. In this work, we do not focus on the security aspects of BB84 in the considered setting but rather on characterizing the completion time of the protocol under the specified physical model. To that end, we consider all steps of BB84 and analyze the completion time in terms of the durations of these steps.

Our first contribution is to provide a formula for the moment generating function (MGF) of the completion time where the relevant generation or communication phases of the BB84 steps last for a non-negligible random time. The MGF can then be used to obtain the Laplace transform of the cumulative distribution function (CDF) of the completion time, which can be subsequently inverted to derive the CDF. We also derive a delay bound on the tail probability using Chernoff's method. This result helps us understand the behavior of the completion time with respect to the parameters of the elementary phases of the protocol without having to go through the tedious exercise of running the protocol multiple times. As a specific application, our analysis can be used to compute the memory lifetime requirement on the involved quantum memories so that the completion time is below a given threshold with a certain probability. More broadly, recent works [12]–[15] have pursued an application-driven approach to quantum-network design by translating target performance metrics into requirements on the underlying quantum-network hardware. While these works formulate hardware requirements using average-case metrics such as entangling rate, teleportation rate, expected end-to-end link fidelity, or expected teleportation fidelity, our analysis can be further used to derive requirements for running the BB84 protocol with probabilistic latency guarantees.

Our second contribution is a fast simulation scheme to generate the completion time of the protocol. There has been a considerable amount of work [16]–[20] on the simulation of the BB84 protocol with or without attack, where the focus has been on obtaining the key rate or the eavesdropping detection probability. All of these works simulate the individual phases of the protocol to arrive at their end objective. In this work, we demonstrate that this is not necessary if we only want to obtain the completion time, and we assume that the durations of the relevant phases follow certain distributions. Under such assumptions, we first show that the total time required to generate an end-to-end link can be well-approximated by a Coxian phase-type distribution, which can be efficiently aggregated to simulate the completion time of the BB84 protocol.

The rest of the paper is structured as follows: we describe the setup and main results on the BB84 protocol in Sect. II, whereas the methods and proofs are given in Sect. III. We next describe the simulation scheme in Sect. IV, while a numerical evaluation of the derived results is given in Sect. V.

II. ASSUMPTIONS AND MAIN RESULTS ON THE COMPLETION TIME

A. SETUP AND ASSUMPTIONS

As mentioned, we consider the entanglement-based version of the BB84 protocol. The protocol proceeds by sharing an entangled pair between the sender, Alice, and the receiver, Bob. In our setup, Alice and Bob are located at the end nodes connected by a repeater. Following the standard terminology, an entangled pair shared between an end node and the repeater is called an *elementary link*, whereas an entangled pair shared between the end nodes is called an *end-to-end link*. We first recall the steps for the BB84 protocol in this setting.

State preparation and distribution At each step, an end-to-end link is generated, whose state description is provided in (3).

Measurement We assume a symmetric implementation of the protocol. For each end-to-end link, Alice selects the measurement basis, either $\mathbb{Z}$ or $\mathbb{X}$, with equal probability for her half of the pair. Independently, Bob applies the same basis-selection rule to his half. Alice and Bob store their outcomes in their respective classical registers.

Sifting After n links are generated and measured, Alice and Bob broadcast their basis choices and keep only the outcomes for which their bases match.

QBER estimation To check the channel quality, Alice and Bob estimate the (QBER), for which they sample α fraction of the sifted outcomes from the previous step without replacement. The estimated QBER is defined as the proportion of disagreements between Bob's and Alice's measurements out of the sampled outcomes. The protocol aborts if the estimated QBER exceeds a predetermined threshold, i.e., if it is greater than $1-\beta$, and the parties begin a fresh BB84 attempt. Otherwise, they proceed to the next step.

Classical post processing Alice and Bob perform error correction and privacy amplification to obtain a key out of the remaining $(1-\alpha)$ fraction of the sifted outcomes.

We now provide the underlying assumptions impacting the end-to-end link generation process.

A1 **Single qubit memory**: Alice and Bob each have a single qubit memory, which implies that they must measure their half of the entangled pair after every successful end-to-end link generation event rather than measuring all pairs together at the end.

A2 **Single repeater and equidistant parties**: Alice and Bob are just far enough to be connected via a single quantum repeater. Further, they are equidistant from the repeater which allows the communication time for classical and quantum signals between the repeater and an end node (i.e., Alice or Bob) to be identically distributed.

A3 **Identical memories**: The qubits at the end nodes and the repeater node are assumed to have identical physical characteristics. This simplifies the noise modelling process, which we describe next.

A4 **Depolarizing noise and link description in Werner form**: The ideal target state for an elementary or end-to-end link is a maximally entangled Bell state. Since one maximally entangled Bell state can be converted to another by applying Pauli operation only at one end, we simply take this state to be

$$|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}. \tag{1}$$

In practice, however, noise at multiple stages of the entanglement generation causes the resulting link state to deviate from this ideal target. A common tractable noise model in this respect is the depolarizing noise model [8], [21], where decoherence of a single qubit is captured by the relation

$$\mathcal{N}_t(\rho) = (1-p(t))\rho + p(t)\frac{\mathbb{I}_2}{2}. \tag{2}$$

Here the density matrix ρ (resp. $\mathcal{N}_t(\rho)$) represents the state of the qubit at time 0 (resp. time t) and $p(t) = 1 - e^{-t/t_c}$. Further, t_c is the memory lifetime constant characterizing the depolarizing effect that depends upon the physical properties of the memory qubit holding the state. We assume that depolarizing noise is the only noise acting on the qubits. Under this assumption, any entangled link is described by a Werner state

$$\rho_w = w\,|\Phi^+\rangle\langle\Phi^+| + (1-w)\mathbb{I}_4/4, \tag{3}$$

where w is the Werner parameter. The Werner form is preserved under depolarizing noise and entanglement swaps through the following multiplicative properties:

- When qubits are identical and each undergoes depolarizing noise as per (2), the Werner parameter $w(t)$ after time t from generation decays as

$$w(t) = w_0 e^{-2t/t_c}, \tag{4}$$

where w_0 is the Werner parameter of a freshly generated link.

- Following a Bell state measurement (BSM) at the repeater with two elementary links having Werner parameters w_A and w_B, an end-to-end link with Werner parameter $w_A w_B$ is produced.

A5 **Individual phase durations**: We generally assume a setting where communication time between the nodes is orders of magnitude higher than the time required for local operations so that the durations for the latter can be considered negligible. This is true in the regime where the parties are, for example, separated by a few hundred of kilometers, leading to a propagation time in the order of milliseconds, whereas local operations such as measurement, qubit initialization for link generation, or two-qubit operations such as BSM for entanglement swap have durations of the order of tens of microseconds [23]–[26].

We assume that the durations of the individual phases of end-to-end link generation and those of the reconciliation phases (sifting and QBER test result communication, classical postprocessing) follow shifted exponential distributions with parameters specified in Tab. 1. We now describe the execution order of the end-to-end link generation process.

- Phase `LINK-GEN`: Link-level entanglement is established between an end node (Alice/Bob) and the repeater.
- Phase `L-COMM`: Immediately after, success/failure of the `LINK-GEN` phase is communicated to the end node and the repeater. While the `LINK-GEN` phase usually involves sending photons to a midpoint station hosting a Bell state analyzer, the `L-COMM` phase involves sending the click pattern back to the end node and the repeater. Thus, the durations of both phases are dominated by the photon travel time between an end node or repeater and the midpoint station, and we take them to be IID. Note that `LINK-GEN` and `L-COMM` run in parallel for Alice and Bob.
- Phase `S-COMM`: As soon as both links are successfully generated, a BSM is performed at the repeater, which, if successful, results in end-to-end entanglement between Alice and Bob. A direct failure of the measurement operation is observed with probability $(1 - p_{\text{swap}})$. Otherwise, the measurement result is communicated to Alice and Bob. Subsequently, Alice and Bob individually choose X or Z basis to measure their half of the EPR pair. The bases and the measurement outcomes are kept secret until the process above is repeated n times.

Recall that once all measurements are done, Alice and Bob sift the measurement outcomes and estimate QBER. If the QBER is below the predetermined threshold β, Bob sends a message to Alice indicating if the protocol is successful and they should go ahead with classical post-processing. We denote this phase as `K-COMM`. *Note that if*

V0: a simpler version of the BB84 protocol where Alice and Bob use the same basis for measuring their halves of each end-to-end link and use all measurement outcome for estimating QBER,
V1: actual version of the BB84 protocol,
n: number of end-to-end link generations (and hence, measurements) in a single BB84 attempt,
α: fraction of qubits sampled for QBER estimation among those for which Alice's and Bob's bases match,
β: parameter for QBER test, the protocol is successful if estimated QBER $\leq 1-\beta$,
$W_{n,c}$: completion time of version V0,
W_n: completion time of version V1,
T_{GA} (resp. T_{GB}): duration of the phase `LINK-GEN` between Alice (resp. Bob) and the repeater, distributed as IID SE(λ_ℓ, a_ℓ),
p_{gen}: success probability of link generation, $p_{\text{gen}} < 1$,
T_{CA} (resp. T_{CB}): duration of the phase `L-COMM` between Alice (resp. Bob) and the repeater, distributed as IID SE(λ_ℓ, a_ℓ),
p_{swap}: BSM success probability at the repeater,
$T_C^{'}$: the duration of the phase `S-COMM`, distributed as SE$(\lambda_{\text{swap}}, a_{\text{swap}})$,
w_0 (resp. w): Werner parameter of a freshly generated (resp. at a general point in time) elementary link,
t_c: coherence time for the depolarizing noise acting on single qubits,
K_C: duration of the sifting and QBER estimation phase `K-COMM`, distributed as SE$(\lambda_{\text{AB}}, a_{\text{AB}})$,
P_C: duration of the postprocessing phase `P-COMM`, distributed as SE$(\lambda_{\text{P}}, a_{\text{P}})$,
T_{HA} (resp. T_{HB}): total time until an elementary link is successfully generated between Alice (resp. Bob) and the repeater,
$V_A := T_{HA} + T_{CA}$, $V_B := T_{HB} + T_{CB}$,
T_γ: cumulative duration of failed swap trials until an end-to-end link is successfully generated,
X: time to generate an end-to-end link successfully, shown as `E2E-GEN` in Fig. 1; i.e., $X = T_\gamma + \max\{V_A, V_B\} + T_C^{'}$,
Y: indicator variable denoting that Alice's and Bob's measurement outcomes on respective halves of the entangled link coincide.

TABLE 1: Primary notations.

there is a failure in any intermediate phase, all the previous phases that lead to the concerned phase restart immediately. If the QBER test succeeds, Alice and Bob perform classical postprocessing. We denote this phase as `P-COMM`.

The set of elementary notations required to state some of the main results is given in Tab. 1. A schematic description of the completion time of the protocol is shown in Fig. 1 in terms of the phases mentioned above.

B. MAIN RESULTS

Our main result is a computable expression for the MGF of the protocol completion time in terms of the MGFs of the elementary phases described in assumption A5. We then invert the MGF to derive the CDF of the completion time and provide a bound for tail probabilities using Chernoff's method as well. Notation-wise, we use $M_Z(t)$ to denote the MGF of a random variable Z, whereas the relevant random variables are defined in Tab. 1.

