Sequential Decision Algorithms for Measurement-Based Impromptu Deployment of a Wireless Relay Network along a Line

Arpan Chattopadhyay, Marceau Coupechoux, and Anurag Kumar

***Abstract*—We are motivated by the need, in some applications, for impromptu or as-you-go deployment of wireless sensor networks. A person walks along a line, starting from a sink node (e.g., a base-station), and proceeds towards a source node (e.g., a sensor) which is at an a priori unknown location. At equally spaced locations, he makes link quality measurements to the previous relay, and deploys relays at some of these locations, with the aim to connect the source to the sink by a multihop wireless path. In this paper, we consider two approaches for impromptu deployment: (i) the deployment agent can only move forward (which we call *a pure as-you-go* approach), and (ii) the deployment agent can make measurements over several consecutive steps before selecting a placement location among them (which we call *an explore-forward* approach). We consider a light traffic regime, and formulate the problem as a Markov decision process, where the trade-off is among the power used by the nodes, the outage probabilities in the links, and the number of relays placed per unit distance. We obtain the structures of the optimal policies for the *pure as-you-go* approach as well as for the *explore-forward* approach. We also consider natural heuristic algorithms, for comparison. Numerical examples show that the explore-forward approach significantly outperforms the pure as-you-go approach. Next, we propose two learning algorithms for the explore-forward approach, based on Stochastic Approximation, which asymptotically converge to the set of optimal policies, without using any knowledge of the radio propagation model. We demonstrate numerically that the learning algorithms can converge (as deployment progresses) to the set of optimal policies reasonably fast and, hence, can be practical, model-free algorithms for deployment over large regions.**

***Index Terms*—Wireless relay placement, measurement based impromptu deployment, as-you-go relay placement, sequential relay placement, Markov decision process, stochastic approximation.**

I. Introduction

There are situations in which a wireless sensor network (WSN) needs to be deployed in an *impromptu or as-you-go* fashion. One such situation is in emergencies, e.g., situational awareness networks deployed by first-responders such as fire-fighters or anti-terrorist squads. As-you-go deployment is also of interest when deploying networks over large terrains, such as forest trails, particularly when the network is temporary and needs to be quickly redeployed in a different part of the forest (e.g., to monitor a moving phenomenon such as groups of wildlife), or when the deployment needs to be stealthy (e.g., to monitor fugitives).

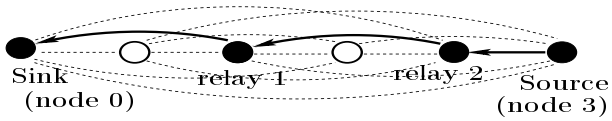

Fig. 1. Two wireless relays (filled dots) deployed along a line to connect a source to a sink by a multihop path. The unfilled dots show other potential relay placement locations, the thin dashed lines indicate all the potential links between the potential placement locations, and the solid lines with arrowheads indicate the links actually used in the deployed network. The distance between two successive potential locations is the step size δ.

Our work in this paper is motivated by the need for as-you-go deployment of wireless relay networks over large terrains, such as forest trails, where planned deployment would be time consuming and difficult. We consider the problem of as-you-go deployment of relay nodes along a line, between a sink node (e.g., the WSN base-station) and a source node (e.g., a sensor) (see Figure 1), where the *single deployment agent* (the person who is carrying out the deployment) starts from the sink node, places relay nodes along the line, and places the source node where required. In applications, the location at which sensor placement is required might only be discovered as the deployment agent walks (e.g., in an animal monitoring application, by finding a concentration of pugmarks, or a watering hole).

In the perspective of an optimal *planned* deployment, we would need to place relay nodes at *all* potential locations (see Figure 1) and measure the qualities of all possible links (between all pairs of potential locations) in order to decide where to place the relays. This approach would provide the global optimal solution, but the time and effort required might not be acceptable in the applications mentioned earlier. With *impromptu* deployment, the next relay placement locations depend on the radio link qualities to the previously placed nodes; these link qualities and also the source location are discovered as the agent walks along the line. Such an approach requires fewer measurements compared to planned deployment, but, in general, is suboptimal.

In this paper, we mathematically formulate the problems of impromptu deployment of relays along a line as optimal

The contents of this paper have been arXived in [1].

Arpan Chattopadhyay and Anurag Kumar are with the Department of ECE, Indian Institute of Science, Bangalore, India; email: arpanc.ju@gmail.com, anurag@ece.iisc.ernet.in.

Marceau Coupechoux is with Telecom ParisTech and CNRS LTCI, Dept. Informatique et Reseaux, 23, avenue d'Italie, 75013 Paris, France; email: marceau.coupechoux@telecom-paristech.fr.

The research reported in this paper was supported by a Department of Electronics and Information Technology (DeitY, India) and NSF (USA) funded project on Wireless Sensor Networks for Protecting Wildlife and Humans, by an Indo-French Centre for Promotion of Advance Research (IFCPAR) funded project, and by the Department of Science and Technology (DST, India), via J.C. Bose Fellowship.

All appendices are provided in the supplementary material.

sequential decision problems. The cost of a deployment is evaluated as a linear combination of three components: the sum transmit power along the path, the sum outage probability along the path, and the number of relays deployed; we provide a motivation for this cost structure. We formulate relay placement problems that minimize the expected average cost per-step. Our channel model accounts for path-loss, shadowing, and fading. We explore deployment with two approaches: (i) the *pure as-you-go* approach and (ii) the *explore-forward* approach. In the pure as-you-go approach, the deployment agent can only move forward; this approach is a necessity if the deployment needs to be quick. Due to shadowing, the path-loss over a link of a given length is random, and a more efficient deployment can be expected if link quality measurements at several locations along the line are compared and an optimal choice is made among these; we call this approach *explore-forward*. Explore-forward would require the deployment agent to retrace his steps; but this might provide a good compromise between deployment speed and deployment efficiency. We formulate each of these problems as a Markov decision process (MDP), obtain the optimal policy structures, illustrate their performance numerically and compare their performance with reasonable heuristics. Next, we propose several learning algorithms and prove that each of them asymptotically converges to the optimal policy if we seek to minimize the average cost per unit distance for deployment over a long line. We also demonstrate the convergence rate of the learning algorithms via numerical exploration.

A. Related Work

In existing literature, problems of impromptu deployment of wireless networks are addressed by heuristics and by experimentation. Howard et al., in [2], provide heuristic algorithms for incremental deployment of sensors in order to cover the deployment area. Souryal et al., in [3], address the problem of impromptu wireless network deployment with experimental study of indoor RF link quality variation; a similar approach is taken in [4] also. The authors of [5] describe a *bread-crumbs* system for aiding firefighting inside buildings. Their work addresses the same class of problems as ours, with the requirement that the deployment agent has to stay connected to k previously placed nodes in the deployment process. Their work considers the trade-off between link qualities and the deployment rate, but does not provide any optimality guarantee of their deployment schemes. Bao and Lee, in [6], study the scenario where a group of first-responders, starting from a command centre, enter a large area where there is no communication infrastructure, and as they walk they place relays at suitable locations in order to stay connected among themselves as well as with the command centre. However, the above described approaches are based on heuristic algorithms, rather than on deriving algorithms from rigorous formulations; hence, these approaches do not provide any provable performance guarantee.

In our work we formulate impromptu deployment as a sequential decision problem, and derive optimal deployment policies. Recently, Sinha et al. ([7]) have provided an algorithm based on an MDP formulation in order to establish a multi-hop network between a sink and an unknown source location, by placing relay nodes along a random lattice path. Their model uses a deterministic mapping between power and wireless link length, and, hence, does not consider the effect of shadowing that leads to statistical variability of the transmit power required to maintain the link quality over links having the same length. The statistical variation of link qualities over space requires measurement-based deployment, in which the deployment agent makes placement decisions at a point based on the measurement of the power required to establish a link (with a given quality) to the previously placed node.

Our previous work on this problem: We view this paper as a continuation of our earlier conference paper [8] which provides the first theoretical formulation of measurement-based impromptu deployment. While the current paper is devoted to problem formulations, derivations of deployment policies and their properties, numerical exploration and comparison of the policies, in another conference paper [9] *we provide results of using the algorithms to carry out actual deployments in a forest-like setting.*

B. Organization

The rest of the paper is organized as follows. The system model and notation have been described in Section II. Impromptu deployment with pure as-you-go approach has been described in Section III. Section IV addresses the problem of impromptu deployment with explore-forward. A numerical comparison between these two approaches are made in Section V. Section VI and Section VII describe the learning algorithms for the explore-forward approach approach. Numerical results are provided in Section VIII on the rate of convergence of the learning algorithms, followed by the conclusion. *All proofs and some discussion are provided in the appendices (i.e., the supplementary material).*

II. System Model and Notation

Throughout this paper, we assume that the line is discretized into steps of length δ (see Figure 1), starting from the sink node. Each point, located at a distance of an integer multiple of δ from the sink node, is considered to be a potential location where a relay can be placed. As the *single* deployment agent walks along the line, at each step or at some subset of steps, he measures the link quality from the current location to the previous node; these measurements are used to decide the location and transmit power of the next relay node.

