Entanglement Distribution in the Quantum Internet: an optimal decision problem formulation

Michele Viscardi*†, Jessica Illiano*, Angela Sara Cacciapuoti *, Marcello Caleffi*
* www.QuantumInternet.it research group, Department of Electrical Engineering and Information Technologies
†Department of Physics *"Ettore Pancini"*
University of Naples *Federico II*, Naples, 80125, Italy
michele.viscardi@unina.it, jessica.illiano@unina.it, angelasara.cacciapuoti@unina.it, marcello.caleffi@unina.it

***Abstract*—The entanglement distribution process is widely envisioned as one of the key functionalities of the Quantum Internet. Its engineering is, thus, foundational to effectively implement communication protocols in quantum networks. In this paper, we exploit the Markov Decision Process formalism to model the entanglement distribution as an optimal decision problem. Furthermore, we analyze the impact of reward functions on two key performance metrics, namely: the average distribution time and the average distributed cluster size. From this analysis, we gain some insights for choosing a reward function that meets suitable figures of merit for an overlaying communication protocol.**

***Index Terms*—Entanglement Distribution, Quantum Internet, Quantum Communications, Markov Decision Process**

I. Introduction

The Quantum Internet, i.e., a network interconnecting heterogeneous quantum networks, is foreseen to enable several applications with no counterpart in the classical world [1]–[5], such as distributed quantum computing [6] and secure communications. In this context, the entanglement distribution process plays a key role. Indeed, the successful distribution of entangled states among remote nodes represents a preliminary condition for any entanglement-based communication protocol. Thus, the ultimate goal is to design an entanglement distribution system that is reliable and efficient, i.e. a system engineered with the aim to account for failures and non-idealities. Understanding how quantum communication systems should function, and how to engineer such systems, has been the focus of research efforts in recent years. Specifically, several theoretical models and designs have been proposed [7]–[9].

In most of the aforementioned models, it is common to assume a small set of super-nodes in the network to be in charge of generating and distributing entangled states [10], [11], due to the current technological limitations. In these scenarios, the super-nodes are therefore responsible for avoiding potential bottlenecks and meeting the overlaying communication protocol's requirements. Hence, optimizing the entanglement distribution process becomes imperative. This can be accomplished by modeling the distribution process at a super-node and integrating some design parameters to control its behavior.

Unfortunately, at the moment, literature is still missing a general yet simple framework that accounts for the aforementioned challenges, enabling to efficiently control the distribution process.

In this paper, we move the first step toward modeling entanglement distribution at the super-nodes as an optimal decision problem, providing quantum network designers with a flexible tool to satisfy their communication needs.

A. Related works

The entanglement distribution process has been modeled in a few different ways in the literature. In [12], the authors model the distribution of entangled pairs as a Discrete Time Markov Chain (DTMC). Specifically, they assume infinite coherence time and infinite resources at the central node – referred to as *"switch"* – with the aim of analyzing the expected capacity of the switch in terms of the number of qubits to be stored for meeting the stability condition of the system. In [13], the distribution of entangled pairs is modeled as a Continuous Time Markov Chain (CTMC). Such a model is based on a Poisson probability distribution for the successful distribution of entangled pairs over the single quantum channel and accounts for some non-idealities, such as decoherence and noisy measurements. Recently, in [14] the Markov decision process formalism has been proposed as a model for the entanglement swapping operations within a quantum repeater chain. Specifically, the resulting policy establishes – for each node belonging to the linear repeater chain and for each time step – which operation should be performed among the set: wait, entanglement distribution with the neighbors nodes, entanglement swapping and measurement. Finally, in [15] some practical figures of merit for entanglement distribution in quantum repeater networks are provided. In particular, the authors define the average connection time and the average size of the largest distributed entangled state for a fixed scenario.

Michele Viscardi acknowledges PNRR MUR project CN00000013, Angela Sara Cacciapuoti acknowledges PNRR MUR NQSTI-PE00000023. The work of M. Caleffi was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART" - E63C22002040007).