Theorem 1 (Completion time of version V1)**.** *The MGF of the completion time (W_n) of the BB84 protocol is given by:*

$$M_{W_n}(t) = \frac{M_{P_C}(t) M_{K_C}(t) D_n(t)}{1 - M_{K_C}(t)(M_X^n(t) - D_n(t))}, \tag{5}$$

where $M_X^n(t) := (M_X(t))^n$ with X being distributed according to the completion time of an end-to-end link generation (i.e., `E2E-GEN` phase) and

$$D_n(t) := \sum_{k=1}^{\lceil \alpha n \rceil} M^{(1)}(t; k, \lceil \beta k \rceil) M_X^{n-k}(t) \frac{1}{2^n} \sum_{\frac{k-1}{\alpha} < j \leq \frac{k}{\alpha}} \binom{n}{j}. \tag{6}$$

For $j \leq l$, $M^{(1)}(t; l, j)$ is defined as follows:

$$\begin{aligned} M^{(1)}(t; l, 0) &:= (m_0(t) + m_1(t))^l, \\ M^{(1)}(t; l, l) &:= m_1^l(t), \quad l \in \mathbb{N}. \end{aligned} \tag{7}$$

Further, for $l \geq 2$, $1 \leq j \leq l-2$,

$$\begin{aligned} M^{(1)}(t; l, j) := &\sum_{k=0}^{1} \binom{l-k}{j} m_0^{l-k-j}(t) m_1^{j+k}(t) + \\ &m_1(t) \sum_{k=j}^{l-2} (m_0(t) + m_1(t))^{l-1-k} \binom{k}{j} m_0^{k-j}(t) m_1^j(t), \end{aligned} \tag{8}$$

and

$$M^{(1)}(t; l, l-1) := m_1^l(t) + \binom{l}{1} m_0(t) m_1^{l-1}(t).$$

Here, $m_0^l(t) := (m_0(t))^l$, $m_1^l(t) := (m_1(t))^l$, and

$$\begin{aligned} m_1(t) &:= \mathsf{E}\left(e^{tX} Y\right) = \mathsf{E}\left(e^{tX} | Y = 1\right) \mathsf{P}\left(Y = 1\right), \\ m_0(t) &:= \mathsf{E}\left(e^{tX}(1-Y)\right) = \mathsf{E}\left(e^{tX} | Y = 0\right) \mathsf{P}\left(Y = 0\right); \end{aligned} \tag{9}$$

where Y is the indicator variable denoting that Alice's and Bob's measurement outcomes on respective halves of the end-to-end link coincide. The RHS of (9) *can be expressed in terms of the durations of the individual phases within an `E2E-GEN` phase as follows:*

$$\begin{aligned} m_1(t) = \frac{1}{2} M_{T_\gamma}(t) \Big(& I(t, 0) M_{T_C^{'}}(t) \\ & + w_0^2 I(t, \frac{2}{t_c}) M_{T_C^{'}}(t - \frac{2}{t_c}) \Big), \text{ with} \\ M_{T_\gamma}(t) = & \frac{p_{swap}}{1 - (1 - p_{swap}) M_{T_C^{'}}(t) I(t, 0)}, \end{aligned} \tag{10}$$

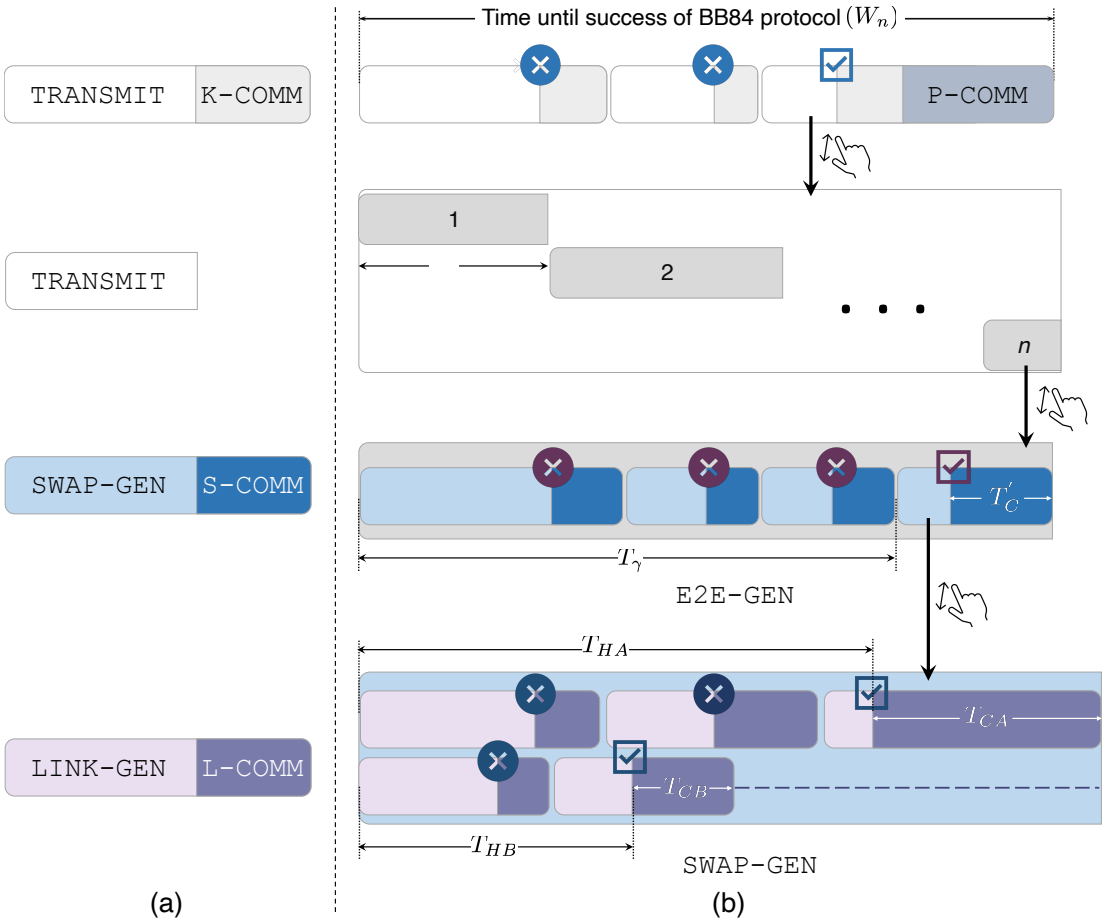


FIGURE 1: Illustration of the completion time of the BB84 protocol, denoted by W_n. Panel (**a**) shows a single trial unit, while panel (**b**) shows independent trials repeated until success, except in the second row. In the first row, a successful trial includes an additional phase. The figure is not to scale. A failed BB84 attempt consists of two phases: qubit transmission (`TRANSMIT`) and sifting with QBER estimation (`K-COMM`); a successful attempt additionally includes a postprocessing phase (`P-COMM`). The `TRANSMIT` phase consists of n sequential `E2E-GEN` trials, each followed by instantaneous measurements at both end nodes. Here, `E2E-GEN` denotes successful end-to-end link generation, whose duration is distributed as X. The MGF of W_n is related to that of X through (28). Each trial within an `E2E-GEN` phase consists of a `SWAP-GEN` phase followed by an `S-COMM` phase. The `S-COMM` phase represents communication of the BSM success or failure outcome from the repeater to an end node. The duration of `SWAP-GEN` is the maximum of the durations of two modules, each corresponding to the completion time of link-level entanglement generation. The trial unit for each such module consists of `LINK-GEN`, i.e., link-level entanglement generation between an end node and the repeater, followed by `L-COMM`, i.e., communication of the `LINK-GEN` success or failure outcome. The dotted line in the last row indicates the waiting time incurred by one link while the other link completes successfully. The variables used for annotation are defined in Tab. 1.

and

$$I(t,u) := \mathsf{E}\left(e^{t\max\{V_A,V_B\}} e^{-u(|V_A-V_B|+T_{CA}+T_{CB})}\right). \quad (11)$$

Also,

$$M_X(t) = M_{T_\gamma}(t) I(t,0) M_{T'_C}(t). \quad (12)$$

Since $m_0(t) = M_X(t) - m_1(t)$, we can compute it by plugging in expressions for $m_1(t)$ and $M_X(t)$. The computation of $I(t,u)$ is described in Sect. III-C.

Using Thm. 1, we can derive the CDF and a bound for the tail probability of W_n as follows.

Corollary 1 (CDF of W_n)**.** *For $s > 0$,*

$$\mathsf{P}(W_n \le s) = \mathcal{L}^{-1}\{F(t)\}(s), \text{ with } F(t) := \frac{M_{W_n}(-t)}{t} \quad (13)$$

where $\mathcal{L}^{-1}$ denotes the inverse Laplace transform. The numerical computation of (13) *is described in Sect. III-C.*

Proof. Let f_n and F_n be the density and CDF of W_n, respectively. Denoting the Laplace transform of a function

W: the completion time of the protocol, $T_1, T_2, \ldots, T_r$: the durations of the constituent tasks of the GENERATE phase, $L_1, L_2, \ldots, L_r$: the latent variable completely determining the duration of the GENERATE phase and success probability of the measurement event, θ: hardware efficiency parameter for measurement, $p(L_1, L_2, \ldots, L_r, \theta)$: the success probability of the measurement event, T_C: COMMUNICATE/LOCC phase duration.

TABLE 2: Notations used to describe a failure-prone protocol in Sect.III-A.

$h(t)$ by $\mathcal{L}\{h(t)\}(s)$, we have $\mathcal{L}\{f_n(t)\}(s) = M_{W_n}(-s)$. Also, $F_n^{'} = f_n$ and $W_n > 0$ implies $F_n(0) = 0$. Thus, $\mathcal{L}\{F_n(t)\}(s) = M_{W_n}(-s)/s$, which gives (13). □

Corollary 2 (Tail bound of the completion time of version V1)**.** *Using Chernoff's method,*

$$\mathsf{P}(W_n > s) \leq \inf_{t\in[0,b)} e^{-ts} M_{W_n}(t), \tag{14}$$

where $b := \sup\{t \in \mathbb{R} : M_{W_n}(t) < \infty\}$. *Note that for any* $t \in [0, b)$, $e^{-ts}M_{W_n}(t)$ *is a valid upper bound for* $\mathsf{P}(W_n > s)$ *which we use for numerically computing the bound in Sect. III-C.*

Proof. This follows directly by applying Chernoff's method to W_n. □

III. METHODS AND PROOFS OF RESULTS

A. BACKGROUND

Following [7]–[9], we first express the completion time W of a failure-prone protocol in terms of the durations of its trial units. We assume that each trial has a GENERATE phase, which depends on r independent tasks. We denote the duration of these tasks by $T_1, T_2, \ldots, T_r$. Once the tasks finish, there is an instantaneous measurement/detection event which succeeds with probability $p(L_1, L_2, \ldots, L_r, \theta)$, where $\{L_i\}_{i\in[r]}$ are latent variables that completely determine $\{T_i\}_{i\in[r]}$ and θ is a parameter reflecting the efficiency of the underlying physical system. If the measurement event results in failure, the process starts over immediately and carries on until success is observed. Note that the success or failure of the event has to be duly communicated so that the process terminates or moves on to the next trial. We call this part of a trial the COMMUNICATE phase and denote its duration by T_C. We assume that T_C is independent of $\{L_i\}_{i\in[r]}$ and that the trials are independent as well. In case the measurement event takes non-negligible time, we can consider T_C to be the total time corresponding to the measurement and communication events. In this case, the COMMUNICATE phase is termed as LOCC phase. For a schematic description of the completion time, see Fig. 2.