A. Channel Model and Outage Probability

We consider the usual aspects of path-loss, shadowing, and fading to model the wireless channel. The received power of a packet (say the k-th packet, $k \geq 1$) in a particular link (i.e., a transmitter-receiver pair) of length r is given by:

$$P_{rcv,k} = P_T c \left(\frac{r}{r_0}\right)^{-\eta} H_k W \tag{1}$$

where P_T is the transmit power, c is the path-loss at the reference distance r_0, η is the path-loss exponent, H_k denotes the fading random variable seen by the k-th packet (e.g., it could be an exponentially distributed random variable for

the Rayleigh fading model), and W denotes the shadowing random variable. H_k captures the variation of the received power over time, and it takes independent values over different coherence times.

The path-loss between a transmitter and a receiver at a given distance can have a large spatial variability around the mean path-loss (averaged over fading), as the transmitter is moved over different points at the same distance from the receiver; this is called shadowing.[1] Shadowing is usually modeled as a log-normally distributed, random, multiplicative path-loss factor; in dB, shadowing is distributed with values of standard deviation as large as 8 to 10 dB. Also, shadowing is spatially uncorrelated over distances that depend on the sizes of the objects in the propagation environment (see [10]); *our measurements in a forest-like region of our Indian Institute of Science campus proved Log-normality of shadowing and gave a shadowing decorrelation distance of* 6 *meters (see [9]).* In this paper, W is assumed to take values from a set $\mathcal{W}$. We will denote by $p_W(w)$ the probability mass function or probability density function of W, depending on whether $\mathcal{W}$ is a countable set or an uncountable set as in the case of log-normal shadowing.

A link is considered to be in *outage* if the received signal power drops (due to fading) below $P_{rcv-min}$ (e.g., below -88 dBm, a figure that we have obtained via experimentation for the popular TelosB "motes," see [11]). Since practical radios can only be set to transmit at a finite set of power levels, the transmit power of each node can be chosen from a discrete set, $\mathcal{S} := \{P_1, P_2, \cdots, P_M\}$, where $P_1 \leq P_2 \leq \cdots \leq P_M$. For a link of length r, a transmit power γ and any particular realization of shadowing $W = w$, the outage probability is denoted by $Q_{out}(r, \gamma, w)$, which is increasing in r and decreasing in γ, w (according to (1)).

Note that $Q_{out}(r, \gamma, w)$ depends on the fading statistics. For a link with shadowing realization w, if the transmit power is γ, the received power of a packet will be $P_{rcv} = \gamma c(\frac{r}{r_0})^{-\eta} wH$. Outage is defined to be the event $P_{rcv} \leq P_{rcv-min}$. If H is exponentially distributed with mean 1 (i.e., for Rayleigh fading), then we have, $Q_{out}(r, \gamma, w) = \mathbb{P}\left(\gamma c(\frac{r}{r_0})^{-\eta} wH \leq P_{rcv-min}\right) = 1 - e^{-\frac{P_{rcv-min}(\frac{r}{r_0})^{\eta}}{\gamma c w}}$.

The outage probability of a randomly chosen link of given length and given transmit power is a random variable, where the randomness comes from the spatial variation of link quality due to shadowing. *Outage probability is measured by sending sufficiently large number of packets over a link and calculating the percentage of packets whose RSSI is below* $P_{rcv-min}$.

B. Deployment Process and Related Notation

In this paper, we consider two approaches for deployment.

Pure as-you-go deployment: In this case, after placing a relay, the agent skips the next A steps ($A \geq 0$), and sequentially estimates shadowing from the locations $(A+1), (A+2), \cdots, (A+B)$. As the agent explores the locations $(A+1), (A+2), \cdots, (A+B-1)$ and estimates the shadowing in those locations, at each step he decides whether to place a relay there, and if the decision is to place a relay, then he also decides at what transmit power the relay will operate. In this process, if he has walked $(A+B)$ steps away from the previous relay, or if he encounters the source location within this distance, then he must place a node. □

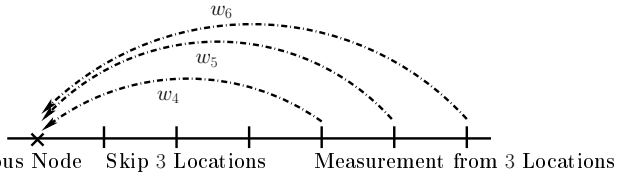

Fig. 2. Explore-forward with $A = 3$ and $B = 3$; the deployment agent skips the first A steps from the previous node and measures the shadowing $w_{A+1}, w_{A+2}, \cdots, w_{A+B}$ from next B locations in order to decide where to place the next relay.

Explore-forward deployment: After placing a node, the deployment agent skips the next A locations ($A \geq 0$) and estimates the shadowing $\underline{w} := (w_{A+1}, w_{A+2}, \cdots, w_{A+B})$[2] to the previous node from locations $(A+1), (A+2), \cdots, (A+B)$. Then he places the relay at one of the locations $(A+1), (A+2), \cdots, (A+B)$ and repeats the same process for placing the next relay. This procedure is illustrated in Figure 2. If the source location is encountered within $(A+B)$ steps from the previous node, then the source is placed. □

We will see later that, in both approaches, it is sufficient to measure the outage probabilities $Q_{out}(r, \gamma, w_r), A+1 \leq r \leq A+B, \gamma \in \mathcal{S}$, *and there is no need to explicitly measure shadowing in the links.*

Choice of A *and* B*:* If the propagation environment is very good, or if we need to place a limited number of relays over a long line, it is very unlikely that a relay will be placed within the first few locations from the previous node. In such cases, link quality measurements from those first few locations are wasted, since shadowing is i.i.d. across links. In such cases, we can skip measurements at locations $1, 2, \cdots, A$ and make measurements from locations $(A+1), (A+2), \cdots, (A+B)$. However, we can simply choose $A = 0$. In general, the choice of A and B will depend on the constraints and requirements for the deployment. A larger value of A will result in faster exploration of the line, since measurements at many locations need not be made. For a fixed A, a larger value of B results in more measurements, and hence we can expect a better performance on an average. However, A and B must be chosen such that the random variable $Q_{out}(A+B, P_M, W)$ is within tolerable limits with high probability; otherwise the deployment agent might measure too many links which have very high outage probability and are not useful.

C. Independence of Shadowing Across Links

As shown in Figure 1, the sink is called Node 0, the relay closest to the sink is called Node 1, and the relays are enumerated as nodes $\{1, 2, 3, \cdots\}$ as we walk away from the

[1] Consider (1). If we transmit a sufficiently large number of packets on a link over multiple coherence times and record the received signal strength of all the packets, we can compute $\overline{P}_{rcv}$ which is the mean received signal power averaged over fading. If the realization of shadowing in that link is w, then $\overline{P}_{rcv} = P_T c(\frac{r}{r_0})^{-\eta} w\mathbb{E}(H)$.

[2] Underlined symbols denote vectors in this paper.

source. The link whose transmitter is Node i and receiver is Node j is called link (i, j). A generic link is denoted by e. Let us recall that the length of each link is an integer multiple of the step size δ.

We assume that the shadowing at any two different links in the network are independent, i.e., $W_{(e_1)}$ is independent of $W_{(e_2)}$ for $e_1 \neq e_2$. Then the independence is a reasonable assumption if δ is chosen to be at least the decorrelation distance (see [10]) of shadowing. For the experimental setting (in the forest inside Indian Institute of Science campus) as in Section II-A, we can safely assume independent shadowing at different potential locations if δ is greater than 6 m (see [9]).

D. Traffic Model

We consider a model where the traffic is so low that there is only one packet in the network at a time; we call this the "lone packet model." As a consequence of this assumption, there are no simultaneous transmissions to cause interference. This permits us to easily write down the communication cost on a path over the deployed relays. Such a traffic model is realistic for sensor networks that carry low duty cycle measurements, or just carry an occasional alarm packet. A design with the lone packet model can be the starting point for a design with desired positive traffic (see [12]). Also, even though the network is designed for the lone packet traffic, it will be able to carry some amount of positive traffic from the source to the sink (see [9] for experimental evidence of this claim; a five-hop line network deployed (using the methodology derived in this paper) over a 500 m long trail in a forest-like environment, was able to carry 127 byte packets at a rate of 4 packets per second, with end-to-end packet loss probability less than 1%).

