On the Stochastic Analysis of a Quantum Entanglement Switch

Gayane Vardoyan
UMass Amherst

Saikat Guha
University of Arizona

Philippe Nain
Inria, France

Don Towsley
UMass Amherst

ABSTRACT

We study a quantum entanglement switch that serves k users in a star topology. We model variants of the system using continuous-time Markov chains (CTMCs) and obtain expressions for switch capacity and the expected number of qubits stored in memory at the switch. Using CTMCs allows us to obtain a number of analytic results for systems in which the links are homogeneous or heterogeneous and for switches that have infinite or finite buffer sizes. In addition, we can easily model the effects of decoherence of quantum states using this technique. From numerical observations, we discover that decoherence has little effect on capacity and expected number of stored qubits for homogeneous systems. For heterogeneous systems, especially those operating close to stability constraints, buffer size and decoherence can significantly affect performance. We also learn that, in general, increasing the buffer size from one to two qubits per link is advantageous to most systems, while increasing the buffer size further yields diminishing returns.

1. INTRODUCTION

Quantum networks can support a variety of distributed computing applications, including coin flipping, leader election, and Byzantine agreement. One of the strongest motivations, however, is quantum key distribution (QKD), of which the two most well-known protocols are E91 [4] and BB84 [1]. The former protocol is intrinsically entanglement-based while the latter was originally formulated based on the idea that one user would encode a bit within a qubit and transmit it to another user over, for example, optical fiber. In QKD, users exploit properties of quantum states to produce a shared secret key that can be used as a one-time pad to encode classical messages. QKD and other applications drive the increasing need for a quantum network that can supply end-to-end entanglements to groups of endpoints that request them [8, 10, 11].

A major challenge of implementation of distributed tasks in quantum networks is the difficulty of error-free transmission of quantum states across large distances. For optical fiber, the channel transmissivity is $\eta = e^{-\gamma L}$, where L is the length of the link and γ the fiber's attenuation coefficient. The probability of successful entanglement generation p on a link is proportional to its transmissivity η. Hence, it decreases exponentially in the length of the link. Transmission through free space poses its own challenges, such as photon loss and phase changes due to scattering [11].

A remedy for this issue is the use of quantum repeaters [3] coupled with the process of teleportation [2]. Quantum teleportation is a method for transferring the state of one qubit to another, possibly across a large distance. The cost of performing one teleportation is exactly one entanglement. To accomplish this process across a larger distance, a quantum repeater is positioned between the users, with a quantum channel connecting each to the repeater. Then, link-level entanglements are created: one between each user and the repeater. The repeater then performs a measurement in the Bell basis on the two locally-held qubits resulting in an end-to-end entanglement between the two users. Now, one of the users can teleport a qubit using this new, longer-distance entanglement. To extend the distance even further, more repeaters may be added, and end-to-end entanglements created using several link-level entanglements. This process is what allows QKD protocols and other distributed quantum algorithms to be of practical use.

In this work, we use the term "quantum switch" instead of "repeater" because we will allow it to connect arbitrary pairs of links. Series of quantum repeaters serving simple applications have been thoroughly studied ([5], [7], [9]), but future quantum networks will need to be able to serve multiple applications and users. Our study of a quantum switch is an initial attempt to understand performance of such applications. To this end, we study a single quantum switch that serves k users in a star topology. Each user has a dedicated link to the switch, and all sets of users of size $n \leq k$, for a fixed n, wish to share an entangled state. To achieve this, link-level entanglements are attempted at a constant rate across each link, and if successful result in two-qubit maximally-entangled states (*i.e.* Bell pairs[1], ebits or EPR states). Because ebit generation is inherently probabilistic, we model it as a Poisson process. Qubits from successfully-generated ebits are stored at local quantum memories: one qubit of a Bell pair at the user and the other at the switch. When enough of these ebits are accrued (at least n of them), the switch performs multi-qubit measurements to provide end-to-end entanglements to user groups of size n. When $n = 2$, the switch uses Bell-state

[1] A Bell pair consists of two entangled qubits, *e.g.* $|\Phi^+\rangle = \frac{|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B}{\sqrt{2}}$, where subscripts A and B signify the two qubits (possibly separated by distance) and $\otimes$ is a tensor product. For Bell pairs, measuring one of them tells us with certainty the state of the other qubit with outcomes 0 and 1 equally likely.