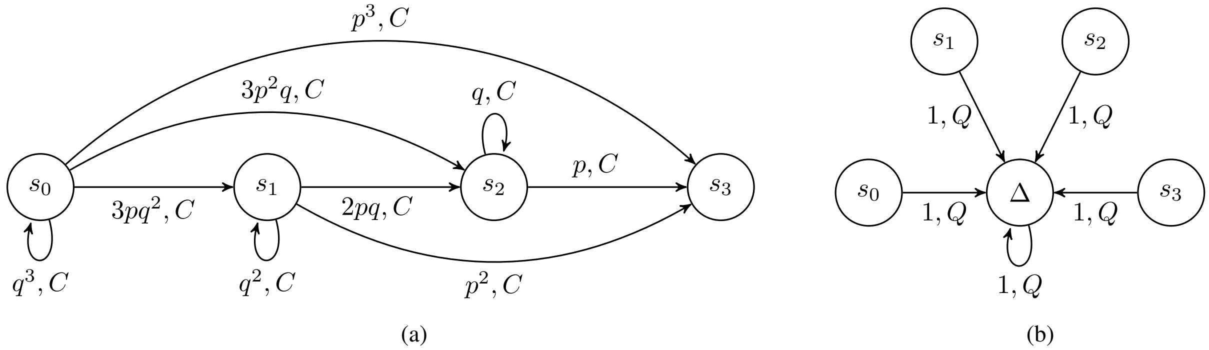


Fig. 1: Representation of the two functioning regimes for a network with $N = 3$ clients: (a): regime of the action C. (b): regime of the action Q.

B. *Our contributions*

Differently from the previously mentioned state-of-the-art proposals, in this work, we focus on engineering the entanglement distribution process, by abstracting from the particular state to be distributed and providing a model that can be tweaked to account for the physical characteristics of the process itself. To this aim:

1) we formulate the entanglement distribution process as a Markov Decision Process (MDP);
2) we analyze the impact of different reward functions on the distribution process through two figures of merit: the average distribution time and the average distributed cluster size;
3) we provide some insights into selecting a suitable reward function for entanglement-based communication protocols.

In short, we provide an easy-to-use tool for modeling and tuning entanglement distribution systems for meeting some performance requirements. However, it is worth noting that the model that we provide in this work is widely flexible, and may be adjusted to be applied in many different scenarios.

II. Preliminaries

In the following, we will focus on the entanglement distribution process from a communication engineering perspective. Specifically, in this section we provide the system model for a generic entanglement distribution system.

A. *System Model*

Without loss of generality, we consider a star network topology, in which a super-node acts as central node and it is responsible for the entanglement distribution process [10], [11]. Specifically, the super-node is in charge of distributing entangled states to N quantum nodes – referred to as clients – through N dedicated quantum channels. During the entanglement distribution process, the super-node and its clients are assumed to interact in a time-slotted fashion.

More into details, we consider the time horizon of the entanglement distribution process constituted by M time slots:

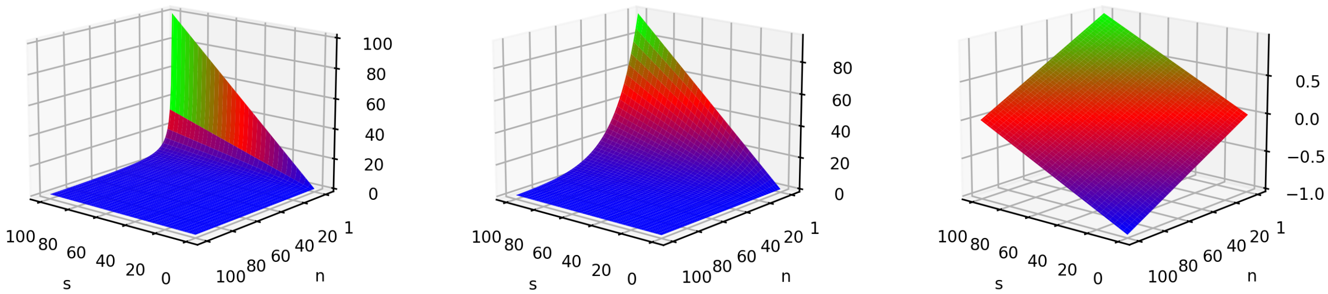
$$T = \{1, 2, \ldots, M\}. \tag{1}$$

with M implicitly accounting for the minimum coherence time. Specifically, the value of M in (1) depends on the particulars of the technology adopted for generating and distributing the entangled states, and it is set such that the decoherence effects can be considered negligible. We also consider N identical and independent quantum channels, where ebit transmissions are assumed to be independent[1] and each quantum channel is assumed to be a quantum absorbing channel. Accordingly, we denote with p the probability of an ebit propagating through a quantum channel without experiencing absorption, and with $q \stackrel{\triangle}{=} 1-p$ the probability of failing an ebit distribution as a consequence of the carrier absorption. Hence, an ebit distribution attempt over a quantum channel can be modeled as a Bernoulli random variable with parameter p.