E. Network Cost Structure

In this section we develop the cost that we use to evaluate the performance of a given deployment policy. Given the current location of the deployment agent with respect to the previous relay, and given the measurements made to the previous relay, a policy will provide the placement decision (in the case of as-you-go deployment, whether or not to place the relay, and if place then at what power, and in the case of explore-forward deployment, where among the B locations to place the relay and at which power). Formal definition of a policy will be given later in this paper.

Let us denote the number of placed relays up to x steps (i.e., $x\delta$ meters) from the sink by N_x ($\leq x$); define $N_0 = 0$. Since deployment decisions are based on measurements to already placed relays, and since the path-loss over a link is a random variable (due to shadowing), we see that $\{N_x\}_{x\geq 1}$ is a random process. In this paper we have assumed that each node forwards each packet to the immediately previously placed relay (e.g., with reference to Figure 1, the source forwards all packets to Relay 2, which, in turn, forwards all packets to Relay 1, etc.). See [8] for the considerably more complex possibility of relay skipping while forwarding packets.

When the node i is placed, the deployment policy also prescribes the transmit power that this node should use, say, Γ_i; then the outage probability over the link $(i, i-1)$, so created, is denoted by $Q_{out}^{(i,i-1)}$. We evaluate the cost of the deployed network, up to $x\delta$ steps, as a linear combination of three cost measures:

(i) The number of relays placed, i.e., N_x.
(ii) The sum outage, i.e., $\sum_{i=1}^{N_x} Q_{out}^{(i,i-1)}$. The motivation for this measure is that, for small values of Q_{out}, the sum-outage is approximately the probability that a packet sent from the point x to the source encounters an outage along the path from the point x back to the sink.
(iii) The sum power over the hops, i.e., $\sum_{i=1}^{N_x} \Gamma_i$. This is a measure of the energy required to operate the network (see discussion later in this section).

These three costs are combined into one cost measure by combining them linearly and taking expectation (under a policy π), as follows:

$$\mathbb{E}_\pi\Big(\sum_{i=1}^{N_x} \Gamma_i + \xi_{out} \sum_{i=1}^{N_x} Q_{out}^{(i,i-1)} + \xi_{relay} N_x\Big) \tag{2}$$

The multipliers $\xi_{out} \geq 0$ and $\xi_{relay} \geq 0$ can be viewed as capturing the emphasis we wish to place on the corresponding measure of cost. For example, a large value of ξ_{out} will aim for a network deployment with smaller end-to-end expected outage. We can view ξ_{relay} as the cost of placing a relay. More formally, these cost multipliers also emerge as "Lagrange" multipliers if we formulate the problem of minimizing the energy cost subject to constraints on the other two costs. We will formalize this in Section II-F.

A Motivation for the Sum Power Objective: In case all the nodes have *wake-on radios,* the nodes normally stay in sleep mode, and each sleeping node draws a very small current from the battery (see [13]). When a node has a packet, it sends a wake-up tone to the intended receiver. The receiver wakes up and the sender transmits the packet. The receiver sends an ACK packet in reply. Clearly, the energy spent in transmission and reception of data packets governs the lifetime of a node, given that the ACK size is negligible compared to the packet size. We assume that a fixed modulation scheme is used, so that the transmission bit rate over all links is the same (e.g., in IEEE 802.15.4 radios, that are commonly used for sensor networking, the standard modulation scheme provides a bit rate of 250 Kbps). We also assume a fixed packet length.

Let t_p be the transmission duration of a packet over a link, and suppose that the node i ($1 \leq i \leq N_x$) uses power Γ_i during transmission. Let P_r denote the packet reception power expended in the electronics at any receiving node. If the packet generation rate ζ at the source is very small, the lifetime of the k-th node ($1 \leq k \leq N_x$) is $T_k := \frac{E}{\zeta(\Gamma_k + P_r)t_p}$ seconds (E is the total energy in a fresh battery). Hence, the rate at which we have to replace the batteries in the network from the sink up to distance x steps is given by $\sum_{k=1}^{N_x} \frac{1}{T_k} = \sum_{k=1}^{N_x} \frac{\zeta(\Gamma_k + P_r)t_p}{E}$. The term $\frac{\zeta P_r t_p}{E}$ can be absorbed into ξ_{relay}. Hence, the battery depletion rate between the sink and the point x is proportional to $\sum_{k=1}^{N_x} \Gamma_k$. Note that, $\sum_{k=1}^{N_x} \Gamma_k$ is the total transmit power to send a packet from node N_x to the sink node, since there is no collision among packets transmitted from various nodes (due to *lone packet* traffic; see Section II-D).

F. Deployment Objective

We assume in this paper that the the distance L to the source from the sink (at the start of the line) is a priori unknown, and no knowledge about its distribution is available. Hence, we assume that $L = \infty$ and use deployment policies that seek to minimize the average cost per step. *This setting can also be useful when L is large (e.g., a long forest trail), or when we seek to deploy relays along various trails (the trails might be interconnected among themselves). Once deployed, such a chain of nodes can be used to realize and connect several source-sink pairs, or even each node could act as a sensor and a relay.*[3]

1) The Unconstrained Problem: Motivated by the cost structure in (2) and the $L = \infty$ model, we seek to solve the following problem:

$$\inf_{\pi\in\Pi} \limsup_{x\to\infty} \frac{\mathbb{E}_\pi \sum_{i=1}^{N_x} (\Gamma_i + \xi_{out} Q_{out}^{(i,i-1)} + \xi_{relay})}{x} \tag{3}$$

where π is a placement *policy* (i.e., deployment strategy), and Π is the set of all possible placement policies (to be formalized later). We formulate (3) as a long-term average cost Markov decision process (MDP).

2) Connection to a Constrained Problem: Note that, (3) is the relaxed version of the following constrained problem where we seek to minimize the mean power per step subject to a constraint on the mean outage per step and a constraint on the mean number of relays per step:

$$\begin{aligned} &\inf_{\pi\in\Pi} \limsup_{x\to\infty} \frac{\mathbb{E}_\pi \sum_{i=1}^{N_x} \Gamma_i}{x} \\ s.t. \quad &\limsup_{x\to\infty} \frac{\mathbb{E}_\pi \sum_{i=1}^{N_x} Q_{out}^{(i,i-1)}}{x} \le \overline{q} \text{ and } \limsup_{x\to\infty} \frac{\mathbb{E}_\pi N_x}{x} \le \overline{N} \end{aligned} \tag{4}$$

The following standard result tells us how to choose the *Lagrange multipliers* ξ_{out} and ξ_{relay} (see [14], Theorem 4.3):

Theorem 1: Consider the constrained problem (4). If there exists a pair $\xi_{out}^* \ge 0$, $\xi_{relay}^* \ge 0$ and a policy π^* such that π^* is the optimal policy of the unconstrained problem (3) under $(\xi_{out}^*, \xi_{relay}^*)$ and the constraints in (4) are met with equality under π^*, then π^* is an optimal policy for (4) also. □

III. Pure As-You-Go Deployment

A. Markov Decision Process (MDP) Formulation

Here we seek to solve problem (3), for the pure as-you-go approach. When the agent is r steps away from the previous node ($A + 1 \le r \le A + B$), he measures the shadowing w on the link from the current location to the previous node. He uses the knowledge of (r, w) to decide whether to place a node at his current location, and what transmit power $\gamma \in \mathcal{S}$ to use if he places a relay. In this case, we formulate the impromptu deployment problem as a Markov Decision Process (MDP) with state space $\{A + 1, A + 2, \cdots, A + B\} \times \mathcal{W}$. At state $(r, w), (A + 1) \le r \le (A + B - 1), w \in \mathcal{W}$, the action is either to place a relay and select a transmit power, or not to place. When $r = A + B$, the only feasible action is to place and select a transmit power $\gamma \in \mathcal{S}$. If, at state (r, w), a relay is placed and it is set to use transmit power γ, a hop-cost of $\gamma + \xi_{out} Q_{out}(r, \gamma, w) + \xi_{relay}$ is incurred.

A deterministic Markov policy π is a sequence of mappings $\{\mu_k\}_{k\ge1}$ from the state space to the action space, and it is called a stationary policy if $\mu_k = \mu$ for all k. Given the state (i.e., the measurements), the placement decision is made according to the policy.