Proposition 1. *(i) Let us assume that the* r *tasks in the* GENERATE *phase run in parallel. The MGF of the completion time is then given by:*

$$M_W(t) = \frac{\mathsf{E}(e^{tT_C})\mathsf{E}\big(e^{t\max_{i\in[r]} T_i} p_\theta(\mathbf{L}^{(1)})\big)}{1 - \mathsf{E}(e^{tT_C})\mathsf{E}\big(e^{t\max_{i\in[r]} T_i}(1 - p_\theta(\mathbf{L}^{(1)}))\big)}, \tag{15}$$

for $\mathsf{E}(e^{tT_C})\mathsf{E}\big(e^{t\max_{i\in[r]} T_i}(1-p_\theta(\mathbf{L}^{(1)}))\big) < 1$. *Here,* $[r] := \{1, 2, \ldots, r\}$ *and* $\mathbf{L}^{(k)} := (L_1^{(k)}, L_2^{(k)}, \ldots, L_r^{(k)})$ *is the vector comprising the latent variables of the* r *constituent tasks of the* GENERATE *phase for the* k*th trial, and*

$$\mathbf{L}^{(k)} \overset{\text{iid}}{\sim} (L_1, L_2, \ldots, L_r).$$

(ii) If the r *tasks run in sequence instead,*

$$M_W(t) = \frac{M_{T_C}(t)\mathsf{E}\big(e^{t\sum_{i=1}^r T_i} p_\theta(\mathbf{L}^{(1)})\big)}{1 - M_{T_C}(t)\mathsf{E}\big(e^{t\sum_{i=1}^r T_i}(1-p_\theta(\mathbf{L}^{(1)}))\big)}. \tag{16}$$

Proof. (i) If the tasks run in parallel, the duration of each trial is given by: $U = \max_{i\in[r]} T_i + T_C$. Further, if we denote the success of the measurement event as Y,

$$Y|\{L_i\}_{i\in[r]} \sim \text{Bern}(p(\{L_i\}_{i\in[r]}, \theta)).$$

Let N denote the number of trials until success. Using the shorthand $p(\mathbf{L}, \theta) = p_\theta(\mathbf{L})$,

$$\begin{aligned}&\mathsf{E}\big(e^{tW}\mathbb{1}_{N=k}|\{\mathbf{L}^{(j)}\}_{j=1}^\infty\big)\\ =&e^{t\sum_{j=1}^k U_j} p_\theta(\mathbf{L}^{(k)}) \prod_{j=1}^{k-1}\big(1 - p_\theta(\mathbf{L}^{(j)})\big),\end{aligned} \tag{17}$$

where $U_j = \max_{i\in[r]}\{T_i^{(j)}\} + T_C^{(j)}$. Since the RHS of (17) is non-negative, summing over k and taking expectation with respect to $\{\mathbf{L}^{(j)}\}_{j=1}^\infty$, we have:

$$\begin{aligned}M_W(t) =& \sum_{k=1}^\infty \mathsf{E}\big(e^{tU_1} p_\theta(\mathbf{L}^{(1)})\big)\mathsf{E}^{k-1}\big(e^{tU_1}(1-p_\theta(\mathbf{L}^{(1)}))\big)\\ =&\frac{\mathsf{E}\big(e^{tU_1} p_\theta(\mathbf{L}^{(1)})\big)}{1 - \mathsf{E}\big(e^{tU_1}(1 - p_\theta(\mathbf{L}^{(1)}))\big)}\\ =&\frac{\mathsf{E}(e^{tT_C})\mathsf{E}\big(e^{t\max_{i\in[r]} T_i} p_\theta(\mathbf{L}^{(1)})\big)}{1 - \mathsf{E}(e^{tT_C})\mathsf{E}\big(e^{t\max_{i\in[r]} T_i}(1 - p_\theta(\mathbf{L}^{(1)}))\big)},\end{aligned}$$

for $\mathsf{E}(e^{tU_1}(1 - p_\theta(\mathbf{L}^{(1)}))) < 1$, as claimed.

(ii) The result follows simply by observing that $U_j = \sum_{i=1}^r T_i^{(j)} + T_C^{(j)}$ when tasks are executed in sequence and imitating the proof of part (i). □

The next results are useful for deriving tail bounds for the completion time W.

Proposition 2. *Let* $I_1, I_2, \ldots, I_r, I_C$ *be the neighborhoods of zero where the MGFs of the RVs* $T_1, T_2, \ldots, T_r, T_C$ *exist, respectively. Further, let* I_0 *denote the neighborhood of zero where* $G(t) = \mathsf{E}(e^{tU_1}(1 - p_\theta(\mathbf{L}^{(1)}))) < 1$. *Then the MGF of* W *exists in* $\bigcap_{l\in[r]\cup\{0,C\}} I_l$.

Proof. See Sect. VI-A. □

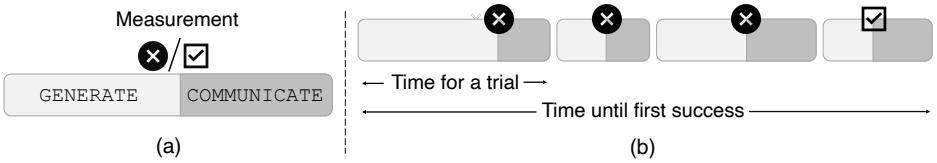


FIGURE 2: Schematic description of the completion time of a failure-prone protocol: (**a**) each trial begins with a GENERATE phase, followed by a measurement/detection event resulting in success (✓) or failure (×). The measurement outcome is in general dependent on granular details of the GENERATE phase and is communicated to relevant parties in the COMMUNICATE/LOCC phase. The durations of the two phases are random and independent. Independent trials are repeated until the measurement event results in success and the completion time is shown in (**b**).

Corollary 3. *If $T_1, T_2, \ldots, T_r, T_C$ are sub-exponential, so is W.*

Proof. See Sect. VI-A. □

Corollary 3 essentially guarantees the existence of the MGF of W in a neighborhood of zero, which implies that we can apply Chernoff's method to obtain an upper bound for the tail probability:

$$\mathsf{P}(W > s) \leq \inf_{t \in [0, \tilde{b})} e^{-ts} \mathsf{E}(e^{tW}), \tag{18}$$

where $\tilde{b} := \sup\{t \in \mathbb{R} : M_W(t) < \infty\}$.

B. PROOFS OF MAIN RESULTS

We now provide the proof of the main results using the methods described in Sect. III-A. We organize this section as follows: we first derive the success probability of an E2E-GEN phase given the durations of its individual phases, where success refers to the event that Alice's and Bob's measurements of their respective halves of the end-to-end link produce the same outcome. This notion of success is unambiguous, as an E2E-GEN phase corresponds to the successful generation of an end-to-end link. This allows us to show that the expressions for m_1 and M_X in (10) and (12) are valid. We then derive the MGF of the completion time of the simplified version V0, where the *main idea* of the proof is to decompose the duration of an E2E-GEN phase into its constituent phases, as shown in Fig. 1, and then apply the methods from Sect. III-A. This serves as a stepping stone for the MGF computation of the completion time of the actual version V1, as stated in Thm. 1.

Proof of (10) *and* (12). First, we refer the reader to Fig. 1, where the duration of the E2E-GEN phase is shown and its constituent phases and respective durations are annotated. Recall that T_{HA} (resp. T_{HB}) denotes the time until Alice (resp. Bob) successfully generates a link-level entanglement and the communication of this event takes T_{CA} (resp. T_{CB}) amount of time. To derive the expression for $M_{T_\gamma}(t)$ in (10), we first see that the swap failure time T_γ comprises $(\text{Geo}(p_{\text{swap}}) - 1)$ independent swap trials (phase SWAP-GEN+S-COMM), each having a duration that is an independent copy of $\max\{T_{HA} + T_{CA}, T_{HB} + T_{CB}\} + T_C^{'}$. Applying (15) now leads to the given expression.

The expression for $M_X(t)$ in (12) now follows from the breakdown of the E2E-GEN phase duration X in Tab. 1:

$$X = T_\gamma + \max\{V_A, V_B\} + T_C^{'}, \tag{19}$$

the definition of $I(t, u)$ in (11), and the fact that each term in the RHS of (19) are independent. We remind the reader of the shorthand $V_A = T_{HA} + T_{CA}$ (resp. for B); see Tab. 1.

Next, we establish the formula for $m_1(t)$. Observe from assumption A4 that after successful generation, the elementary link between Alice (resp. Bob) and the repeater decoheres for a duration distributed as T_{CA} (resp. T_{CB}). Before this, the link generated earlier would have decohered for a duration distributed as $|T_{HA} + T_{CA} - T_{HB} - T_{CB}|$. Thus, a successful swap leads to an end-to-end link between Alice and Bob with Werner parameter:

$$w_0^2 e^{-2(|T_{HA}+T_{CA}-T_{HB}-T_{CB}|+T_{CA}+T_{CB})/t_c},$$

where w_0 denotes the Werner parameter of a freshly generated link and t_c denotes the coherence time of a single quantum memory. The swap success message takes further $T_C^{'}$ time to reach the parties and thus measurement is performed on each qubit of a Werner state with parameter:

$$w = w_0^2 e^{-2(|V_A - V_B| + T_{CA} + T_{CB} + T_C^{'})/t_c}.$$

By a standard property of Werner states (3), given an end-to-end link with Werner paramter w, the probability that Alice and Bob measure the same outcome is given by [21], [22]:

$$\mathsf{P}(Y = 1|w) = \frac{1}{2}(1 + w). \tag{20}$$

Accordingly,

$$\begin{aligned} m_1(t) =& \mathsf{E}\left(e^{tX} Y\right) \\ =& \mathsf{E}\left(\mathsf{E}\left(e^{tX} Y | T_{HA}, T_{CA}, T_{HB}, T_{CB}, T_C^{'}\right)\right) \\ =& \mathsf{E}\Bigg(e^{t(T_\gamma + \max\{V_A, V_B\} + T_C^{'})} \\ & \frac{1 + w_0^2 e^{-2(|V_A - V_B| + T_{CA} + T_{CB} + T_C^{'})/t_c}}{2}\Bigg). \end{aligned}$$

Simple calculations now lead to (10). □

We now prove Lemma 1, which will help us express the MGF of the completion time of the actual version V1 as formalized in Thm. 1.

Lemma 1 (Completion time of version V0)**.** *Let $c := \lceil \beta n \rceil$. The MGF of the completion time ($W_{n,c}$) of version V0 is then given by:*

$$M_{W_{n,c}}(t) = \frac{M_{P_C} M_{K_C}(t) M^{(1)}(t;n,c)}{1 - M_{K_C}(t) M^{(0)}(t;n,c)},$$

where $M^{(1)}$ is defined in Thm. 1 and

$$M^{(0)}(t;l,j) = (m_0(t) + m_1(t))^l - M^{(1)}(t;l,j). \quad (21)$$

Proof. Recall that in version V0, Alice and Bob use the same basis for measuring their halves for each end-to-end link and all measurement outcomes are used for QBER estimation. Let us denote by X_i the duration of the ith `E2E-GEN` phase within a BB84 attempt. Further, let Y_i be the indicator variable assuming the value 1 when Alice's and Bob's measurement results for the corresponding end-to-end link agree. Therefore, $(X_i, Y_i)_i \overset{\text{iid}}{\sim} (X, Y)$. Barring the last attempt, which ends with success, the duration of each BB84 attempt are IID copies of: $K_C + \sum_{i=1}^{n} X_i$, where K_C denotes the duration of the `K-COMM` phase. The last attempt has an additional phase `P-COMM`, whose duration is distributed as P_C and is independent of all phases preceding it. Now, let c be the success threshold of the protocol, i.e., a BB84 attempt succeeds iff $\sum_{i=1}^{n} Y_i \geq c$. Then, using part (ii) of Prop. 1, we have:

$$M_{W_{n,c}}(t) = \frac{M_{P_C}(t) M_{K_C}(t) \mathsf{E}(e^{t\sum_{i=1}^{n} X_i} \mathbb{1}_{\sum_{i=1}^{n} Y_i \geq c})}{1 - M_{K_C}(t) \mathsf{E}(e^{t\sum_{i=1}^{n} X_i} \mathbb{1}_{\sum_{i=1}^{n} Y_i < c})}. \quad (22)$$

Let us define

$$\bar{M}^{(1)}(t;l,j) := \mathsf{E}(e^{t\sum_{i=1}^{l} X_i} \mathbb{1}_{\sum_{i=1}^{l} Y_i \geq j}),$$
$$\bar{M}^{(0)}(t;l,j) := \mathsf{E}(e^{t\sum_{i=1}^{l} X_i} \mathbb{1}_{\sum_{i=1}^{l} Y_i < j}).$$