Copyright is held by author/owner(s).

measurements[2] (BSMs) and when $n \geq 3$, it uses n-qubit Greenberger-Horne-Zeilinger (GHZ) basis measurements [6]. The reason we make the assumption that any n users want to share an entangled state is twofold: first, we are interested in the *maximum* achievable capacity of this system, hence it is helpful if all resources can be used up as quickly as possible; and second, a possible application is a scenario in which any n users wish to share a secret key, using for example the E91 protocol. In this case, many ebits are required to create a key of sufficient length, and the end-to-end entanglement demand is higher than the entanglement generation rate.

We consider a scenario in which links may generate entanglements at different rates. The switch can store B qubits (each entangled with another qubit held by a user) per link; we study the effect of B on performance. Throughout this work, we will refer to these pairs of stored qubits as *stored entanglements*. Another factor that impacts performance is decoherence of quantum states; we model it and study its effect. We focus on two metrics, switch capacity C, *i.e.*, the number of end-to-end entanglements served by the switch per time unit, and the expected number of qubits Q in memory at the switch, $E[Q]$. Both C and $E[Q]$ depend on values of k, n, B, entanglement generation and decoherence rates, measurement success probability, and the switching mechanism, including the scheduling policy used by the switch.

The contributions of this work are as follows: using CTMCs, we derive C and $E[Q]$ for $n = 2$ for a particular scheduling policy and study how they change as functions of k, buffer size, and decoherence rate. For $n \geq 3$, we derive C for $B = \infty$ and statistically identical links, under the assumption of no decoherence. From our analysis, we gain valuable insight into which factors most influence capacity. For instance, we learn that for $n = 2$, when links are identical, the number of links and their entanglement generation rate are the most impactful, while decoherence and buffer size have little effect. However, the same is not true in the heterogeneous-link case, where the distribution of entanglement generation rates, combined with finite coherence time, can drastically affect both C and $E[Q]$ for systems that operate close to stability constraints.

2. MODEL ASSUMPTIONS

Since we are interested in obtaining achievability results for C, we assume that the switch utilizes a scheduling policy that enables it to use resources as efficiently as possible. In this policy, the switch adheres to the *Oldest Link Entanglement First (OLEF)* rule, wherein the oldest link-level ebits have priority to be used in entangling measurements. A practical reason for this rule is that quantum states are subject to decoherence, which is a function of time; hence, our goal is to make use of link-level ebits as soon as possible.

Both link-level entanglement generation and entangling measurements can be modeled as probabilistic phenomena [5]. In this work, we assume that link-level entanglements are successfully generated according to a Poisson process with rate μ_l for link l. We assume that measurements performed by the switch succeed with probability q. We also incorporate decoherence by modeling coherence time as an exponential r.v. with mean $1/\alpha$. The state space can be represented by a vector $\boldsymbol{Q}(t) \in \{0, 1, \dots, B\}^k$, where the l-th element corresponds to the number of stored entanglements at link l at time t. Note that since any set of n users always wish to share an entangled state, at most $n-1$ distinct users can have stored entanglements at any time.

[2] A measurement in the Bell basis is an operation that takes as an input a Bell pair and outputs two classical bits. The bits are then used to decode the Bell state.

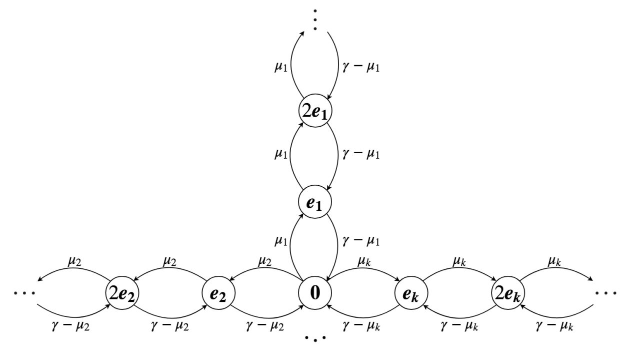

Figure 1: A CTMC model with k users, infinite buffer, and heterogeneous links. μ_l is the entanglement generation rate of link l, while γ is the aggregate entanglement generation rate of all links. $\boldsymbol{e}_l$ is a vector of all zeros except for the l-th position, which is equal to one.