B. Formulation for $L \sim Geometric(\theta)$

Under the pure as-you-go approach, we will first minimize the expected total cost for $L \sim Geometric(\theta)$, and then take $\theta \to 0$; this approach provides the policy structure for the average cost problem (see [15], Chapter 4).

In the $L \sim Geometric(\theta)$ case, the deployment process regenerates (probabilistically) after placing a relay, because of the memoryless property of the geometric distribution, and because of the fact that deployment of a new node will involve measurement of qualities of new links not measured before, and the new links have i.i.d. shadowing independent of the previously measured links. The state of the system at such regeneration points is denoted by $\mathbf{0}$ (also, there are states of the form (r, w)). When the source is placed at the end of the line, the process terminates. Suppose N is the (random) number of relays placed, and node $N + 1$ is the source node (as shown in Figure 1). We first seek to solve the following:

$$\min_{\pi\in\Pi} \mathbb{E}_\pi \left(\sum_{i=1}^{N+1} \Gamma_i + \xi_{out} \sum_{i=1}^{N+1} Q_{out}^{(i,i-1)} + \xi_{relay} N \right) \tag{5}$$

We will first investigate this approach assuming finite $\mathcal{W}$, and later generalize it for the case when $\mathcal{W}$ is a Borel subset of the positive real line.

C. Bellman Equation

Let us denote the optimal expected cost-to-go at state (r, w) and at state $\mathbf{0}$ be $J(r, w)$ and $J(\mathbf{0})$ respectively. Note that here we have an infinite horizon total cost MDP with a finite state space and finite action space. The assumption P of Chapter 3 in [15] is satisfied, since the single-stage costs are nonnegative. Hence, by the theory developed in [15], we can restrict ourselves to the class of stationary deterministic Markov policies.

By Proposition 3.1.1 of [15], the optimal value function $J(\cdot)$ satisfies the Bellman equation which is given by, for all $(A + 1) \le r \le (A + B - 1)$ (explanation follows after the equations),

[3] In [8], we considered the scenario where L is unknown, but there is prior information (e.g., the mean $\overline{L}$) on its distribution. This led us to model L as a geometrically distributed number of steps and minimize the expected total cost of the network. The step length δ and the mean $\overline{L}$, can be used to obtain the parameter of the geometric distribution, i.e., the probability θ that the line ends at the next step. In the current paper, we consider the case $L \sim Geometric(\theta)$ (in steps) only for pure as-you-go case, the reason being to exploit the connection between this model and the $L = \infty$ model.

$$
\begin{aligned}
J(r,w) &= \min\Big\{ \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w))+\xi_{relay}+J(\mathbf{0}),\\
&\quad \theta\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r+1,\gamma,W))\\
&\quad +(1-\theta)\mathbb{E}_W J(r+1,W)\Big\},\\
J(A+B,w) &= \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(A+B,\gamma,w)+\xi_{relay})+J(\mathbf{0})\\
J(\mathbf{0}) &= \sum_{k=1}^{A+1}(1-\theta)^{k-1}\theta\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(k,\gamma,W))\\
&\quad +(1-\theta)^{A+1}\mathbb{E}_W J(A+1,W) \qquad (6)
\end{aligned}
$$

These equations are understood as follows. If the current state is $(r,w), (A+1) \le r \le (A+B-1)$ and the line has not ended yet, we can either place a relay and set its transmit power to $\gamma \in \mathcal{S}$, or we may not place. If we place, the cost $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w)+\xi_{relay})$ is incurred at the current step, and the cost-to-go from there is $J(\mathbf{0})$. If we do not place a relay, the line will end with probability θ in the next step, in which case a cost $\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r+1,\gamma,W))$ will be incurred. If the line does not end in the next step, the next state will be a random state $(r+1,W)$ and a mean cost of $\mathbb{E}_W J(r+1,W)$ will be incurred. At state $(A+B,w)$ the only possible decision is to place a relay. At state $\mathbf{0}$, the deployment agent starts walking until he encounters the source location or location $(A+1)$; if the line ends at step $k, 1 \le k \le A+1$ (with probability $(1-\theta)^{k-1}\theta$), a cost of $\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(k,\gamma,W))$ is incurred. If the line does not end within $(A+1)$ steps (this event has probability $(1-\theta)^{A+1}$), the next state will be $(A+1,W)$.

D. Value Iteration

The value iteration for (5) is obtained by replacing $J(\cdot)$ in (6) by $J^{(k+1)}(\cdot)$ on the L.H.S (left hand side) and by $J^{(k)}(\cdot)$ on the R.H.S (right hand side), and by taking $J^{(0)}(\cdot)=0$ for all states. The standard MDP theory says that $J^{(k)}(\cdot) \uparrow J(\cdot)$ for all states as $k \to \infty$.

E. Policy Structure: OptAsYouGo Algorithm

Lemma 1: $J(r,w)$ is increasing in r, ξ_{out} and ξ_{relay}, decreasing in w, and jointly concave in ξ_{out} and ξ_{relay}. $J(\mathbf{0})$ is increasing and jointly concave in ξ_{out} and ξ_{relay}.

Proof: See Appendix A. ∎

Next, we propose an optimal algorithm for impromptu deployment under the pure as-you-go approach.

Algorithm 1: (OptAsYouGo Algorithm) At state (r,w) (where $A+1 \le r \le A+B-1$), place a relay if and only if $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w)) \le c_{th}(r)$ where $c_{th}(r)$ is a threshold increasing in r. If the decision is to place a relay, the optimal power to be selected is given by $\operatorname{argmin}_{\gamma\in\mathcal{S}}\Big(\gamma+\xi_{out}Q_{out}(r,\gamma,w)\Big)$. At state $(A+B,w)$, select the transmit power $\operatorname{argmin}_{\gamma\in\mathcal{S}}\Big(\gamma+\xi_{out}Q_{out}(A+B,\gamma,w)\Big)$. □

Theorem 2: Under the pure as-you-go approach, Algorithm 1 provides the optimal policy for Problem (3).

Proof: See Appendix A. ∎

From now on, we will call Algorithm 1 as OptAsYouGo (Optimal algorithm with pure As-You-Go approach).[4]

Remark: Note that, in order to make a placement decision, one need not explicitly measure the shadowing w in a given link; measuring the outage probabilities at each transmit power level $\gamma \in \mathcal{S}$ for a given link will suffice to make the decision. In fact, we have taken (r,w) as a typical state for simplicity of representation; so long as the channel model given by (1) is valid, we can take $(r, \{Q_{out}(r,\gamma,w)\}_{\gamma\in\mathcal{S}})$ as a typical state.

Remark: The trade-off in the impromptu deployment problem is that if we place relays far apart, the cost due to outage increases, but the cost of placing the relays decreases. The intuition behind the threshold structure of the policy is that if at distance r we get a good link with the combination of power and outage less than a threshold, then we should accept that link because moving forward is unlikely to yield a better link. $c_{th}(r)$ is increasing in r. Since $Q_{out}(r,\gamma,w)$ is increasing in r for any γ, w, and since shadowing is i.i.d across links, the probability of a link (to the previous node) having desired QoS decreases as we move away from the previous node. Hence, the optimal policy will try to place relays as soon as possible if r is large, and this explains why $c_{th}(r)$ is increasing in r. Note that the threshold $c_{th}(r)$ does not depend on w, due to the fact that shadowing is i.i.d. across links.

F. Computation of the Optimal Policy

Let us write $V(r) := \mathbb{E}_W J(r,W) = \sum_{w\in\mathcal{W}} p_W(w) J(r,w)$, and $V(\mathbf{0}) := J(\mathbf{0})$. Also, for each stage $k \ge 0$ of the value iteration, define $V^{(k)}(r) := \mathbb{E}_W J^{(k)}(r,W)$ and $V^{(k)}(\mathbf{0}) := J^{(k)}(\mathbf{0})$. Multiplying both sides of the value iteration by $p_W(w)$ and summing over $w \in \mathcal{W}$, we obtain an iteration in terms of $V^{(k)}(\cdot)$ and this iteration does not involve $J^{(k)}(\cdot)$. Since $J^{(k)}(r,w) \uparrow J(r,w)$ for each r, w and $J^{(k)}(\mathbf{0}) \uparrow J(\mathbf{0})$ as $k \uparrow \infty$, we can argue that $V^{(k)}(r) \uparrow \mathbb{E}_W J(r,W) = V(r)$ for all r (by Monotone Convergence Theorem) and $V^{(k)}(\mathbf{0}) \uparrow J(\mathbf{0}) = V(\mathbf{0})$. Then we can compute $c_{th}(r)$ by knowing $V(\cdot)$ itself (see the expression of $c_{th}(r)$ in the proof of Theorem 2); we need not keep track of the cost-to-go values $J^{(k)}(r,w)$ for each state (r,w), at each stage k. Here we simply need to keep track of $V^{(k)}(\cdot)$.