To complete the proof, we need to show that $\bar{M}^{(1)} \equiv M^{(1)}$ and $\bar{M}^{(0)} \equiv M^{(0)}$. Note that $\bar{M}^{(1)}(t;1,0) = m_0(t) + m_1(t)$ and $\bar{M}^{(1)}(t;1,1) = m_1(t)$. Since $(X_i, Y_i) \overset{\text{iid}}{\sim} (X, Y)$,

$$\begin{aligned} \bar{M}^{(1)}(t;l,0) &= (m_0(t) + m_1(t))^l, \\ \bar{M}^{(1)}(t;l,l) &= m_1^l(t), \quad l \in \mathbb{N}. \end{aligned} \quad (23)$$

Now, for $l \geq 2$, $1 \leq j \leq l$,

$$\begin{aligned} &\bar{M}^{(1)}(t;l,j) \\ &= \mathsf{E}\big(e^{t\sum_{i=1}^{l} X_i} (\mathbb{1}_{Y_1=0} \mathbb{1}_{\sum_{i=2}^{l} Y_i \geq j} + \mathbb{1}_{Y_1=1} \mathbb{1}_{\sum_{i=2}^{l} Y_i \geq j-1})\big) \\ &= \mathsf{E}(e^{tX_1} \mathbb{1}_{Y_1=0}) \mathsf{E}(e^{t\sum_{i=2}^{l} X_i} \mathbb{1}_{\sum_{i=2}^{l} Y_i \geq j}) \\ &\quad + \mathsf{E}(e^{tX_1} \mathbb{1}_{Y_1=1}) \mathsf{E}(e^{t\sum_{i=2}^{l} X_i} \mathbb{1}_{\sum_{i=2}^{l} Y_i \geq j-1}) \\ &= m_0(t) \bar{M}^{(1)}(t;l-1,j) + m_1(t) \bar{M}^{(1)}(t;l-1,j-1). \end{aligned} \quad (24)$$

Here, the second and the third equality follow from the fact that $(X_i, Y_i) \overset{\text{iid}}{\sim} (X, Y)$.

To solve the recurrence relation, let us introduce the following shorthand: $a_{l,j} = \bar{M}^{(1)}(t;l,j)$, $m_0 = m_0(t)$, and $m_1 = m_1(t)$ Therefore, $a_{1,0} = m_0 + m_1$, $a_{1,1} = m_1$, and for $l \geq 2$, $1 \leq j \leq l-1$,

$$a_{l,j} = m_0 a_{l-1,j} + m_1 a_{l-1,j-1}. \quad (25)$$

Also, (23) can be rewritten as:

$$a_{l,0} = (m_0 + m_1)^l, \quad a_{l,l} = m_1^l, \quad l \in \mathbb{N}. \quad (26)$$

We now define the generating function:

$$A_l(x) := \sum_{j=0}^{l} a_{l,j} x^j, \quad l \in \mathbb{N}.$$

Multiplying both sides of (25) by x^j and summing over $1 \leq j \leq l-1$, we have for $l \geq 2$:

$$A_l(x) = (m_0 + m_1 x) A_{l-1}(x) + (m_0 + m_1)^{l-1} m_1, \quad (27)$$

with $A_1(x) = m_0 + m_1 + m_1 x$. Note that we have used (26) to arrive at (27). For $l \geq 2$, recursive substitution in (27) leads to:

$$\begin{aligned} A_l(x) &= (m_0 + m_1 x)^{l-1} A_1(x) \\ &\quad + m_1 \sum_{k=0}^{l-2} (m_0 + m_1 x)^k (m_0 + m_1)^{l-1-k}, \end{aligned}$$

which, after some calculations, yields:

$$\begin{aligned} a_{l,l-1} &= m_1^l + \binom{l}{1} m_0 m_1^{l-1}, \\ a_{l,j} &= \binom{l}{j} m_0^{l-j} m_1^j + \binom{l-1}{j} m_0^{l-1-j} m_1^{j+1} \\ &\quad + m_1 \sum_{k=j}^{l-2} (m_0 + m_1)^{l-1-k} \binom{k}{j} m_0^{k-j} m_1^j, \end{aligned}$$

for $1 \leq j \leq l-2$. It is now straightforward to see that $\bar{M}^{(1)} \equiv M^{(1)}$. Further, $\mathsf{E}(e^{tX}) = m_0(t) + m_1(t)$, which establishes $\bar{M}^{(0)} \equiv M^{(0)}$. □

Proof of Thm. 1. Recall that out of the $B \sim \text{Bin}(n, 1/2)$ measurements where the bases agree, $B_1 = \lceil \alpha B \rceil$ are sampled without replacement for QBER estimation, and the protocol is deemed successful if the corresponding measurement outcomes match for at least β fraction of the sample. Note that the protocol has to be rerun until $B_1 \geq 1$.

Let Z_i denote the Bernoulli RV assuming value 1 iff the measurement outcome corresponding to the end-to-end link generated during ith `E2E-GEN` phase is sampled, i.e., $B_1 = \sum_{i=1}^{n} Z_i$. Further, let k_i's for $i \in [B_1]$ be the indices such that $Z_{k_i} = 1$. Therefore, the protocol is successful when $B_1 \geq 1$ and $\sum_{i=1}^{B_1} Y_{k_i} \geq \beta B_1$, where Y_j is the indicator variable that takes value 1 when Bob's and Alice's measurements on their respective halves of the jth end-to-end link agree. Denoting

$$\bar{D}_n(t) := \mathsf{E}(e^{t\sum_1^n X_i} \mathbb{1}_{\sum_{i=1}^{B_1} Y_{k_i} \geq \beta B_1} \mathbb{1}_{B_1 \geq 1}), \quad (28)$$

part (ii) of Prop. 1 gives

$$M_{W_n}(t) = \frac{M_{P_C}(t)M_{K_C}(t)\bar{D}_n(t)}{1-M_{K_C}(t)(M_X^n(t) - \bar{D}_n(t))},$$

We now show that $\bar{D}_n \equiv D_n$ indeed. Observe that

$$\begin{aligned}\bar{D}_n(t) &= \sum_{k=1}^{\lceil \alpha n \rceil} M^{(1)}(t; k, \lceil \beta k \rceil) M_X^{n-k}(t) \mathsf{P}(B_1 = k) \\ &= \sum_{k=1}^{\lceil \alpha n \rceil} M^{(1)}(t; k, \lceil \beta k \rceil) M_X^{n-k}(t) \frac{1}{2^n} \sum_{\frac{k-1}{\alpha} < j \le \frac{k}{\alpha}} \binom{n}{j},\end{aligned}$$

which completes the proof. □

C. NUMERICAL COMPUTATION OF THE CDF AND THE TAIL PROBABILITY OF THE COMPLETION TIME

We first describe the numerical computation of $I(t,u)$ in (11), which is subsequently used to numerically compute $M_{W_n}(t)$. Recall that T_{HA} (resp. T_{HB}) is given by the sum of the durations of N_A (resp. N_B) `LINK-GEN` and $N_A - 1$ (resp. $N_B - 1$) `L-COMM` trials, where N_A (resp. N_B) denotes the number of trials needed for Alice (resp. Bob) to establish a link-level entanglement successfully, i.e., $N_A, N_B \overset{\text{iid}}{\sim} \text{Geo}(p_{\text{gen}})$. Moreover, T_{HA} and T_{HB} are IID.

Under assumption A5, the durations of `LINK-GEN` and `L-COMM` have the same shifted exponential distribution. We denote by $\underline{G}$ (resp. $\underline{C}$) an RV following the common distribution of T_{GA} and T_{GB} (resp. T_{CA} and T_{CB}). Their MGFs are given by

$$M_{\underline{G}}(t) = M_{\underline{C}}(t) = e^{a_\ell t}\frac{\lambda_\ell}{\lambda_\ell - t}. \tag{29}$$

Similarly, we denote by $\underline{H}$ an RV following the common distribution of T_{HA} and T_{HB}. Conditioning on the number of `LINK-GEN` trials gives

$$\begin{aligned}M_{\underline{H}}(t) &= \sum_{j=1}^{\infty} p_{\text{gen}}(1-p_{\text{gen}})^{j-1} M_{\underline{G}}(t)^j M_{\underline{C}}(t)^{j-1} \\ &= \frac{p_{\text{gen}} M_{\underline{G}}(t)}{1-(1-p_{\text{gen}})M_{\underline{G}}(t)M_{\underline{C}}(t)}.\end{aligned} \tag{30}$$

We further define

$$b_H := \sup\{t \in \mathbb{R} : M_{\underline{H}}(t) < \infty\}. \tag{31}$$

We now derive an integral expression for $I(t,u)$ that can be evaluated directly. Since V_A and V_B are IID and continuous,

$$\begin{aligned}I(t,u) &= 2\mathsf{E}\left(e^{tV_B} e^{-u(V_B - V_A + T_{CA} + T_{CB})} \mathbb{1}_{V_B > V_A}\right) \\ &= 2\mathsf{E}\left(e^{uT_{HA}} e^{(t-u)T_{HB}} e^{(t-2u)T_{CB}} \mathbb{1}_{V_B > V_A}\right).\end{aligned} \tag{32}$$

Following [27], we use the Bromwich representation of the indicator function to evaluate $I(t,u)$. Specifically,

$$\mathbb{1}_{x>0} = \lim_{R\to\infty} \frac{1}{2\pi \mathrm{i}} \int_{\sigma - \mathrm{i}R}^{\sigma + \mathrm{i}R} \frac{e^{zx}}{z}\,\mathrm{d}z, \qquad x \neq 0, \tag{33}$$

for any $\sigma > 0$. Using the independence of T_{HA}, T_{CA}, T_{HB}, and T_{CB}, we have

$$\begin{aligned}I(t,u) = \frac{1}{\pi \mathrm{i}} \int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty} \frac{1}{z} &M_{\underline{H}}(u-z)M_{\underline{C}}(-z) \\ &\times M_{\underline{H}}(t-u+z)M_{\underline{C}}(t-2u+z)\,\mathrm{d}z.\end{aligned} \tag{34}$$

To ensure that the MGFs in the integrand of (34) are well-defined, we choose the contour such that

$$\max\{0, u - b_H\} < \sigma < \min\{u + b_H - \mathrm{Re}(t), 2u + \lambda_\ell - \mathrm{Re}(t)\}. \tag{35}$$

We include a short discussion in Appendix VI-B as to why this is possible in our case and the validity of the interchange of expectation and integration in (34).

For numerical evaluation, letting $z = \sigma + \mathrm{i}\omega$ in (34),

$$\begin{aligned}I(t,u) = \frac{1}{\pi} \int_{-\infty}^{\infty} &\frac{M_{\underline{H}}(u-\sigma-\mathrm{i}\omega)M_{\underline{C}}(-\sigma-\mathrm{i}\omega)}{\sigma+\mathrm{i}\omega} \\ &\times M_{\underline{H}}(t-u+\sigma+\mathrm{i}\omega)M_{\underline{C}}(t-2u+\sigma+\mathrm{i}\omega)\,\mathrm{d}\omega.\end{aligned} \tag{36}$$

CDF computation (Cor. 1): For each value of t required by the numerical inverse Laplace transform, we need to evaluate $I(t,u)$ only for $u = 0$ and $u = 2/t_c$; see (10) and (12). The resulting values are substituted in Thm. 1 to evaluate $M_{W_n}(t)$. To calculate the CDF of W_n according to (13), we use the de Hoog algorithm, as implemented by `invertlaplace` in the `mpmath` library of Python, at a working precision of 30 decimal digits. We choose the inversion degree to be either 22 or 26 by checking the numerical convergence of $\log_{10} \mathsf{P}(W_n > s)$ against the preceding degree, using degree 18 as a baseline.