B. *Problem Statement*

The overall goal is to distribute a multipartite entangled state. Due to the current hardware technology limitations and the existence of different classes of multipartite entanglement (not all characterized by the persistence property), to give generality to the proposal, we assume the super-node to be distributing EPRs to the clients. This allows the super-node to eventually distribute the multipartite entangled state through teleportation [10]. Accounting for Sec. II-A, during the first timeslot the super-node simultaneously transmits N ebits to the N clients, by exploiting an heralded scheme. This enables the super node to recognize which client – if any – experienced an absorption over the channel. In case of absorption, further distributions can be attempted. These require additional time, thus challenging the decoherence constraints as well as impacting the overall distribution rate. Hence, there exists a trade-off between the number of clients that successfully received the ebit – which we refer to as *distributed cluster size* – and the *distribution time*, i.e., the number of time slots in which the distribution process is completed. As a consequence, the problem of optimizing the aforementioned trade-off arises. Solving such a problem is not a trivial task,

[1]Specifically, successive ebit transmission over the same channel and transmissions over different quantum channels are assumed to be independent.



(a) $g(s,n) = \frac{s}{n}$ (b) $g(s,n) = \lambda^n s$, with $\lambda = 0.95$ (c) $g(s,n) = \frac{s}{N} - \frac{n}{M}$

Fig. 2: Reward function $g_n(s)$ plots.

and, indeed, it deeply affects the performance of the overlaying communication protocol.

To this aim, in the following we introduce a theoretical model of the entanglement distribution process that accounts for several key parameters - such as the coherence time, the channel absorption probability and the number of clients - and introduces some design parameters to be tweaked for meeting some communication performance requirements.

III. Optimal Decision Problem

With the discussion of Sec. II in mind, we propose to model the entanglement distribution process as an optimal decision problem by exploiting the Markov Decision Processes (MDPs) formalism [16]. We first describe the model, then we distinguish the fixed parameters from the design parameters. With the former we mean fixed quantities that represent constraints arising from the underlying technology and\or system architecture. With the latter we mean the system degrees of freedom that can be exploited to engineer the entanglement distribution process. By doing so, we are able to understand the roles, the relationships and the impact of different parameters in the decision process by conducting a performance analysis.

A. Mathematical model

In the following formulation, the central node acts as a decision maker, deciding at any given time-slot whether or not to attempt the distribution. From now on, the consequence of the decision, i.e. attempting or not the distribution, will be referred to as *action*.

We model the distribution process through the quintuple:

$$\{S, T, A_{(s,n)}, p(\cdot|s,a), r_n(s,a)\} \tag{2}$$

where:

- S denotes the finite and discrete state space associated to the entanglement distribution process;
- T, as previously described in (1), is the finite and discrete time horizon in which the distribution takes place;
- $A_{(s,n)}$ denotes the set of actions available to the decision-maker – i.e., the super-node – when the system is in state $s \in S$ at the time $n \in T$;
- $p(s_j|s_i,a)$ denotes the transition probability from state s_i to s_j according to action a;
- $r_n(s,a)$ denotes the reward function.

More into details, we refer to $s \in S$ as the state of the system, where:

$$S = S' \cup \{\Delta\} \quad \text{with} \quad S' = \{0,1,2,\ldots,N\} \tag{3}$$

Specifically, $s_i \in S'$ denotes the state where i clients successfully received an ebit and Δ denotes the absorption state, i.e., no further ebits are transmitted.

Hence, the couple (s,n), with $s \in S$ and $n \in T$, fully describes the state of the system.

For any given state $s \in S$ and time-slot $n \in T$, the set of the available actions $A_{(s,n)}$ is:

$$A_{(s,n)} = \begin{cases} \{C,Q\} & s \in S' \setminus \{N\} \wedge n < M \\ \{Q\} & s \in \{N,\Delta\} \vee n = M \end{cases} \tag{4}$$

with C denoting the action of attempting the distribution, and Q denoting the action of not attempting the distribution. Remarkably, the only action available in the state $s = \Delta$ and/or at the time-slot M is Q.