Similar iterations were proposed in our prior work [8] for a slightly different model; please see [8], Section III-A-5 for a detailed derivation.

G. Average Cost Problem: Optimality of OptAsYouGo

Note that the problem (5) can be considered as an infinite horizon discounted cost problem with discount factor $(1-\theta)$. Hence, keeping in mind that we have finite state and action spaces, we observe that for the discount factor sufficiently close to 1, i.e., for θ sufficiently close to 0, *the optimal*

[4] Similar approach as in this paper can be used to analyze the case where the length of the line is constant and known. The only difference will be that the optimal policy will be nonstationary.

policy for problem (5) is optimal for the problem (3) (see [15], Proposition 4.1.7). In particular, the optimal average cost per step with pure as-you-go approach, λ^*, is given by $\lambda^* = \lim_{\theta \to 0} \theta J_\theta(\mathbf{0})$ (see [15], Section 4.1.1), where $J_\theta(\mathbf{0})$ is the optimal cost for problem (5) under pure as-you-go with the probability of the line ending in the next step is θ.

Now suppose that $\mathcal{W}$ is a Borel subset of the real line. In this case, we still have a finite action space, and bounded, nonnegative cost per step. We can still write the Bellman Equation (6) for the case $L \sim Geometric(\theta)$. We see that $0 \leq J(A+B,w) - J(\mathbf{0}) \leq P_M + \xi_{out} + \xi_{relay}$. Now, by using the fact that $0 \leq \theta J(\mathbf{0}) \leq P_M + \xi_{out} + \xi_{relay}$, we can prove by induction that $|J(r,w) - J(\mathbf{0})|$ is uniformly bounded across $\theta \in (0,1)$, $r \in \{A+1, A+2, \cdots, A+B\}$, $w \in \mathcal{W}$ and it is also equicontinuous in w for all $\theta \in (0,1)$. Hence, by Theorem 5.5.4 of [16], the optimal average cost per step is again $\lambda^* = \lim_{\theta \to 0} \theta J_\theta(\mathbf{0})$. As $\theta \downarrow 0$, we will obtain a sequence of optimal policies (i.e., mappings from the state space to the action space), and a limit point of them will be an average cost optimal policy.

H. HeuAsYouGo: A Suboptimal Pure As-You-Go Heuristic

This is a modified version of the deployment algorithm proposed in [3]. The algorithm is just a natural heuristic; it has not been derived from any sequential optimization formulation.

Algorithm 2: (HeuAsYouGo) The power used by the relays is set to a fixed value. At each potential location, the deployment agent checks whether the outage to the previous relay meets a certain predetermined target with this fixed transmit power level. After placing a relay, the next relay is placed at the last location where the target outage is met; or place at the $(A+1)$-st location (after the previously placed relay) in the unlikely situation where the target outage is violated in the $(A+1)$-st location itself. If the agent reaches the $(A+B)$-th step and if all previous locations violate the outage target, he must place the next relay at step $(A+B)$. □

This algorithm requires the deployment agent to move back by one step and place in case the outage target is violated for the first time in $(A+2)$-nd step or beyond.

IV. Explore Forward Deployment

A. Semi-Markov Decision Process (SMDP) Formulation

Here we seek to solve the unconstrained problem (3). We formulate our problem as a Semi-Markov Decision Process (SMDP) with state space $\mathcal{W}^B$ and action space $\{A+1, A+2, \cdots, A+B\} \times \mathcal{S}$. The vector $\underline{w} := (w_{A+1}, w_{A+2}, \cdots, w_{A+B})$, i.e., the shadowing from B locations, is the state in our SMDP. In the state $\underline{w}$, an action $(u,\gamma) \in \{A+1, A+2, \cdots, A+B\} \times \mathcal{S}$ is taken where u is the distance of the next relay (from the previous relay) that would be placed and γ is the transmit power that this relay will use. In this case, a hop-cost of $\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay}$ is incurred. After placing a relay, the next state becomes $\underline{w}' := (w'_{A+1}, w'_{A+2}, \cdots, w'_{A+B})$ with probability $g(\underline{w}') := \prod_{r=A+1}^{A+B} p_{W_r}(w'_r)$ (since shadowing is i.i.d. across links).

Let us denote, by the vector-valued random variable $\underline{W}(k)$, the (random) state at the k-th decision instant, and by $\mu_k(\underline{W}(k))$ the action at the k-th decision instant. For a deterministic Markov policy $\{\mu_k\}_{k \geq 1}$, let us define the functions $\mu_k^{(1)} : \mathcal{W}^B \to \{A+1, A+2, \cdots, A+B\}$ and $\mu_k^{(2)} : \mathcal{W}^B \to \mathcal{S}$ as follows: if $\mu_k(\underline{w}) = (u,\gamma)$, then $\mu_k^{(1)}(\underline{w}) = u$ and $\mu_k^{(2)}(\underline{w}) = \gamma$.

B. Policy Structure: Algorithm OptExploreLim

Note that, under any policy, $\underline{W}(k)$ is i.i.d across $k, k \geq 1$. The state space is a Borel space and the action space is finite. The hop cost and hop length (in number of steps) are uniformly bounded across all state-action pairs. Hence, we can work with stationary deterministic policies (see [17] for finite state space, i.e., finite $\mathcal{W}$, and [18] for a general Borel state space, i.e., when $\mathcal{W}$ is a Borel set). Under our current scenario, the optimal average cost per step, λ^*, exists (in fact, the limit exists) and is same for all states, i.e. for all $\underline{w} \in \mathcal{W}^B$. *For simplicity, we work with finite $\mathcal{W}$ in this section, but the policy structure holds for Borel state space also.*

We next present a deployment algorithm called "OptExploreLim," an optimal algorithm for limited exploration.

Algorithm 3: (OptExploreLim Algorithm:) In the state $\underline{w}$ which is captured by the measurements $\{Q_{out}(u,\gamma,w_u)\}$ for $A+1 \leq u \leq A+B$, $\gamma \in \mathcal{S}$, place the new relay according to the policy μ^* (later we will also use the notation π^* or $\pi^*(\xi_{out}, \xi_{relay})$ to denote the same policy) as follows:

$$\mu^*(\underline{w}) = \underset{u,\gamma}{\operatorname{argmin}} \left(\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda^* u \right) \tag{7}$$

where λ^* (or $\lambda^*(\xi_{out}, \xi_{relay})$) is the optimal average cost per step for the Lagrange multipliers (ξ_{out}, ξ_{relay}). □

Theorem 3: The policy μ^* given by Algorithm 3 is optimal for the problem (3) under the explore-forward approach.

Proof: The optimality equation for the SMDP is given by (see [17], Equation 7.2.2):

$$\begin{aligned} v^*(\underline{w}) \quad = \quad & \min_{u,\gamma} \Big\{ \gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} \\ & -\lambda^* u + \sum_{\underline{w}' \in \mathcal{W}^B} g(\underline{w}') v^*(\underline{w}') \Big\} \end{aligned} \tag{8}$$

$v^*(\underline{w})$ is the optimal differential cost corresponding to state $\underline{w}$. The structure of the optimal policy is obvious from (8), since $\sum_{\underline{w}' \in \mathcal{W}^B} g(\underline{w}') v^*(\underline{w}')$ does not depend on (u,γ) (note that, $v^*(\underline{w})$ in (8) is obtained after taking minimum over (u,γ)). ■

Later we will also use the notation $\pi^*(\xi_{out}, \xi_{relay})$ to denote the OptExploreLim policy under the pair (ξ_{out}, ξ_{relay}).

Remark 1: Same optimality equation and optimal policy structure will hold when we have a Borel state space (e.g., for Log-Normal shadowing), by the theory presented in [18].

Remark 2: Note that, the optimal policy depends on the state $\underline{w}$ only via the outage probabilities which can be easily measured by the agent.

Remark 3: If we take an action (u,γ), a cost $(\gamma + \xi_{out}Q_{out}(u,\gamma,w_u) + \xi_{relay})$ will be incurred. On the other hand, if we incur a cost of λ^* over each one of those u steps, the total cost incurred will be $\lambda^* u$. The policy selects the placement point that minimizes the difference between these two. Note that, the deployment process regenerates at each placement point (due to i.i.d shadowing across links).