Chernoff's bound (Cor. 2): Recall that for calculating Chernoff's bound, we minimize $e^{-ts}M_{W_n}(t)$ over $[0, b)$ for a given s, with b as in Cor. 2. Numerically evaluating (36) requires truncating the infinite integration interval, which introduces a truncation error. To ensure that the resulting quantity is indeed an upper bound on the tail probability, we first perform numerical integration of (36) on a finite interval and then calculate a bound on the absolute truncation error. This yields a lower and an upper bound for (36) and we subsequently obtain an upper bound $M^+_{W_n}(t)$ of $M_{W_n}(t)$ via interval arithmetic. Note that we ignore the numerical integration error on the finite interval.

To compute the Chernoff's bound at point s, we numerically select $t^* \in [0, (1-\epsilon)b]$ that minimizes $e^{-ts}M^+_{W_n}(t)$, where $\epsilon > 0$ is chosen to ensure that the numerical evaluation remains strictly inside the domain of convergence. In our implimentations, we use $\epsilon = 10^{-3}$. Since $M^+_{W_n}(t^*) \geq M_{W_n}(t^*)$, we obtain a valid upper bound

$$\mathsf{P}(W_n > s) \leq e^{-t^* s} M^+_{W_n}(t^*). \tag{37}$$

The details of the truncation error bound calculation are provided in Appendix VI-B.

IV. SYNTHETIC SIMULATION OF THE COMPLETION TIME OF THE BB84 PROTOCOL

As seen in Sect. II, a successful realization of the BB84 protocol requires repeated execution of E2E−GEN trials, each consisting of generation and communication phases. This makes simulating the completion time (W_n) of the protocol time-consuming. To that end, we propose an algorithm to efficiently simulate the duration W_n when the distributions of the duration of individual phases follow shifted exponential distribution (assumption A5) and the respective parameters are known or can be estimated.

Main idea: Recall that as part of the BB84 protocol, a fraction of the measurement outcomes with matching bases are used for QBER estimation, and the remaining are left unchecked. Also, an E2E−GEN trial where Alice and Bob use the same basis for measurement is deemed successful if their measurement outcomes agree and failed otherwise. The central idea of the proposed algorithm is to find the distribution of the count of E2E−GEN trials not used for QBER estimation, successful E2E−GEN trials, and failed E2E−GEN trials. Subsequently, we fit a distribution to the generation time of a E2E−GEN trial (i) unconditionally, (ii) given it is successful, and (iii) given it failed. The final duration W_n is then given by the sum total of (i) the aggregated unconditional duration of E2E−GEN trials, (ii) the aggregated duration of successful E2E−GEN trials, (iii) the aggregated duration of failed E2E−GEN trials, (iv) the aggregated sifting and QBER estimation (K−COMM) time, and (v) postprocessing (P−COMM) time. The acceleration of our scheme is due to the fact that such an aggregated duration can be simulated in a constant number of steps, irrespective of the number of E2E−GEN trials n by appropriately choosing the distributions for a single unconditional, successful, and failed E2E−GEN trial. Note that this method is advantageous only when we assume that the simulation exercise will be performed a large number of times. In that case, distribution fitting can be done once, and the output can be reused for subsequent simulation exercises, which would make this approach advantageous.

The next result states that the completion time of a failure-prone protocol can be simulated backwards if the distributions of the individual phase durations are known given the success/failure of an attempt. The result, stated for generic random variables, follows simply by rearranging the terms of the MGF from (15).

Lemma 2. *Let $\underline{Y}$ be a Bernoulli RV and $\underline{X}, \underline{W}, T_C$ be RVs such that*

$$\mathsf{E}\left(e^{t\underline{W}}\right) = \frac{M_{T_C}(t)\mathsf{E}\left(e^{t\underline{X}}\mathbb{1}_{\underline{Y}=1}\right)}{1 - M_{T_C}(t)\mathsf{E}\left(e^{t\underline{X}}\mathbb{1}_{\underline{Y}=0}\right)},$$

where $M_{T_C}(t) := \mathsf{E}\left(e^{tT_C}\right)$ and the MGFs of $\underline{X}$ and T_C exist in a neighbourhood of zero. Then,

$$\underline{W} \overset{d}{=} \tilde{W} := \sum_{j=1}^{N} T_j + \sum_{j=1}^{N-1} X_j^{(0)} + X_1^{(1)},$$

where T_j's, $X_j^{(0)}$'s, $X_j^{(1)}$'s, and N are drawn independently from the following known distributions: $T_j \overset{\text{iid}}{\sim} T_C$, $X_j^{(0)} \overset{\text{iid}}{\sim} \underline{X}|\underline{Y}=0$, $X_j^{(1)} \overset{\text{iid}}{\sim} \underline{X}|\underline{Y}=1$ for $j \in \mathbb{N}$, and $N \sim$ Geo$(\mathsf{P}(\underline{Y}=1))$.

Proof. See Appendix. VI-A. □

Let us now recall a few features of the BB84 protocol from Sect. II: it is deemed successful when at least one measurement outcome pair is sampled for QBER estimation, and out of the sampled outcome pairs, a certain fraction (β) corresponds to same measurement outcome by Alice and Bob (i.e., successful E2E−GEN trial). Reusing notations from Lemma 2 and Sect. II, let N denote the number of times the protocol has to be run to see the first success. That is,

$$N \sim \text{Geo}(\mathsf{P}(U=1)) \text{ , where } U = \mathbb{1}_{\sum_{i=1}^{B_1} Y_{k_i} \geq \beta B_1} \mathbb{1}_{B_1 \geq 1} \text{ ,}$$

and k_i's for $i \in [B_1]$ denote the exhaustive set of indices for which $Z_{k_i} = 1$. Applying Lemma 2, (5) gives:

$$W_n \overset{d}{=} \tilde{W}_n := P_C + \sum_{j=1}^{N} T_j + \sum_{j=1}^{N-1} V_j^{(0)} + V_1^{(1)} \text{ , } \quad (38)$$

where $T_j \overset{\text{iid}}{\sim} K_C$, $V_j^{(l)} \overset{\text{iid}}{\sim} \sum_{i=1}^{n} X_i|U=l$, for $l \in \{0,1\}$ and $j \in \mathbb{N}$, and $N \sim$ Geo$(\mathsf{P}(U=1))$.

Now for a single BB84 trial, let N_S and N_F respectively denote the number of measurement outcome pairs out of the sampled ones that correspond to same outcome for Alice and Bob, and the pairs that correspond to different outcomes, i.e.,

$$N_S = \sum_{i=1}^{B_1} Y_{k_i} \text{ , } \quad N_F = \sum_{i=1}^{B_1} (1 - Y_{k_i}) \text{ .}$$

The following result provides a way to simulate $V_j^{(0)}$'s and $V_j^{(1)}$'s.

Proposition 3. *Let $(N_S^{(l)}, N_F^{(l)}) \sim (N_S, N_F)|U = l$ for $l \in \{0,1\}$. Then,*

$$V_j^{(l)} \overset{d}{=} \tilde{V}_l = \sum_{i=1}^{N_S^{(l)}} X_i^{(1)} + \sum_{i=1}^{N_F^{(l)}} X_i^{(0)} + \sum_{i=1}^{n-N_S^{(l)}-N_F^{(l)}} X_i \text{ ,}$$

where $X_j^{(l)} \overset{\text{iid}}{\sim} X|Y = l$, $l \in \{0,1\}$ and all summands on the RHS are drawn independently of each other and of $(N_S^{(l)}, N_F^{(l)})$ as well.

Proof. See Appendix VI-A. □

We can now use (38) together with Prop. 3 to simulate W_n, the completion time of the protocol, provided we can simulate (i) the number of protocol attempts N, (ii) the conditional counts of successful and failed E2E−GEN trials

$(N_S, N_F)|U$ in a single BB84 attempt, and (iii) the unconditional and conditional E2E-GEN durations distributed as X, $X|Y=1$, and $X|Y=0$. For simulating N, we observe that

$$\begin{aligned}\mathsf{P}(U=1) &= \sum_{k=1}^{\lceil \alpha n \rceil} \mathsf{P}\Big(\sum_{j=1}^{k} Y_{i_j} \geq \beta k\Big)\mathsf{P}(B_1=k) \\ &= \sum_{k=1}^{\lceil \alpha n \rceil} \mathsf{P}\Big(\mathrm{Bin}(k, m_1(0)) \geq \beta k\Big) \frac{1}{2^n} \sum_{\frac{k-1}{\alpha} < j \leq \frac{k}{\alpha}} \binom{n}{j},\end{aligned}$$

where we have used the fact that Y_i's are IID Bernoulli with $\mathsf{P}(Y_i=1)=m_1(0)$. Using the shorthand $p_1 := P(U=1)$,

$$\begin{aligned}&\mathsf{P}(N_S=s, N_F=f|U=0) \\ &= (1-\mathbb{1}_{\frac{s}{s+f}\geq\beta})\ \mathsf{P}\Big(\mathrm{Bin}(s+f, m_1(0))=s\Big) \\ &\quad \times \frac{1}{2^n} \sum_{\frac{s+f-1}{\alpha} < j \leq \frac{s+f}{\alpha}} \binom{n}{j} \quad (1-p_1)^{-1}, \end{aligned} \tag{39}$$

where $\mathsf{P}(\mathrm{Bin}(0,p')=0)=1$ for $p' \in [0,1]$ by convention. Similarly,

$$\begin{aligned}&\mathsf{P}(N_S=s, N_F=f|U=1) \\ &= \mathsf{P}(N_S=s, N_F=f, U=1) p_1^{-1} \\ &= \mathbb{1}_{\frac{s}{s+f}\geq\beta}\ \mathsf{P}\Big(\mathrm{Bin}(s+f, m_1(0))=s\Big) \\ &\quad \times \frac{1}{2^n} \sum_{\frac{s+f-1}{\alpha} < j \leq \frac{s+f}{\alpha}} \binom{n}{j} \quad p_1^{-1}\ . \end{aligned} \tag{40}$$

Finally, we need a way to simulate from the distributions of unconditional and conditional E2E-GEN durations: X, $X|Y=1$, and $X|Y=0$. We approximate them by three shifted Coxian phase-type distributions using the method of moments, which requires calculating the moments of the unconditional and conditional E2E-GEN durations. We estimate these moments by first approximating the distributions of T_{HA} and T_{HB}, the time it takes for the parties to generate an elementary link successfully, by a suitable shifted Gamma distribution. This allows fast approximation of $I(t,u)$ in (11) and the moments of the unconditional and conditional E2E-GEN durations can then be estimated using (10) and (12). The same equations also help us determine the shifts, which turn out to be $a_W = 2a_\ell + a_{\text{swap}}$ for all three distributions. We show in Sect. V that even with the Gamma approximation, we have good agreement between the protocol completion time distributions from the full-scale and synthetic simulations. Since a Coxian phase-type distribution with d phases has $(2d-1)$ parameters, we can compute the first $(2d-1)$ moments as functions of the parameters and solve for the parameters by equating them to the moments derived from (10) and (12). Note that the distribution fitting is done only once, and the output can be reused for subsequent simulation exercises.