Assuming the system being in the state $s_i \in S$ at the time-slot $n \in T$, when an action $a \in A_{(s_i,n)}$ is performed, the system will evolve into state $s_j \in S$ with probability $p(s_j|s_i,a)$. Accounting for the system model described in Sec. II-A, the transition probability $p(s_j|s_i,a)$ can be expressed as follows:

$$p(s_j|s_i,a) = \begin{cases} p(s_j|s_i) & s_i,s_j \in S' : j \geq i \wedge a = C \\ 1 & s_i \in S, s_j = \Delta \wedge a = Q \\ 0 & s_i,s_j \in S' : j < i \vee s_i = \Delta, s_j \in S' \end{cases} \tag{5}$$

where, in our problem:

$$p(s_j|s_i) = p(s_{i+l}|s_i) = \binom{N-i}{l} q^{N-i-l} p^l \tag{6}$$

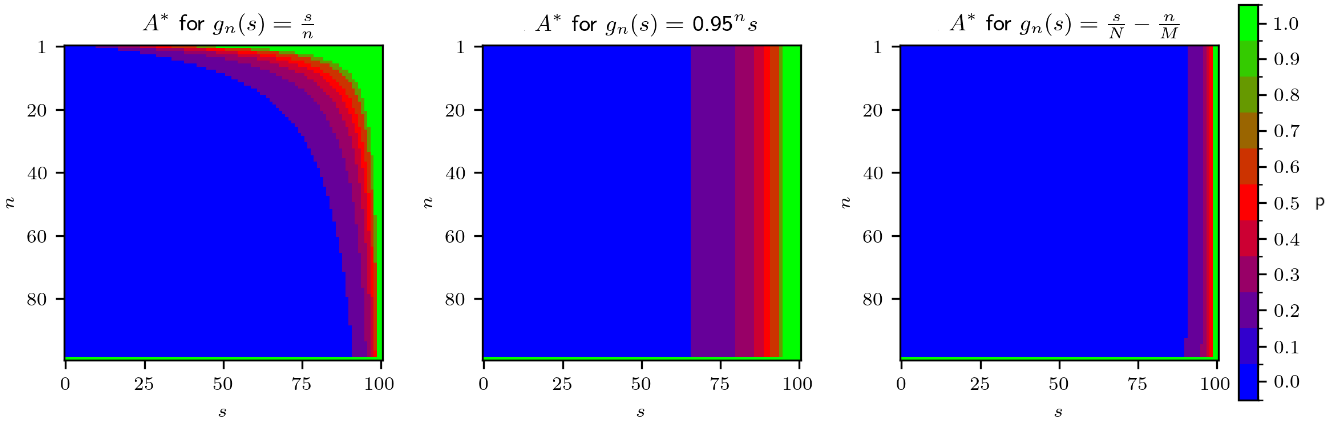


Fig. 3: Compact representation of the action matrices parameterized – according to the color scale on the right – in p with $N = M = 100$, $\lambda = 0.95$, $f_n(s) = 0$, $h(s) = g_n(s)$.

with $j = i + l$ and $0 \leq l \leq N - i$.

As represented in Fig. 1 with reference to a system with $N = 3$ clients, the available actions establish two disjoint functioning regimes for the system, namely, the regime of action C and the regime of action Q. Specifically, Fig. 1a represents the regime of action C. Here, the system evolves according to the transition probabilities $p(s_j|s_i)$ in (6). Remarkably, only the action Q enables the state of the system to change into Δ. Hence, there exist no transition towards the absorbing state through action C. Whereas, Fig. 1b represents the region of action Q. As also expressed in (5), once the super-node decides to perform action Q, the system will only evolve towards the absorbing state Δ, where no further ebit transmission is attempted.

The action to perform is chosen upon a criterion that accounts for a reward function $r_n(s,a)$. In particular, considering the system being in state s at a certain time-slot n, when an action a is chosen and performed, the decision-maker is rewarded with a quantity $r_n(s,a)$.

More precisely, when the system is in a given state s at a given time-slot $n < M$, we define the reward function as:

$$r_n(s,a) = \begin{cases} -f_n(s) & s \in S', a = C \\ g_n(s) & s \in S', a = Q \\ 0 & s = \Delta \end{cases} \tag{7}$$

where:

- $-f_n(s)$ denotes the *cost function*. It represents the cost of attempting the ebit distribution in (s,n);
- $g_n(s)$ denotes the *gain function*. It represents the gain obtained when not attempting the ebit distribution in (s,n).