Remark 4: Also, note that, the policy requires the deployment agent to know λ^*. But computation of λ^* will require perfect knowledge of propagation environment (e.g., the path-loss exponent η in (1), the distribution of shadowing in a link, etc.); see Section IV-C. Later we will propose two learning algorithms in Section VI and Section VII, which will not require such knowledge of the propagation environment.

Theorem 4: The optimal average cost per step $\lambda^*(\xi_{out},\xi_{relay})$ is jointly concave, increasing and continuous in ξ_{out} and ξ_{relay}.

Proof: See Appendix B. ∎

Let us consider a sub-class of stationary deployment policies (parameterized by $\lambda \geq 0$, $\xi_{out} \geq 0$ and $\xi_{relay} \geq 0$) given by:

$$\mu(\underline{w}) = \underset{u,\gamma}{\operatorname{argmin}} \left(\gamma + \xi_{out}Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda u \right) \tag{9}$$

where λ is not necessarily equal to $\lambda^*(\xi_{out},\xi_{relay})$.

Under the class of policies given by (9), let $(U_k,\Gamma_k,Q_{out}^{(k,k-1)}), k \geq 1$, denote the sequence of inter-node distances, transmit powers and link outage probabilities that the optimal policy yields during the deployment process. By the assumption of i.i.d. shadowing across links, it follows that $(U_k,\Gamma_k,Q_{out}^{(k,k-1)}), k \geq 1$, is an i.i.d. sequence.

Let $\overline{\Gamma}(\lambda,\xi_{out},\xi_{relay})$, $\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})$ and $\overline{U}(\lambda,\xi_{out},\xi_{relay})$ denote the mean power per link, mean outage per link and mean placement distance (in steps) respectively, under the policy given by (9), where λ is not necessarily equal to $\lambda^*(\xi_{out},\xi_{relay})$. Also, let $\overline{\Gamma}^*(\xi_{out},\xi_{relay})$, $\overline{Q}^*_{out}(\xi_{out},\xi_{relay})$ and $\overline{U}^*(\xi_{out},\xi_{relay})$ denote the optimal mean power per link, the optimal mean outage per link and the optimal mean placement distance (in steps) respectively, under the OptExploreLim algorithm (i.e., policy $\pi^*(\xi_{out},\xi_{relay})$ when λ in (9) is replaced by $\lambda^*(\xi_{out},\xi_{relay})$). By the Renewal-Reward theorem, the optimal mean power per step, the optimal mean outage per step, and the optimal mean number of relays per step are given by $\frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$, $\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$ and $\frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}$.

Theorem 5: For a given ξ_{out}, the mean number of relays per step under the OptExploreLim algorithm (Algorithm 3), $\frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}$, decreases with ξ_{relay}. Similarly, for a given ξ_{relay}, the mean outage probability per step, $\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$, decreases with ξ_{out} under the optimal policy.

Proof: See Appendix B. ∎

Remark: The proof of Theorem 5 is quite general; the results hold for the pure as-you-go approach also.

Theorem 6: For Problem (3), under the optimal policy (with explore-forward approach) characterized by λ^* (i.e., under the OptExploreLim algorithm), we have $\mathbb{E}_{\underline{W}} \min_{u,\gamma}(\gamma + \xi_{out}Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda^* u) = 0$.

Proof: See Appendix B. ∎

Remark 5: The result provided in Theorem 6 will be used to develop the learning algorithms in Section VI and Section VII.

C. Policy Computation

We adapt a policy iteration (from [17]) based algorithm to calculate λ^*. The algorithm generates a sequence of stationary policies $\{\mu_k\}_{k\geq 1}$ (note that the notation μ_k was used for a different purpose in Section IV-A; here each μ_k is a stationary, deterministic, Markov policy), such that for any $k \geq 1$, $\mu_k(\cdot) : \mathcal{W}^B \rightarrow \{A+1,\cdots,A+B\} \times \mathcal{S}$ maps a state into some action. Define the sequence $\{\mu_k^{(1)},\mu_k^{(2)}\}_{k\geq 1}$ of functions as follows: if $\mu_k(\underline{w}) = (u,\gamma)$, then $\mu_k^{(1)}(\underline{w}) = u$ and $\mu_k^{(2)}(\underline{w}) = \gamma$.

Algorithm 4: The policy iteration based algorithm is as follows:

Step 0 (Initialization): Start with an initial stationary deterministic policy μ_0.

Step 1 (Policy Evaluation): Calculate the average cost λ_k corresponding to the policy μ_k, for $k \geq 0$. λ_k is equal to the following quantity (by the Renewal Reward Theorem; see [19], Proposition 7.3):

$$\frac{\xi_{relay} + \sum_{\underline{w}} g(\underline{w}) \left(\mu_k^{(2)}(\underline{w}) + \xi_{out}Q_{out}(\mu_k^{(1)}(\underline{w}),\mu_k^{(2)}(\underline{w}),w_{\mu_k^{(1)}(\underline{w})}) \right)}{\sum_{\underline{w}} g(\underline{w})\mu_k^{(1)}(\underline{w})}$$

Step 2 (Policy Improvement): Find a new policy μ_{k+1} by solving the following:

$$\mu_{k+1}(\underline{w}) = \underset{(u,\gamma)}{\operatorname{argmin}} \left(\gamma + Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda_k u \right) \tag{10}$$

If μ_k and μ_{k+1} are the same policy (i.e., if $\lambda^{(k-1)} = \lambda_k$), then stop and declare $\mu^* = \mu_k$, $\lambda^* = \lambda_k$. Otherwise, go to Step 1. □

Remark: It was shown in [17] that this policy iteration will converge in a finite number of iterations, for finite state and action spaces. The policy iteration will provide λ^* in finite number of steps. The convergence requires that under any stationary policy, the state evolves as an irreducible Markov chain, which is satisfied here. When we have a general Borel state space (e.g., for log-normal shadowing), convergence may not happen in a finite number of states, but large enough number of iterations will provide a value close to λ^*.

Computational Complexity: The finite state space has cardinality $|\mathcal{W}|^B$. Then, $O(|\mathcal{W}|^B)$ addition operations are required to compute λ_k from the policy evaluation step. However, careful manipulation leads to a drastic reduction in this computational requirement, as shown by the following theorem.

Theorem 7: In the policy evaluation step in Algorithm 4, we can reduce the number of computations in each iteration from $|\mathcal{W}|^B$ to $O(B^2M^2|\mathcal{W}|^2)$.

Proof: See Appendix B. ∎

D. HeuExploreLim: An Intuitive but Suboptimal Heuristic

A natural heuristic for (3) under the explore-forward approach is the following HeuExploreLim Algorithm (Heuristic Algorithm for Limited Explore-Forward):

Algorithm 5: (HeuExploreLim Algorithm) Under the explore-forward setting as discussed in Section IV, at state $\underline{w}$, make the decision according to the following rule:

$$(u^*, \gamma^*) = \underset{u,\gamma}{\operatorname{argmin}} \frac{\gamma + \xi_{out} Q_{out}(u, \gamma, w_u) + \xi_{relay}}{u}$$

□

Lemma 2: HeuExploreLim solves $\inf_\mu \mathbb{E}_\mu\left(\frac{C_\mu}{U_\mu}\right)$.

Proof: See Appendix B. ■

Remark: This heuristic is not optimal. Under any stationary deterministic policy μ, let us denote the cost of a link by C_μ (a random variable) and the length of a link by U_μ (under any stationary deterministic policy μ, the deployment process regenerates at the placement points). Our optimal policy given in Theorem (3) solves $\inf_\mu \frac{\mathbb{E}_\mu(C_\mu)}{\mathbb{E}_\mu(U_\mu)}$. However, HeuExploreLim solves $\inf_\mu \mathbb{E}_\mu\left(\frac{C_\mu}{U_\mu}\right)$, which is, in general, different from $\inf_\mu \frac{\mathbb{E}_\mu(C_\mu)}{\mathbb{E}_\mu(U_\mu)}$. Note that $\mathbb{E}_\mu\left(\frac{C_\mu}{U_\mu}\right) = \frac{\mathbb{E}_\mu(C_\mu)}{\mathbb{E}_\mu(U_\mu)}$ *if and only if* the variance U_μ is zero, i.e., we always place at the same distance from the previous node. But this does not happen in practice due to the variability in shadowing over space. Hence, HeuExploreLim is suboptimal.