3. RESULTS

Consider first the case for $n = 2$, and assume first that entanglements do not decohere ($\alpha = 0$). Let μ_l be the successful entanglement generation rate of link l. Define the aggregate entanglement generation rate over all links,

$$\gamma := \sum_{l=1}^{k} \mu_l, \quad \text{then let} \quad \rho_l \equiv \frac{\mu_l}{\gamma - \mu_l}, \ \forall l.$$

Note that $\boldsymbol{Q}(t)$ can only take on values of $\mathbf{0}$ or $j\boldsymbol{e}_l$, $l \in \{1, \dots, k\}$, $j \in \{1, 2, \dots\}$, where $\boldsymbol{e}_l$ is a vector of all zeros except for the l-th position, which equals 1. Figure 1 presents the CTMC for the case of heterogeneous links and infinite buffer. Consider state $\mathbf{0}$ (no stored entanglements). From there, a transition along one of the k "arms" of the CTMC occurs with rate μ_l, when the l-th link successfully generates an entanglement. For a BSM to occur, any of the $k-1$ other links must successfully generate an entanglement; this occurs with rate $\gamma - \mu_l$.

Define the following limits when they exist:

$$\pi_0 = \lim_{t \to \infty} P(\boldsymbol{Q}(t) = \mathbf{0}),$$
$$\pi_l^{(j)} = \lim_{t \to \infty} P(\boldsymbol{Q}(t) = j\boldsymbol{e}_l).$$

Once we obtain an expression for π_0, we can derive expressions for capacity and the expected number of stored qubits. The balance equations for the CTMC in Figure 1 are

$$\pi_0 \mu_l = \pi_l^{(1)}(\gamma - \mu_l), \ l \in \{1, \dots, k\},$$
$$\pi_l^{(j-1)} \mu_l = \pi_l^{(j)}(\gamma - \mu_l), \ l \in \{1, \dots, k\}, \ j \in \{2, 3, \dots\},$$
$$\pi_0 + \sum_{l=1}^{k} \sum_{j=1}^{\infty} \pi_l^{(j)} = 1.$$

For the stability of a heterogeneous system with infinite buffer, ρ_l must be strictly less than one for all links l. For a homogeneous system with infinite size buffer, the system is stable as long as there are at least three links (a stationary distribution does not exist for $k = 2$). Next, we present analytical results for C and $E[Q]$ for the case where links are heterogeneous, the rate of decoherence $\alpha > 0$, and each link can store up to B entanglements. Complete results of our analysis, including proofs, are presented in [12]. All other variants of the problem are special cases of the aforementioned; for instance, to obtain analytic expressions for a homogeneous system, simply equate all μ_l. Then,

$$\pi_0 = \left(1 + \sum_{l=1}^{k}\sum_{j=1}^{B}\prod_{i=1}^{j}\frac{\mu_l}{\gamma - \mu_l + i\alpha}\right)^{-1},$$

$$C = q\pi_0\sum_{l=1}^{k}\sum_{j=1}^{B}(\gamma - \mu_l)\prod_{i=1}^{j}\frac{\mu_l}{\gamma - \mu_l + i\alpha},$$

$$E[Q] = \pi_0\sum_{j=1}^{B} j\sum_{l=1}^{k}\prod_{i=1}^{j}\frac{\mu_l}{\gamma - \mu_l + i\alpha}.$$

We also obtain an expression for capacity of a system with homogeneous links, where the switch has infinite buffer, serves n-partite entanglements, and the states do not decohere: $C = q\mu k/n$ whenever the system is stable and $E[|\boldsymbol{Q}|] < \infty$, where $|\boldsymbol{Q}|$ denotes the number of stored qubits across all links in steady state. If in addition $E[|\boldsymbol{Q}|^2] < \infty$, then $E[|\boldsymbol{Q}|] = (n-1)k/(2(k-n))$.

In [12], for $n = 2$, we numerically compare switches with identical links with infinite and finite buffer sizes as the number of links k is varied. Interestingly, the convergence from finite-buffer to infinite-buffer models seems to occur quite rapidly, even for the smallest value of k (3), and the maximum relative difference between the two capacities never exceeds 0.25 (even as μ increases). From this, we conclude that buffer size does not play a major role in determining capacity for homogeneous systems under the OLEF policy and only a small quantum memory is required.