When $n = M$, regardless of the state of the system, the distribution process ends. We set the final reward, i.e., the value of the reward function at the last available time-slot, as follows:

$$r_M(s,Q) = \begin{cases} h(s) & s \in S' \\ 0 & s = \Delta \end{cases} \tag{8}$$

Specifically, we refer to $h(s)$ as the *boundary reward function*.

B. Design parameters

In the model introduced in Sec. III-A we can easily distinguish the fixed parameters from the design parameters. Specifically, T, S and $p(\cdot|s,a)$ explicitly depend on the characteristics of the system, such as the underlying technology or architecture. Indeed, the time horizon M in which the distribution must take place may depend on the minimum coherence time of the nodes in the network. Also, the state space S depends on the number N of client nodes in the network. Moreover, the transition probabilities $p(\cdot|s,a)$ also result as fixed, since they depend on the particular communication channels in the network.

In contrast, the reward function is not necessarily technologically dependent, and can be chosen to reflect the decision maker preferences and satisfy potential performance metrics.

As an example, in quantum networks, the reward function $r_n(s,a)$ should account for the requirements of the particular communication protocol to be performed, thus orienting the decision process towards the most convenient states for our communication purposes. Such preferences are often reflected in the expression and the properties of the reward function itself.

In the following we assume to enclose the effects of the gain and cost functions of (7) in one and only function $g_n(s)$ which satisfies the following properties:

Property 1 (Monotonicity with s):

$$g_n(s_i) \leq g_n(s_{i+l}) \quad \text{for } 0 \leq l \leq N - i$$

$$\begin{cases} u_n(s_i) = \max\{\sum_{l=0}^{N-i} \binom{N-i}{l} q^{N-i-l} p^l u_{n+1}(s_{i+l}) \ , \ g_n(s_i)\} & ,\forall s_i \in S', \forall n \in T \setminus \{M\} \\ u_M(s_i) = h(s_i) & ,\forall s_i \in S' \end{cases} \tag{12}$$

$g_n(s)$ is a monotonic increasing function of s. Property (1) lets the reward function orient the system toward larger distributed cluster sizes. That is, given the time-slot $n \in T$, the decision maker prefers larger distributed cluster sizes.

Property 2 (Monotonicity with n):

$$g_n(s) \geq g_{n+k}(s) \quad \text{for } 0 \leq k \leq M - n$$

$g_n(s)$ is a monotonic decreasing function of n. Property (2) lets the reward function orient the system toward shorter distribution times. This implies that the cost for longer distribution times is implicitly accounted for within $g_n(s)$, and hence we can set the cost function $f_n(s) = 0$.

Assuming the system being in the state s_i at the time-slot n, we can define the corresponding expected reward $u_n(s_i))$ as:

$$u_n(s_i) = \max_{a \in A_{(s_i,n)}} \{r_n(s_i, a) + \sum_{s_j \in S} p(s_j|s_i, a) u_{n+1}(s_j)\} \tag{9}$$

Then, we set the boundary condition:

$$u_M(s_i) = h(s_i) \quad ,\forall s_i \in S' \tag{10}$$

By recalling the expression of $r_n(s, a)$ in (7) and (8), we can write the optimality system as in E(12). Specifically, the maximum in (12) is chosen between the expected reward corresponding to the action C and the current reward corresponding to the action Q. This is crucial to compute the optimal decision a^* in (s_i, n):

$$a^* = \arg \max_{a \in A_{(s_i,n)}} \{r_n(s_i, a) + \sum_{s_j \in S} p(s_j|s_i, a) u_{n+1}(s_j)\} \tag{11}$$

The solution of (11) for any $(s_i, n) \in S \times T$ is the optimal action matrix A^*, i.e., the $N \times M$ matrix whose element (s_i, n) is the action a^* corresponding to the optimal decision.

IV. Performance evaluation

In the following section, we analyze the impact of the reward function on the system's performance. Specifically, we define two metrics:

- Average distribution time: the average amount of time-slots before the distribution is arrested;
- Average distributed cluster size: the average number of nodes to whom an ebit has been successfully distributed;

and, by performing numerical simulations, we study how they vary in dependence of the reward function. The main objective is to draft some guidelines for choosing a reward function that lets the system meet some performance requirements.