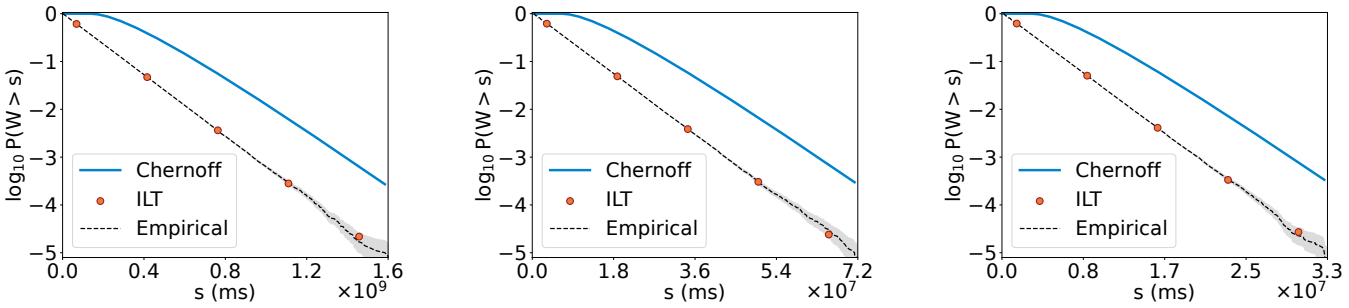
Acceleration summary: The choice of Coxian phase-type distribution is motivated by the fact that their IID sum can be expressed in a compact form and thus can be efficiently simulated. Note that the sum of k IID Coxian phase-type RVs with parameters $(\lambda_1, \ldots, \lambda_d, q_1, \ldots, q_{d-1})$ is distributed as $\sum_{l=1}^{d} \mathrm{Gamma}(D_l, \lambda_l)$ where $D_l \sim \mathrm{Bin}(D_{l-1}, q_{l-1})$ for $l \geq 2$ and $D_1 = k$. The efficiency of our approach follows from the fact that, on average, it needs to simulate $(6d + {}^2/_{p_1})$ RVs per observation vis-à-vis $3 + {}^n/_{p_1}(2 + {}^6/_{p_{\text{swap}}})$ required for the full-scale approach[1]. Recall that $p_1 = \mathsf{P}(U=1)$ and $(1-p_{\text{swap}})$ denotes the observable BSM failure probability at the repeater.

The complete simulation scheme is formalized in Al-

Algorithm 1: Synthetic Simulation of the completion time of the BB84 protocol

Data: $n \in \mathbb{N}$, $\alpha > 0$, $\beta > 0$, $d \in \mathbb{N}$, $p_{\text{gen}} \in (0,1)$, $p_{\text{swap}} \in (0,1)$, $w_0 \in [0,1]$, $t_c > 0$, $a_\ell > 0$, $a_{\text{swap}} > 0$, $a_{\text{AB}} > 0$, $a_{\text{P}} > 0$, $\lambda_\ell > 0$, $\lambda_{\text{swap}} > 0$, $\lambda_{\text{AB}} > 0$, $\lambda_{\text{P}} > 0$.
Result: W_n
Preprocessing:
$p_1 \leftarrow \sum_{k=1}^{\lceil \alpha n \rceil} \mathsf{P}\Big(\mathrm{Bin}(k, m_1(0)) \geq \beta k\Big) \frac{1}{2^n} \sum_{\frac{k-1}{\alpha} < j \leq \frac{k}{\alpha}} \binom{n}{j}$;
Estimate $(\lambda_1^{(1)}, \ldots, \lambda_d^{(1)}, q_1^{(1)}, \ldots, q_{d-1}^{(1)})$ for $X - a_W | Y=1$ via moment-matching;
Estimate $(\lambda_1^{(0)}, \ldots, \lambda_d^{(0)}, q_1^{(0)}, \ldots, q_{d-1}^{(0)})$ for $X - a_W | Y=0$ via moment-matching;
Estimate $(\lambda_1^{(u)}, \ldots, \lambda_d^{(u)}, q_1^{(u)}, \ldots, q_{d-1}^{(u)})$ for $X - a_W$ via moment-matching;
Simulation:
Draw $N \sim \mathrm{Geo}(p_1)$;
Draw (S_0, F_0) according to (40);
for $i = 1 : N-1$ **do**
  Draw (S_i, F_i) according to (39);
end
$S \leftarrow \sum_{i=0}^{N-1} S_i$;
$F \leftarrow \sum_{i=0}^{N-1} F_i$;
$\tilde{N} \leftarrow nN - S - F$;
$A_0 \leftarrow S$;
Draw $W_A^{(0)} \sim \mathrm{Gamma}(A_0, \lambda_1^{(1)})$;
$B_0 \leftarrow F$;
Draw $W_B^{(0)} \sim \mathrm{Gamma}(B_0, \lambda_1^{(0)})$;
$C_0 \leftarrow \tilde{N}$;
Draw $W_C^{(0)} \sim \mathrm{Gamma}(C_0, \lambda_1^{(u)})$;
for $i = 1 : d-1$ **do**
  Draw $A_i \sim \mathrm{Bin}(A_{i-1}, q_i^{(1)})$;
  Draw $W_A^{(i)} \sim \mathrm{Gamma}(A_i, \lambda_{i+1}^{(1)})$;
  Draw $B_i \sim \mathrm{Bin}(B_{i-1}, q_i^{(0)})$;
  Draw $W_B^{(i)} \sim \mathrm{Gamma}(B_i, \lambda_{i+1}^{(0)})$;
  Draw $C_i \sim \mathrm{Bin}(C_{i-1}, q_i^{(u)})$;
  Draw $W_C^{(i)} \sim \mathrm{Gamma}(C_i, \lambda_{i+1}^{(u)})$;
end
Draw $C \sim \mathrm{Gamma}(N, \lambda_{\text{AB}})$;
Draw $\tilde{C} \sim \mathrm{Exp}(\lambda_{\text{P}})$;
$W_n \leftarrow a_{\text{P}} + \tilde{C} + N a_{\text{AB}} + C + n N a_W + \sum_{i=0}^{d-1} (W_A^{(i)} + W_B^{(i)} + W_C^{(i)})$;

[1] Assuming shifted Exponentials are aggregated as shifted Gamma whenever possible.



(a) $(p_{\text{gen}}, w_0) = (10^{-4}, 0.981)$ (b) $(p_{\text{gen}}, w_0) = (5 \times 10^{-4}, 0.904)$ (c) $(p_{\text{gen}}, w_0) = (10^{-3}, 0.809)$

FIGURE 3: Numerical estimates of the CDF via inverse Laplace transform ('ILT') and the Chernoff's bound ('Chernoff') vis-à-vis the empirical tail probability ('Empirical') of the completion time (W_n) for $n = 50$ qubits; the shaded region represents 95% confidence interval for the CCDF computed according to [28, Thm. 2.4]. The empirical distribution is obtained by running a full-scale simulation; see Tab. 3 for an exhaustive list of parameters of the simulation. Time unit: milliseconds.

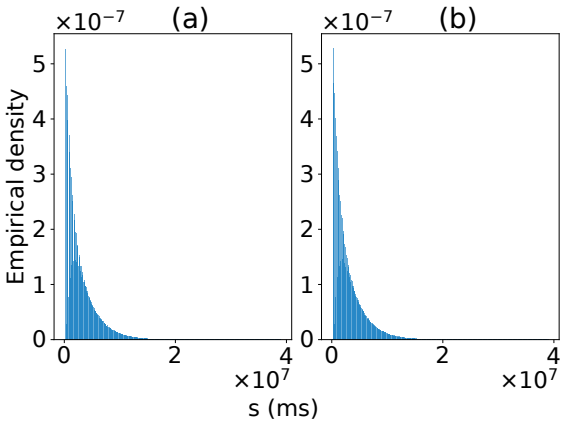


FIGURE 4: Comparison of the empirical densities of the protocol completion time W_n from a (**a**) full-scale and a (**b**) synthetic simulation. Parameters: $n = 50$ qubits, elementary link generation success probability and Werner paramter combination $(p_{\text{gen}}, w_0) = (10^{-3}, 0.809)$; other parameters are provided in Tab. 3.

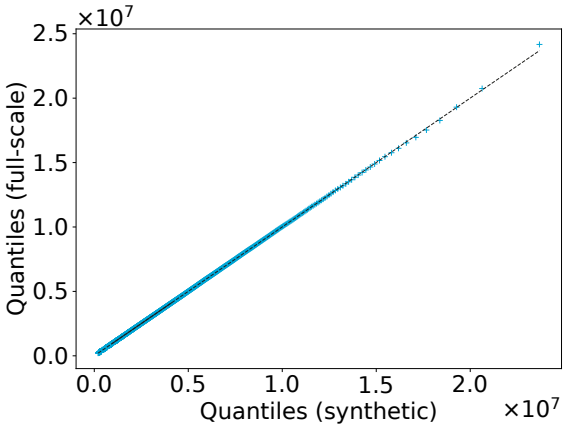


FIGURE 5: Q-Q plot of the protocol completion times from a full-scale and a synthetic simulation; the experimental setup is identical to Fig. 4.

TABLE 3: Parameters used for numerical evaluation.

Parameter	Description	Value
System and duration parameters		
(p_{gen}, w_0)	rate-Werner parameter combination for elementary link generation	$(10^{-4}, 0.981)$, $(5\times10^{-4}, 0.904)$, $(10^{-3}, 0.809)$
t_c	memory coherence time	10 s
p_{swap}	swap success probability	$1/2$
a_ℓ	`LINK-GEN`/`L-COMM` shift	$1/2$ ms
λ_ℓ	`LINK-GEN`/`L-COMM` rate	$2\,\text{ms}^{-1}$
a_{swap}	`S-COMM` shift	1 ms
λ_{swap}	`S-COMM` rate	$1\,\text{ms}^{-1}$
$a_{\text{AB}}, a_{\text{P}}$	`K-COMM`/`P-COMM` shift	2 ms
$\lambda_{\text{AB}}, \lambda_{\text{P}}$	`K-COMM`/`P-COMM` rate	$1/2\,\text{ms}^{-1}$
BB84 parameters		
n	total number of measurements	50
α	sample fraction for QBER estimation	30%
β	parameter determining QBER threshold	89%

gorithm 1. We also compare our approach to a full-scale simulation based on 10^6 observations in Fig. 4, where the histograms of the outputs from the two approaches are compared, and in Fig. 5, where a QQ-plot of these outputs is presented. In the synthetic approach, we use Coxian phase-type distributions with 7 phases as they turn out to be reasonably good fits. The exhaustive list of parameters used for the simulations can be found in Sect.V.

V. NUMERICAL EVALUATION

In this section, we evaluate the CDF estimate and the tail bound proposed in Cor. 1 and 2 numerically. In our simulation setup, we assume that the distance between Alice or Bob and the repeater is $200\,\text{km}$ and that the elementary links are generated using the single-click protocol [29], with the Bell-state analyzer placed at a midpoint station. Commensurate with the speed of light in optical fiber, $2 \times 10^5\,\text{km/s}$, the parameters for the shifted exponentials modelling the durations of the phases `LINK-GEN`, `L-COMM`, `S-COMM`, `K-COMM`, and `P-COMM` are set. Since `S-COMM` traverses

twice the distance of LINK−GEN or L−COMM, and K−COMM and P−COMM twice that again, we scale the corresponding shifted exponentials with distance: the shift doubles and the rate halves at each step. The time unit throughout the numerical evaluation is milliseconds. We consider probabilistic entanglement swapping at the repeater, rather than gate-based deterministic swapping. The system and BB84 protocol parameters used in the evaluation are summarized in Tab. 3.

For the CDF of the completion time and the corresponding tail probability, we first run a full-scale simulation and compare the CDF estimate and the corresponding bound with the empirical distribution. We consider three combinations of the elementary link generation probability p_{gen} and the Werner parameter w_0 of a freshly generated link, as listed in Tab. 3. These combinations reflect projected near-term performance estimates for trapped-ion-based implementations of the single-click protocol over a $200\,\text{km}$ distance; see [30, App. B]. We similarly use the projected memory coherence time $t_c = 10\,\text{s}$ [30, Tab. 2]. For each BB84 run, a total of $n = 50$ measurements are performed, and $\alpha = 30\%$ of the outcomes corresponding to matching measurement bases are sampled for QBER estimation. The protocol is deemed successful if at least one outcome is sampled and the measurement results agree for at least $\beta = 89\%$ of the sampled outcomes. In the resulting plots in Fig. 3, we see that our estimate of the CDF matches closely with the empirical CDF for sampled points.