We also compare C and $E[Q]$ for heterogeneous systems with infinite and finite buffer sizes. We vary the number of links in order to observe the speed of convergence of the finite-buffer metrics to their infinite-buffer counterparts. For each value of k, links are split into two classes; links in the first class successfully generate entanglements at rate μ_1 and those in the second class at rate μ_2. We set $\mu_1 = 1.9\mu_2$. Values of μ_1 and μ_2 are chosen in a manner that satisfies the stability condition for heterogeneous systems. For each value of k, the ratio of class 1 to class 2 links is 1/2. For all experiments, $q = 1$ since it only scales capacity. Figure 2 presents C and $E[Q]$ for $k = 3$ and $k = 9$ (more results are presented in [12]). As with the homogeneous-link systems, we observe that the slowest convergence is for smaller values of k and the largest relative difference is for smaller values of B. However, the rate of convergence speeds up quickly as k increases from 3 to 6. With the latter, convergence is already observed for $B < 10$. Meanwhile, when $k = 9$, even for $B = 2$ the difference between finite and infinite buffer metrics is already small. Another interesting observation is that quantum memory usage is large when $k = 3$ but not for larger values of k. This is due to the system operating closer to the stability constraint for $k = 3$ than for larger k.

In [12], we study the effects of decoherence on C and $E[Q]$. In practice, α is expected to be at least an order of magnitude smaller than the link-level entanglement generation rate. For homogeneous systems, we observe that even as

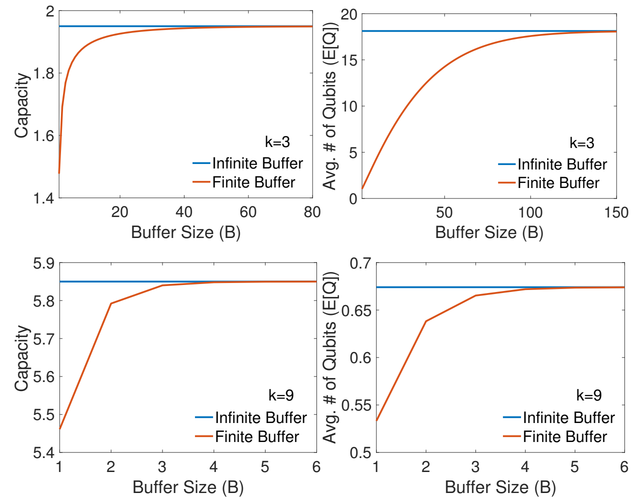

Figure 2: Capacity and expected number of qubits in memory $E[Q]$ for heterogeneous systems with varied number of links and buffer sizes.

α approaches μ decoherence does not cause major degradation in capacity, and likewise does not introduce drastic variations in $E[Q]$. For heterogeneous systems operating close to the stability constraint, however, decoherence and buffer size may have significant effects on performance.

4. REFERENCES

[1] C. H. Bennett and G. Brassard. Quantum Cryptography: Public Key Distribution and Coin Tossing. *Theor. Comput. Sci.*, 560(12):7–11, 2014.

[2] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an Unknown Quantum State Via Dual Classical and Einstein-Podolsky-Rosen Channels. *PRL*, 1993.

[3] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum Repeaters: the Role of Imperfect Local Operations in Quantum Communication. *PRL*, 1998.

[4] A. K. Ekert. Quantum Cryptography Based on Bell's Theorem. *PRL*, 1991.

[5] S. Guha, H. Krovi, C. A. Fuchs, Z. Dutton, J. A. Slater, et al. Rate-loss Analysis of an Efficient Quantum Repeater Architecture. *Phys. Rev. A*, 2014.

[6] M. A. Nielsen and I. Chuang. Quantum Computation and Quantum Information, 2002.

[7] M. Pant, H. Krovi, D. Englund, and S. Guha. Rate-Distance Tradeoff and Resource Costs for All-Optical Quantum Repeaters. *Phys. Rev. A*, 2017.

[8] M. Pant, H. Krovi, D. Towsley, L. Tassiulas, L. Jiang, P. Basu, D. Englund, and S. Guha. Routing Entanglement in the Quantum Internet. 2019.

[9] S. Pirandola. Capacities of Repeater-Assisted Quantum Communications. *arXiv:1601.00966*, 2016.

[10] E. Schoute, L. Mancinska, T. Islam, I. Kerenidis, and S. Wehner. Shortcuts to Quantum Net. Routing. 2016.

[11] R. Van Meter. *Quantum Networking*. Wiley, 2014.

[12] G. Vardoyan, S. Guha, P. Nain, and D. Towsley. On the Stochastic Analysis of a Quantum Entanglement Switch. *arXiv:1903.04420*, 2019.