To this aim, we consider three different expressions $g_n(s)$ as our reward functions:

$$g(s, n) = \frac{s}{n} \tag{13}$$

$$g(s, n) = \lambda^n s \tag{14}$$

$$g(s, n) = \frac{s}{N} - \frac{n}{M} \tag{15}$$

where in (14) $\lambda \in (0, 1)$ is an additional parameter acting as a discount factor.
As showed in Fig. 2, the above reward functions meet Property (1) and Property (2). Moreover, every $g(s, n)$ differently relates s and n: sometimes *"valuing"* one parameter more than the other (Fig. 2a, Fig. 2b), sometimes *"valuing"* them equally (Fig. 2c).

A. Simulations and Numerical Results

In order to understand the impact of the reward function on the system's performance, we first computed the action matrix A^* for different values of p. The action matrices are reported in a compact representation in Fig. 3. According to the colormap in Fig. 3, associated with each value $p_k \in \mathcal{P} = \{0, 0.1, \ldots, 0.9, 1\}$ is a color c_k.

Each element[2] $a^*_{(s,n)}$ of the action matrix A^* is colored such that:

$$\text{if } a^*_{(s,n)} \text{ is colored in } c_k \Rightarrow a^{k'}_{(s,n)} = \begin{cases} Q & \forall k' : p_{k'} \leq p_k \\ C & \forall k' : p_{k'} > p_k \end{cases} \tag{16}$$

with $a^{k'}_{(s,n)}$ being the optimal action to take when the system is in (s, n) and the succesful ebit propagation probability p is equal to $p_{k'} \in \mathcal{P}$.

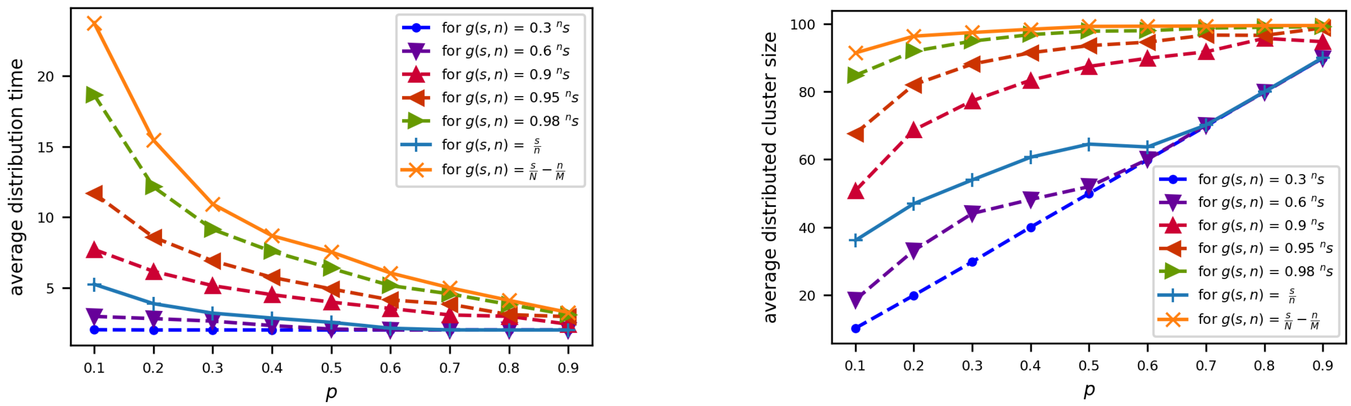
For any given p_k, Fig. 3 shows two functioning regimes[3] in the action matrix A^*. In the regime of action Q, the optimal action is always Q for any state (s, n) within this regime. Similarly, in the regime of action C, the optimal action is always C for any state (s, n) within this regime.

It is worth noting that different reward functions may lead to different action matrices. As an example, the reward function $g_n(s) = \frac{s}{n}$ seems to be accounting more for the distribution time than the others.

As a result, the system may exhibit different performances depending on the reward function. This is showed in Fig. 4, where the average distribution time and the average cluster size are computed for different values of p and for different reward functions $g_n(s)$. Interestingly, the reward functions significantly impact the performances for lower values of p.

[2]The elements $a^*_{(s,n)}$ of the action matrix A^* are computed exploiting backward induction. Therefore, they should be intended as the best action to take in case the system is in state (s, n).

[3]With the term regimes, we mean two contiguous regions of the action matrix.



(a) Average distribution time

(b) Average distributed cluster size

Fig. 4: Plot of entanglement distribution process performance metrics as a function of p.

Indeed, both in Fig. 4a and Fig. 4b, as p increases, the distance between the curves in the graph tends to reduce. Thus, different reward functions result in vastly different ebit distribution performances under bad transmission conditions.