Remark: The advantage of HeuExploreLim is that, given ξ_{out} and ξ_{relay}, HeuExploreLim does not require any propagation model parameter such as η or σ. However, a learning algorithm reported in Section VI also has the same advantage, and provides near-optimal performance if the deployment continues for a sufficient number of steps.

V. Comparison between Explore-Forward and Pure As-You-Go Approaches

Let us denote the optimal average cost per step (for a given ξ_{out} and ξ_{relay}) under the explore-forward and pure as-you-go approaches by λ^*_{ef} and λ^*_{ayg}.

Theorem 8: $\lambda^*_{ef} \leq \lambda^*_{ayg}$.

Proof: See Appendix C. The proof is done by arguing that pure as-you-go approach is a special case explore-forward. ■

In Appendix C, we have presented some numerical work which illustrates the structure of the OptAsYouGo algorithm (we have shown the variation of the threshold $c_{th}(r)$ as a function of r, for various values of ξ_{out} and ξ_{relay}; see Appendix C, Section A), and also numerically compared various deployment algorithms (see Appendix C, Section B). A detailed explanation of the numerical results has also been provided in Appendix C. The purpose of the comparison is to provide insights into the performance of various algorithms, and to select the algorithm which is best suited for practical deployment. In this section, we will just discuss the choice of parameter values in our numerical work in Appendix C.

A. Parameter Values

We consider deployment for a given ξ_{out} and a given ξ_{relay}, for the objective in (3). In Appendix C, we provide numerical results for deployment with iWiSe motes ([20]) (based on the Texas Instrument (TI) CC2520 which implements the IEEE 802.15.4 PHY in the 2.4 GHz ISM band, yielding a bit rate of 250 Kbps, with a CSMA/CA medium access control (MAC); 9 dBi antennas were used in the experiments. The set of transmit power levels $\mathcal{S}$ is taken to be $\{-18, -7, -4, 0, 5\}$ dBm, which is a subset of the transmit power levels available in the chosen device. For the channel model as in (1), our measurements in a forest-like environment inside the Indian Institute of Science Campus gave path-loss exponent $\eta = 4.7$ and $c = 10^{0.17}$ (i.e., 1.7 dB); see [9]. Shadowing W was found to be log-normal; $W = 10^{\frac{Y}{10}}$ with $Y \sim \mathcal{N}(0, \sigma^2)$, where $\sigma = 7.7$ dB. Shadowing decorrelation distance was found to be 6 meters. Fading is assumed to be Rayleigh; $H \sim Exponential(1)$.

We define outage to be the event when the received signal power of a packet falls below $P_{rcv-min} = 10^{-9.7}$ mW (i.e., -97 dBm); for a commercial implementation of the PHY/MAC of IEEE 802.15.4, -97 dBm received power corresponds to a 2% packet loss probability for 127 byte packets for iWiSe motes, as per our measurements.

We consider deployment along a line with step size $\delta = 20$ meters, $A = 0$, $B = 5$.

Choice of B: There is no specific rule as to how to choose B; the choice can be arbitrary. A higher value of B will result in a lower value of the optimal average cost, since the deployment agent will have more choices for larger B. We chose B is the following way. Define a link to be good if its outage probability is less than 3%, and choose B to be the largest integer such that the probability of finding a good link of length $B\delta$ is more than 20%, when the highest transmit power is used. For the measured parameters $\eta = 4.7$, $c = 10^{0.17}$, $\sigma = 7.7$ dB, and 5 dBm transmit power, B turned out to be 5. If B is increased further, the probability of getting a good link will be very small. Hence, exploring beyond 5 steps would be wasteful in terms of measurement effort and deployment time. □

The main conclusion from the comparison among various algorithms in Appendix C is that the algorithms based on the explore-forward approach significantly outperform the algorithms based on the pure-as-you-go approach, at the cost of slightly more number of measurements per step (see Appendix C for a detailed discussion). Hence, for applications that do not require rapid deployment, such as deployment along a long forest trail for wildlife monitoring, explore-forward is a better approach to take. Thus, for the learning algorithms presented later, we will consider only explore-forward approach.

VI. OptExploreLimLearning: Learning with Explore-Forward, for Given ξ_{out} and ξ_{relay}

Based on the discussion in Section V, we proceed, in the remaining paper, with developing learning algorithms based

on the optimal policy for OptExploreLim. Let us recall problem (3). It is obvious from Section V that Explore-Forward approach is much better suited (compared to pure as-you-go approach) for deployment over large terrains, because it provides a good compromise between the network cost and the number of measurements to be made.

We observe that the optimal policy (given by Algorithm 3) can be completely specified by the optimal average cost per step λ^*, for given values of ξ_{out} and ξ_{relay}. But the computation of λ^* requires policy iteration. Policy iteration requires the channel model parameters η and σ, and it is computationally intensive. In practice, these parameters of the channel model might not be available. Under this situation, the agent measures $\{Q_{out}(u,\gamma,w_u) : A+1 \leq u \leq A+B, \gamma \in \mathcal{S}\}$ before deploying each relay, but he has to *learn* the optimal average cost per step in the process of deployment, and, use the corresponding updated policy each time he places a new relay. In order to address this requirement, we propose an algorithm which will maintain a running estimate of λ^*, and update it each time a relay is placed. The algorithm is motivated by the theory of Stochastic Approximation (see [21]), and it uses, as input, the measurements made for each placement, in order to improve the estimate of λ^*. We prove that, as the number of deployed relays goes to infinity, the running estimate of average network cost per step converges to λ^* almost surely. Note that, this algorithm is much more convenient than the traditional Q-learning algorithm; the Q-learning algorithm maintains a value function for each state of an MDP and updates each of them over time, whereas this algorithm updates only a scalar value, namely an estimate of the optimal average cost per step.

Let us recall that the sink is called node 0, and the subsequent relays are called nodes $1, 2, \cdots$. After the deployment is over, let us denote the length, transmit power and outage values of the link between node k and node $(k-1)$ by u_k, γ_k and $Q_{out}^{(k,k-1)}$. After placing the $(k-1)$-st node, we will place node k, and consequently u_k, γ_k and $Q_{out}^{(k,k-1)}$ will be decided according to the following algorithm.

Algorithm 6: (OptExploreLimLearning Algorithm) Let $\lambda^{(k)}$ be the estimate of the optimal average cost per step after placing the k-th relay (the sink is called node 0), and let $\lambda^{(0)}$ be the initial estimate. In the process of placing relay $(k+1)$, if the measured outage probabilities are $\{Q_{out}(u,\gamma,w_u) : A+1 \leq u \leq A+B, \gamma \in \mathcal{S}\}$, then place relay $(k+1)$ according to the following policy:

$$(u_{k+1},\gamma_{k+1}) = \underset{u,\gamma}{\operatorname{argmin}} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda^{(k)} u\Big)$$

After placing relay $(k+1)$, update $\lambda^{(k)}$ as follows (using the measurements made in the process of placing relay $(k+1)$):

$$\begin{aligned} & \lambda^{(k+1)} \\ = \;& \lambda^{(k)} + a_{k+1} \min_{u,\gamma} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda^{(k)} u\Big) \\ = \;& \lambda^{(k)} + a_{k+1} \Big(\gamma_{k+1} + \xi_{out} Q_{out}^{(k+1,k)} + \xi_{relay} - \lambda^{(k)} u_{k+1}\Big) \end{aligned} \quad (11)$$

$\{a_k\}_{k\geq 1}$ is a decreasing sequence such that $a_k > 0 \,\forall\, k \geq 1$, $\sum_k a_k = \infty$ and $\sum_k a_k^2 < \infty$. One example is $a_k = \frac{1}{k}$. □

Theorem 9: Suppose that the channel model is given by (1), and that shadowing is i.i.d. across links. If we employ Algorithm 6 in the deployment process, we will have $\lambda^{(k)} \to \lambda^*$ almost surely.

Proof: By Theorem 6, under the optimal policy specified by λ^*, we have $\mathbb{E}_{\underline{W}} \min_{u,\gamma}(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda^* u) = 0$, where $\mathbb{E}_{\underline{W}}$ denotes expectation over shadowing from locations $(A+1, \cdots, A+B)$. Writing $f(\underline{w},\lambda) = \min_{u,\gamma}(\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda u)$, we have $\mathbb{E}_{\underline{W}} f(\underline{W},\lambda^*) = 0$, thus leading to the stochastic approximation update in Algorithm 6. The detailed proof can be found in Appendix D. ■

While Algorithm 6 utilizes the general stochastic approximation update, Algorithm 7 ensures that the iterate $\lambda^{(k)}$ is the actual average network cost per step up to the k-th relay.