Next, we demonstrate the effectiveness of the simulation scheme described in Algorithm 1. We compare protocol completion times generated using our scheme with those obtained from a full-scale simulation for a system where the rate-fidelity tradeoff for the elementary link generation is set as $(p_{\text{gen}}, w_0) = (10^{-3},\ 0.809)$, while keeping the remaining parameters the same as in Fig. 3. The resulting empirical densities are compared in Fig. 4, while the corresponding Q-Q plot is shown in Fig. 5. We observe good agreement between the two simulation approaches.

VI. CONCLUSION

In this work, we have done a performance analysis of the completion time of the BB84 protocol. Our setup assumes that the communicating parties are connected by a single quantum repeater, and we focus on the performance of the protocol in terms of its completion time rather than its security aspects. To reflect the current quantum hardware standards, we, however, consider the possibility of failure at every individual phase of end-to-end link generation and take into account the resulting effect of decoherence on the performance of the protocol via a simplified noise model. We subsequently provide a method to calculate the MGF of the completion time, which lets us calculate an estimate of the CDF via inverse Laplace transform and a numerical bound for the corresponding tail probability. Making certain assumptions on the distributions of the durations of individual phases, we also propose an efficient simulation scheme for the completion time. The simulation scheme is based on the idea that the duration of a successful end-to-end link generation event can be well-approximated by a Coxian phase-type distribution, which can be efficiently aggregated to arrive at the completion time of the whole protocol.

APPENDIX

A. DEFERRED PROOFS

Proof of Prop. 2. For $t \leq 0$,

$$e^{t \max_{i\in[n]} T_i} p_\theta(\mathbf{L}_1) \leq e^{tT_1} \ .$$

Similarly, for $t > 0$,

$$e^{t \max_{i\in[n]} T_i} p_\theta(\mathbf{L}_1) \leq e^{t \sum_{j=1}^n T_j} \ .$$

Since $M_{T_C}(t) \geq 0$, for $t \in \mathbb{R}$ we have:

$$\begin{aligned} &M_{T_C}(t)\mathsf{E}\big(e^{t \max_{i\in[n]} T_i} p_\theta(\mathbf{L}_1)\big) \\ \leq &M_{T_C}(t) \max\big\{\mathsf{E}(e^{tT_1}), \prod_{j=1}^n \mathsf{E}(e^{tT_j})\big\} \ . \end{aligned}$$

The RHS is finite on $\bigcap_{l\in[n]\cup\{c\}} I_l$. The same holds for the denominator of the RHS of (15). Further, the infinite sum on the second line of (15) converges on I_0, which proves the claim. □

Proof of Cor. 3. Recall that an RV is sub-exponential iff its MGF exists in a neighborhood of zero. Thus, given the hypothesis, it is enough to show that I_0 is non-empty. For $t_1 < t_2$ and $t_1, t_2 \in \cap_{i\in[n]\cup\{C\}} I_l$,

$$\begin{aligned} G(t_2) - G(t_1) =&\mathsf{E}\big((e^{t_2 X_1} - e^{t_1 X_1})(1 - p_\theta(\mathbf{L}_1))\big) \\ \leq&\mathsf{E}(e^{t_2 X_1}) - \mathsf{E}(e^{t_1 X_1}) \ . \end{aligned}$$

Thus, continuity of the MGF of X_1 in $\cap_{i\in[n]\cup\{C\}} I_l$ implies continuity of G. Further, barring the trivial case $p_\theta(\mathbf{L}_1) \equiv 0$, $G(0) = \mathsf{E}(1 - p_\theta(\mathbf{L}_1)) < 1$. Since G is continuous, I_0 is non-empty as claimed. □

Proof of Lemma 2. Denoting $p_l = P(Y = l)$ and $\phi_l(t) = \mathsf{E}\Big(e^{tX_1^{(l)}}\Big) = \mathsf{E}\big(e^{tX}\mathbb{1}_{Y=l}\big)/p_l$ for $l \in \{0, 1\}$,

$$\begin{aligned} \mathsf{E}\Big(e^{t\tilde{W}}\Big) =&M_{T_C}(t)\phi_1(t) \sum_{j=0}^{\infty} p_0^j p_1 (M_{T_C}(t)\phi_0(t))^j \\ =&\frac{p_1 M_{T_C}(t)\phi_1(t)}{1 - p_0 M_{T_C}(t)\phi_0(t)} = \mathsf{E}\big(e^{tW}\big) \ . \end{aligned}$$

Following the argument of Corr. 3, MGFs of W and $\tilde{W}$ exist in a neighborhood of zero, implying that $W \overset{d}{=} \tilde{W}$. □

Proof of Prop. 3. Observe that

$$\begin{aligned}\mathsf{E}\Big(e^{tV_j^{(l)}}\Big) =&\mathsf{E}\Big(e^{t\sum_1^n X_i}|U=l\Big)\\ =&\sum_{k,d}\mathsf{E}\big(e^{t\sum_1^n X_i}|\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d,\\ &U=l\big)\mathsf{P}\big(\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d|U=l\big)\\ =&\sum_{k,d}\mathsf{E}\big(e^{t\sum_1^n X_i}|\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d\big)\\ &\mathsf{P}\big(\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d|U=l\big)\ ,\end{aligned}$$

where the last equality follows from the definition of U. Now, for $X\sim X_1$ and $Y\sim Y_1$,

$$\begin{aligned}&\mathsf{E}\big(e^{t\sum_1^n X_i}|\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d\big)\\ =&\sum_{r=k}^{k+n-d}\mathsf{E}\big(e^{t\sum_1^n X_i}|\sum_1^n Y_i=r,\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d\big)\\ &\mathsf{P}\big(\sum_1^n Y_i=r|\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d\big)\\ =&\sum_{r=k}^{k+n-d}\mathsf{E}\big(e^{t\sum_1^n X_i}|\sum_1^n Y_i=r\big)\\ &\mathsf{P}\big(\sum_1^n Y_i=r|\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d\big)\\ =&\mathsf{E}^k\big(e^{tX}|Y=1\big)\mathsf{E}^{d-k}\big(e^{tX}|Y=0\big)\sum_{r=0}^{n-d}\binom{n-d}{r}\mathsf{E}^r\big(e^{tX}|Y=1\big)\\ &(\mathsf{P}\left(Y=1\right))^r\mathsf{E}^{n-d-r}\big(e^{tX}|Y=0\big)(\mathsf{P}\left(Y=0\right))^{n-d-r}\\ =&\mathsf{E}^k\big(e^{tX}|Y=1\big)\mathsf{E}^{d-k}\big(e^{tX}|Y=0\big)\mathsf{E}^{n-d}\big(e^{tX}\big).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathsf{E}\Big(e^{tV_j^{(l)}}\Big)\\ =&\sum_{k,d}\mathsf{E}^k\big(e^{tX}|Y=1\big)\mathsf{E}^{d-k}\big(e^{tX}|Y=0\big)\\ &\mathsf{E}^{n-d}\big(e^{tX}\big)\mathsf{P}\big(\sum_1^n Y_iZ_i=k,\sum_1^n Z_i=d|U=l\big)\\ =&\sum_{k,d}\mathsf{E}^k\big(e^{tX}|Y=1\big)\mathsf{E}^{d}\big(e^{tX}|Y=0\big)\\ &\mathsf{E}^{n-d-k}\big(e^{tX}\big)\mathsf{P}\big(N_S=k,N_F=d|U=l\big)\ ,\end{aligned}$$

which establishes $V_j^{(l)}\overset{d}{=}\tilde{V}_l$ for $l\in\{0,1\}$. □

B. DETAILS OF THE NUMERICAL COMPUTATION

Interchange of expectation and integration in (34): We first justify the use of the Bromwich representation. For $\sigma>0$ and $R>0$, let

$$B_R(x):=\frac{1}{2\pi\mathrm{i}}\int_{\sigma-\mathrm{i}R}^{\sigma+\mathrm{i}R}\frac{e^{zx}}{z}\,\mathrm{d}z. \tag{41}$$

For $x\neq 0$, the Bromwich inversion formula gives

$$\lim_{R\to\infty}B_R(x)=\mathbb{1}_{x>0}. \tag{42}$$

Moreover, parameterizing $z=\sigma+\mathrm{i}\omega$ gives

$$B_R(x)=\frac{e^{\sigma x}}{\pi}\int_0^R\frac{\sigma\cos(\omega x)+\omega\sin(\omega x)}{\sigma^2+\omega^2}\,\mathrm{d}\omega.$$

It can be directly verified via a Dirichlet test that there exist K_σ independent of R and x such that

$$|B_R(x)|\le K_\sigma e^{\sigma x},\qquad x\neq 0. \tag{43}$$

Since V_A and V_B are continuous, $\mathsf{P}(V_A=V_B)=0$. Let

$$Y_R:=2e^{uT_{HA}}e^{(t-u)T_{HB}}e^{(t-2u)T_{CB}}B_R(V_B-V_A).$$

For $z=\sigma+\mathrm{i}\omega$ and finite R,

$$\begin{aligned}&\int_{-R}^R\frac{1}{\pi|z|}\mathsf{E}\Big[\Big|e^{uT_{HA}}e^{(t-u)T_{HB}}e^{(t-2u)T_{CB}}e^{z(V_B-V_A)}\Big|\Big]\,\mathrm{d}\omega\\ &\le\frac{2R}{\pi\sigma}M_{\underline{H}}(u-\sigma)M_{\underline{C}}(-\sigma)M_{\underline{H}}(\mathrm{Re}(t)-u+\sigma)\\ &\quad\times M_{\underline{C}}(\mathrm{Re}(t)-2u+\sigma)<\infty,\end{aligned}$$

when σ satisfies (35). Thus, Fubini's theorem gives

$$\begin{aligned}\mathsf{E}(Y_R)=\frac{1}{\pi\mathrm{i}}\int_{\sigma-\mathrm{i}R}^{\sigma+\mathrm{i}R}&\frac{1}{z}M_{\underline{H}}(u-z)M_{\underline{C}}(-z)\\ &\times M_{\underline{H}}(t-u+z)M_{\underline{C}}(t-2u+z)\,\mathrm{d}z.\end{aligned} \tag{44}$$

Further, (43) gives

$$\begin{aligned}|Y_R|\le 2K_\sigma e^{(u-\sigma)T_{HA}}&e^{-\sigma T_{CA}}e^{(\mathrm{Re}(t)-u+\sigma)T_{HB}}\\ &\times e^{(\mathrm{Re}(t)-2u+\sigma)T_{CB}}=:D,\end{aligned} \tag{45}$$

with $\mathsf{E}(D)<\infty$. Therefore, by (42) and the dominated convergence theorem,

$$I(t,u)=\lim_{R\to\infty}\mathsf{E}[Y_R],$$

and (34) follows. The tail bound in (50) below also shows that the resulting infinite contour integral is absolutely convergent.

Choice of the contour: We next show that a σ satisfying (35) exists at all points required for the numerical computations. Recall from Cor. 1 that

$$\mathsf{P}(W_n\le s)=\mathcal{L}^{-1}\left\{\frac{M_{W_n}(-t)}{t}\right\}(s). \tag{46}$$

Numerical inverse Laplace transform algorithms such as the de Hoog method evaluate the transform in (46) at complex points $p=\gamma+\mathrm{i}\omega$ with $\gamma>0$. Hence, the argument supplied to M_{W_n}, and consequently to $I(t,u)$, is $t=-p$, so that $\mathrm{Re}(t)=-\gamma<0$. For $u=0$ or $u=2/t_c$, we have

$$\begin{aligned}\max\{0,u-b_H\}&\le u<u+b_H-\mathrm{Re}(t),\\ \max\{0,u-b_H\}&\le u<2u+\lambda_\ell-\mathrm{Re}(t),\end{aligned}$$

where we have used $b_H>0$, $\lambda_\ell>0$, and $u\ge 0$. Therefore, a σ satisfying (35) exists at every point required by the inverse Laplace transform.