In real communication scenarios, the probability of successful ebit distribution p may be fixed, since it is dependent on the particular communication channel. Moreover, we may be interested in performing a communication protocol with particular performance requirements in terms of average distribution time or average distributed cluster size. In most of these cases, it is possible to meet the performance requirements by choosing a suitable reward function. As an example, we may adjust the value of λ in $g(s,n) = \lambda^n s$ in order to meet some average distribution time requirements[4].

Thus, our formulation of the entanglement distribution process as an optimal decision problem may result handy for quantum network designers whenever some performance requirements must be met.

V. Conclusion

In this work, we provided a formulation of the entanglement distribution process as a Markov Decision Process. Our formal model jointly accounts for the constraints arising from the underlying technologies and the overlaying communication protocol requirements. We exploited this formulation to discuss the trade-off between two performance metrics, i.e., the average distribution time and the average distributed cluster size, and to analyze the impact of the reward function on the entanglement distribution performances. The numerical simulations proved the role of the reward function as a powerful design parameter and showed the flexibility of our model. Indeed, by properly designing the reward function, in most cases, we can let the system meet the aforementioned performance requirements. This analysis paves the way towards the design of robust and efficient entanglement distribution systems for quantum networks and, more broadly, the Quantum Internet.

[4]Some λ values may cause degenerate decisions. For instance, when $\lambda = 0.3$ in Fig. 4, the average distribution time is approximately one, leading to a linear trend in the average distributed cluster size.

References

[1] H. J. Kimble, "The quantum internet," Nature, vol. 453, no. 7198, pp. 1023–1030, 2008.

[2] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum internet: Networking challenges in distributed quantum computing," IEEE Network, vol. 34, no. 1, pp. 137–143, 2020.

[3] W. Dür, R. Lamprecht, and S. Heusler, "Towards a quantum internet," European Journal of Physics, vol. 38, no. 4, p. 043001, 2017.

[4] S. Wehner, D. Elkouss, and R. Hanson, "Quantum Internet: a Vision for the Road Ahead," Science, vol. 362, no. 6412, 2018.

[5] R. V. M. et al., "A quantum internet architecture," 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), pp. 341–352, sep 2022.

[6] M. Caleffi, M. Amoretti, D. Ferrari, D. Cuomo, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, "Distributed quantum computing: a survey," arXiv preprint arXiv:2212.10609, 2022.

[7] J. Miguel-Ramiro, A. Pirker, and W. Dür, "Genuine quantum networks with superposed tasks and addressing," npj Quantum Information, vol. 7, p. 135, 09 2021.

[8] W. Kozlowski, S. Wehner, R. V. Meter, B. Rijsman, A. S. Cacciapuoti, M. Caleffi, and S. Nagayama, "Architectural Principles for a Quantum Internet," RFC 9340, Mar. 2023.

[9] J. Illiano, M. Caleffi, A. Manzalini, and A. S. Cacciapuoti, "Quantum internet protocol stack: a comprehensive survey," Computer Networks, vol. 213, 2022.

[10] G. Avis, F. Rozpedek, and S. Wehner, "Analysis of multipartite entanglement distribution using a central quantum-network node," Phys. Rev. A, vol. 107, p. 012609, Jan 2023.

[11] J. Illiano, M. Viscardi, S. Koudia, M. Caleffi, and A. S. Cacciapuoti, "Quantum internet: from medium access control to entanglement access control," arXiv preprint arXiv:2205.11923, 2022.

[12] G. Vardoyan, S. Guha, P. Nain, and D. Towsley, "On the exact analysis of an idealized quantum switch," ACM SIGMETRICS Performance Evaluation Review, vol. 48, no. 3, pp. 79–80, 2021.

[13] ——, "On the stochastic analysis of a quantum entanglement distribution switch," IEEE Transactions on Quantum Engineering, vol. 2, pp. 1–16, 2021.

[14] Á. G. Iñesta, G. Vardoyan, L. Scavuzzo, and S. Wehner, "Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs," npj Quantum Information, vol. 9, no. 1, may 2023.

[15] S. Khatri, C. T. Matyas, A. U. Siddiqui, and J. P. Dowling, "Practical figures of merit and thresholds for entanglement distribution in quantum networks," Phys. Rev. Research, vol. 1, p. 023032, Sep 2019.

[16] M. L. Puterman, Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014.