For Chernoff bound calculation, $t \in \mathbb{R}$ and satisfies $M_{W_n}(t) < \infty$, which implies $M_{\underline{H}}(t) < \infty$ and $M_{\underline{C}}(t) < \infty$. Therefore, $t < b_H$ and $t < \lambda_\ell$. Thus for any non-negative u and specifically for $u = 0$ or $u = 2/t_c$,

$$\max\{0, u - b_H\} < \min\left\{u + b_H - t,\, 2u + \lambda_\ell - t\right\},$$

which allows us to pick σ according to (35) at every point used in the Chernoff calculation.

Truncation error and Chernoff's bound: We finally bound the error resulting from truncating the infinite integral in (36). We consider $t \in [0, b)$, as required for the Chernoff calculation. For $\Omega > 0$, let $I_\Omega(t,u)$ denote the integral in (36) restricted to $-\Omega \le \omega \le \Omega$. We define

$$\kappa(x) := \lambda_\ell e^{a_\ell x}, \qquad q_{\text{gen}} := 1 - p_{\text{gen}}.$$

For $|\omega| \ge \Omega$,

$$\left|M_{\underline{C}}(x + \mathrm{i}\omega)\right| = \frac{\lambda_\ell e^{a_\ell x}}{\sqrt{(\lambda_\ell - x)^2 + \omega^2}} \le \frac{\kappa(x)}{|\omega|}. \tag{47}$$

Further, from (30),

$$\left|M_{\underline{H}}(x + \mathrm{i}\omega)\right| \le \frac{p_{\text{gen}}\kappa(x)/|\omega|}{1 - q_{\text{gen}}\kappa(x)^2/\omega^2}.$$

Consequently, provided $\Omega^2 > q_{\text{gen}}\kappa(x)^2$,

$$\left|M_{\underline{H}}(x+\mathrm{i}\omega)\right| \le \frac{h_\Omega(x)}{|\omega|},\ h_\Omega(x) := \frac{p_{\text{gen}}\kappa(x)}{1 - q_{\text{gen}}\kappa(x)^2/\Omega^2}. \tag{48}$$

For fixed t, u, and σ, let

$$x_1 = u - \sigma, \quad x_2 = -\sigma, \quad x_3 = t - u + \sigma, \quad x_4 = t - 2u + \sigma.$$

Choose Ω such that

$$\Omega > \sqrt{q_{\text{gen}}}\max\{\kappa(x_1), \kappa(x_3)\}. \tag{49}$$

Using (47), (48), and $|\sigma + \mathrm{i}\omega| \ge |\omega|$, for $|\omega| \ge \Omega$ the magnitude of the integrand in (36) is bounded by

$$\frac{h_\Omega(x_1)\kappa(x_2)h_\Omega(x_3)\kappa(x_4)}{\pi|\omega|^5}.$$

Consequently,

$$\begin{aligned} |I(t,u) - I_\Omega(t,u)| &\le 2\int_\Omega^\infty \frac{h_\Omega(x_1)\kappa(x_2)h_\Omega(x_3)\kappa(x_4)}{\pi\omega^5}\,\mathrm{d}\omega \\ &= \frac{h_\Omega(x_1)\kappa(x_2)h_\Omega(x_3)\kappa(x_4)}{2\pi\Omega^4} \\ &=: \Delta_\Omega(t,u). \end{aligned} \tag{50}$$

Thus,

$$I(t,u) \in \mathcal{I}_u(t) := [\max\{0, I_\Omega(t,u) - \Delta_\Omega(t,u)\},\, I_\Omega(t,u) + \Delta_\Omega(t,u)]. \tag{51}$$

Note that $\Delta_\Omega(t,u) \to 0$ as $\Omega \to \infty$, since $h_\Omega(x) \to p_{\text{gen}}\kappa(x)$ while the denominator in (50) grows as Ω^4.

For a fixed t, we compute the intervals $\mathcal{I}_0(t)$ and $\mathcal{I}_{2/t_c}(t)$ and propagate them through (10), (12), and (5) using interval arithmetic. The MGFs not involving $I(t,u)$ are evaluated directly. This yields

$$M_{W_n}(t) \in \mathcal{M}_{W_n}(t) = \left[M^-_{W_n}(t), M^+_{W_n}(t)\right]. \tag{52}$$

Importantly, we ensure that the lower end points of the intervals for all denominators arising in this propagation are strictly positive. Otherwise, Ω is increased until positivity is ensured. We control the effect of contour truncation on the final MGF as follows. For a prescribed tolerance $\tau_{\text{tr}} > 0$, we increase Ω successively by a factor of two, until

$$\frac{M^+_{W_n}(t) - M^-_{W_n}(t)}{M^+_{W_n}(t)} \le \tau_{\text{tr}}. \tag{53}$$

In our numerical calculations, we use $\tau_{\text{tr}} = 10^{-4}$.

REFERENCES

[1] Pirandola, S., Laurenza, R., Ottaviani, C. and Banchi, L., "Fundamental limits of repeaterless quantum communications", Nature Communications, vol. 8, 15043, 2017.

[2] Wootters, W.K. and Zurek, W.H., "A single quantum cannot be cloned", Nature, vol. 299, no. 5886, pp. 802–803, 1982.

[3] Briegel, H.-J., Dür, W., Cirac, J.I. and Zoller, P., "Quantum repeaters: The role of imperfect local operations in quantum communication", Physical Review Letters, vol. 81, no. 26, p. 5932, 1998.

[4] Van Meter, R., "Quantum networking", John Wiley & Sons, 2014.

[5] Munro, W.J., Azuma, K., Tamaki, K. and Nemoto, K., "Inside quantum repeaters", IEEE Journal of Selected Topics in Quantum Electronics, 21(3), pp.78-90, 2015.

[6] Azuma, K., Economou, S.E., Elkouss, D., Hilaire, P., Jiang, L., Lo, H.-K. and Tzitrin, I., "Quantum repeaters: From quantum networks to the quantum internet", Reviews of Modern Physics, vol. 95, no. 4, 045006, 2023.

[7] Brand, S., Coopmans, T. and Elkouss, D., "Efficient computation of the waiting time and fidelity in quantum repeater chains", IEEE Journal on Selected Areas in Communications, 38(3), pp.619-639, 2020.

[8] Li, B., Coopmans, T. and Elkouss, D., "Efficient optimization of cut-offs in quantum repeater chains", In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), pp.158-168, 2020.

[9] Coopmans, T., Brand, S. and Elkouss, D., "Improved analytical bounds on delivery times of long-distance entanglement", Physical Review A, vol. 105, no. 1, 012608, 2022.

[10] Bennett C.H. and Brassard G., "Quantum cryptography: Public key distribution and coin tossing", International Conference on Computers, Systems & Signal Processing, pp.175-179, 1984.

[11] Scarani, V., Bechmann-Pasquinucci, H., Cerf, N.J., Dušek, M., Lütkenhaus, N. and Peev, M., "The security of practical quantum key distribution", Reviews of Modern Physics, vol. 81, no. 3, pp. 1301-1350, 2009.

[12] Avis, G., Ferreira da Silva, F., Coopmans, T., Dahlberg, A., Jirovská, H., Maier, D., Rabbie, J., Torres-Knoop, A. and Wehner, S., "Requirements for a processing-node quantum repeater on a real-world fiber grid", npj Quantum Information, vol. 9, 100, 2023.

[13] Ferreira da Silva, F., Avis, G., Slater, J.A. and Wehner, S., "Requirements for upgrading trusted nodes to a repeater chain over 900 km of optical fiber", Quantum Science and Technology, vol. 9, no. 4, 045041, 2024.

[14] van Dam, J., Avis, G., Propp, T.B., Ferreira da Silva, F., Slater, J.A., Northup, T.E. and Wehner, S., "Hardware requirements for trapped-ion based verifiable blind quantum computing with a measurement-only client", Quantum Science and Technology, vol. 9, no. 4, 045031, 2024.

[15] Maiti, S., Avis, G., Kar, S. and Wehner, S., "Requirements for Teleportation in an Intercity Quantum Network", arXiv preprint arXiv:2602.04869, 2026.

[16] Qiao, H. and Chen, X.Y., "Simulation of BB84 Quantum Key Distribution in depolarizing channel", In Proceedings of 14th Youth Conference on Communication, pp.123-129, 2009.

[17] Sahoo, J.R. and Satapathy, S., "Simulation and analysis of BB84 protocol by model checking", International Journal of Engineering Science and Technology (IJEST), 3(7), 2011.

[18] Halip, N.H.M., Mokhtar, M. and Buhari, A., "Simulation of Bennet and Brassard 84 protocol with Eve's attacks", In 2014 IEEE 5th International Conference on Photonics (ICP), pp.29-31, IEEE, 2014.
[19] Jasim, O.K., Abbas, S., El-Horbaty, E.S.M. and Salem, A.B.M., "Quantum key distribution: simulation and characterizations", Procedia Computer Science, 65, pp.701-710, 2015.
[20] Mina, M.Z. and Simion, E., "A scalable simulation of the BB84 protocol involving eavesdropping", In Innovative Security Solutions for Information Technology and Communications: 13th International Conference, SecITC, Bucharest, Romania, November 19–20, Revised Selected Papers 13, pp.91-109, Springer International Publishing, 2021.
[21] Nielsen, M.A. and Chuang, I., "Quantum computation and quantum information", 2002.
[22] Werner, R.F., "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model", Physical Review A, 40(8), p.4277, 1989.
[23] Stas, P.-J., Huan, Y.Q., Machielse, B., Knall, E.N., Suleymanzade, A., Pingault, B., Sutula, M., Ding, S.W., Knaut, C.M., Assumpcao, D.R., Wei, Y.-C., Bhaskar, M.K., Riedinger, R., Sukachev, D.D., Park, H., Lončar, M., Levonian, D.S. and Lukin, M.D., "Robust multi-qubit quantum network node with integrated error detection", Science, vol. 378, no. 6619, pp. 557–560, 2022.
[24] Bradley, C.E., de Bone, S.W., Möller, P.F.W., Baier, S., Degen, M.J., Loenen, S.J.H., Bartling, H.P., Markham, M., Twitchen, D.J., Hanson, R., Elkouss, D. and Taminiau, T.H., "Robust quantum-network memory based on spin qubits in isotopically engineered diamond", npj Quantum Information, vol. 8, 122, 2022.
[25] Hermans, S.L.N., Pompili, M., Beukers, H.K.C., Baier, S., Borregaard, J. and Hanson, R., "Qubit teleportation between non-neighbouring nodes in a quantum network", Nature, vol. 605, pp. 663–668, 2022.
[26] Bartling, H.P., Abobeih, M.H., Pingault, B., Degen, M.J., Loenen, S.J.H., Bradley, C.E., Randall, J., Markham, M., Twitchen, D.J. and Taminiau, T.H., "Entanglement of spin-pair qubits with intrinsic dephasing times exceeding a minute", Physical Review X, vol. 12, 011048, 2022.
[27] Butler, R.W. and Wood, A.T.A., "Saddlepoint approximation for moment generating functions of truncated random variables", Annals of Statistics, vol. XX, no. XX, pp. XX–XX, 2004.
[28] Le Boudec, J.Y., "Performance evaluation of computer and communication systems", EPFL Press, 2010.
[29] Cabrillo, C., Cirac, J.I., Garcia-Fernandez, P. and Zoller, P., "Creation of entangled states of distant atoms by interference", Physical Review A, vol. 59, no. 2, p. 1025, 1999.
[30] Maiti, S., Kar, S., van Hooft, M. and Wehner, S., "Entanglement Buffers for Heterogeneous Quantum Networks (To be updated)", arXiv preprint, 2026.