Algorithm 7: Start with any $\lambda^{(0)} > 0$. Let, for $k \geq 1$, $\lambda^{(k)}$ be the average cost per step for the portion of the network already deployed between the sink and the k-th relay, i.e.,

$$\lambda^{(k)} = \frac{\sum_{i=1}^{k} (\gamma_i + \xi_{out} Q_{out}^{(i,i-1)} + \xi_{relay})}{\sum_{i=1}^{k} u_i}$$

Place the $(k+1)$-st relay according to the following policy:

$$(u_{k+1},\gamma_{k+1}) = \underset{u,\gamma}{\operatorname{argmin}} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,w_u) + \xi_{relay} - \lambda^{(k)} u\Big) \square$$

Corollary 1: Under Algorithm 7 in the deployment process, we will have $\lambda^{(k)} \to \lambda^*$ almost surely.

Proof: See Appendix D. ■

VII. OptExploreLimAdaptiveLearning with Constraints on Outage Probability and Relay Placement Rate

In Section VI, we provided a stochastic approximation algorithm for relay deployment, with given multipliers ξ_{out} and ξ_{relay}, without knowledge of the propagation parameters. The multipliers have to be chosen appropriately in order to enforce performance targets in a constrained sequential optimization formulation. Let us recall that Theorem 1 tells us how to choose the Lagrange multipliers ξ_{out} and ξ_{relay} (if they exist) in (3) in order to solve the problem given in (4).

However, we need to know the radio propagation parameters (e.g., η and σ) in order to compute an optimal pair $(\xi_{out}^*, \xi_{relay}^*)$ (if it exists) so that both constraints in (4) are met with equality. In real deployment scenarios, these propagation parameters might not be known to the deployment agent. Hence, in this section, we provide a sequential placement and learning algorithm such that, as the relays are placed, the placement policy iteratively converges to the set of optimal policies for the constrained problem displayed in (4). The policy is of the OptExploreLim type, and the cost of the deployed network converges to the optimal cost. We modify the OptExploreLimLearning algorithm so that a running estimate $(\lambda^{(k)}, \xi_{out}^k, \xi_{relay}^k)$ gets updated each time a new relay is placed. The objective is to make sure that the running estimate $(\lambda^{(k)}, \xi_{out}^k, \xi_{relay}^k)$ eventually converges to

the set of optimal $(\lambda^*(\xi_{out},\xi_{relay}),\xi_{out},\xi_{relay})$ tuples as the deployment progresses, and that the constraints are satisfied asymptotically. Our approach is via two time-scale stochastic approximation.

A. OptExploreLim: Effect of Multipliers ξ_{out} and ξ_{relay}

Consider the constrained problem in (4) and its relaxed version in (3). We will seek a policy for the problem in (4) in the class of OptExploreLim policies (see (7)). Clearly, there exists at least one tuple $(\overline{q},\overline{N})$ for which there exists a pair $\xi^*_{out}>0,\xi^*_{relay}>0$ such that, under the optimal policy $\pi^*(\xi^*_{out},\xi^*_{relay})$, both constraints are met with equality. In order to see this, choose any $\xi_{out}>0,\xi_{relay}>0$ and consider the corresponding optimal policy $\pi^*(\xi_{out},\xi_{relay})$ (provided by OptExploreLim). Suppose that the mean outage per step and mean number of relays per step, under the policy $\pi^*(\xi_{out},\xi_{relay})$, are q_0 and n_0, respectively. Now, if we set the constraints $\overline{q}=q_0$ and $\overline{N}=n_0$ in (4), we obtain one instance of such a tuple $(\overline{q},\overline{N})$.

On the other hand, there exist $(\overline{q},\overline{N})$ pairs which are not feasible. One example is the case $\overline{N}=\frac{1}{A+B}$ (i.e., inter-node distance is always $(A+B)$), along with $\overline{q}<\frac{\mathbb{E}_W Q_{out}(A+B,P_M,W)}{A+B}$, where P_M is the maximum available transmit power level at each node. In this case, the outage constraint cannot be satisfied while meeting the constraint on the mean number of relays per step, since even use of the highest transmit power P_M at each node will not satisfy the per-step outage constraint.

Definition 1: Let us denote the optimal mean power per step for problem (4) by γ^*, for a given $(\overline{q},\overline{N})$. The set $\mathcal{K}(\overline{q},\overline{N})$ is defined as follows:

$$\begin{aligned}\mathcal{K}(\overline{q},\overline{N}) \quad = \quad & \Big\{(\lambda^*(\xi_{out},\xi_{relay}),\xi_{out},\xi_{relay}):\\ & \frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}=\gamma^*,\\ & \frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\leq\overline{q}\\ & \frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}\leq\overline{N},\\ & \xi_{out}\geq 0,\xi_{relay}\geq 0\Big\}\end{aligned}$$

where the optimal average cost per step of the unconstrained problem (3) under OptExploreLim is $\lambda^*(\xi_{out},\xi_{relay})$. □

$\mathcal{K}(\overline{q},\overline{N})$ can possibly be empty (in case $(\overline{q},\overline{N})$ is not a feasible pair). *Hence, we make the following assumption which ensures the non-emptiness of* $\mathcal{K}(\overline{q},\overline{N})$.

Assumption 1: The constraint parameters $\overline{q}$ and $\overline{N}$ in (4) are such that there exists at least one pair $\xi^*_{out}\geq 0,\xi^*_{relay}\geq 0$ for which $(\lambda^*(\xi^*_{out},\xi^*_{relay}),\xi^*_{out},\xi^*_{relay})\in\mathcal{K}(\overline{q},\overline{N})$. □

Remark: Assumption 1 implies that the constraints are consistent (in terms of achievability). If $\xi^*_{out}>0,\xi^*_{relay}>0$, it would imply that both of the constraints are active. If $\xi^*_{out}=0$, it would imply that we can keep the mean outage per step strictly less than $\overline{q}$ by using the minimum available power at each node, while meeting the constraint on the mean number of relays per step. The optimal policy in Algorithm 3, under $\xi_{out}=0$, will place relays with inter-relay distance $(A+B)$ steps, and use the minimum available power level at each node. $\xi^*_{out}=\infty$ implies that the outage constraint cannot be met even with the highest power level at each node, under the relay placement rate constraint. Similar arguments apply to ξ^*_{relay}. □

We now establish some structural properties of $\mathcal{K}(\overline{q},\overline{N})$.

Theorem 10: If $\mathcal{K}(\overline{q},\overline{N})$ is non-empty, then the following are true:

- Suppose that there exists $\xi^*_{out}>0$, $\xi^*_{relay}>0$ such that the policy $\pi^*(\xi^*_{out},\xi^*_{relay})$ satisfies both constraints in (4) with equality. Then, there does not exist $\xi'_{out}\geq 0,\xi'_{relay}\geq 0$ satisfying (i) $(\lambda^*(\xi'_{out},\xi'_{relay}),\xi'_{out},\xi'_{relay})\in\mathcal{K}(\overline{q},\overline{N})$, and (ii) $\frac{\overline{Q}^*_{out}(\xi'_{out},\xi'_{relay})}{\overline{U}^*(\xi'_{out},\xi'_{relay})}<\overline{q}$ or $\frac{1}{\overline{U}^*(\xi'_{out},\xi'_{relay})}<\overline{N}$.
- If there exists a $\xi'_{relay}\geq 0$ such that $(\lambda^*(0,\xi'_{relay}),0,\xi'_{relay})\in\mathcal{K}(\overline{q},\overline{N})$, then, for any $\xi_{relay}\geq 0$, we have $(\lambda^*(0,\xi_{relay}),0,\xi_{relay})\in\mathcal{K}(\overline{q},\overline{N})$. □

Proof: See Appendix E, Section A. ■

Assumption 2: The shadowing random variable W has a continuous probability density function (p.d.f.) over $(0,\infty)$; for any $w\in(0,\infty)$, $\mathbb{P}(W=w)=0$. One example could be log-normal shadowing. □

Theorem 11: Suppose that Assumption 2 holds. Under the OptExploreLim algorithm, the optimal mean power per step $\frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$, the optimal mean number of relays per step $\frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}$ and the optimal mean outage per step $\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$, are continuous in ξ_{out} and ξ_{relay}.

Proof: See Appendix E, Section B. ■

Remark: Note that, by Theorem 11, we need not do any randomization (see [22] for reference) among deterministic policies in order to meet the constraints with equality.

B. OptExploreLimAdaptiveLearning Algorithm

OptExploreLimAdaptiveLearning is based on the theory of two-timescale stochastic approximation (see [21], Chapter 6).

Algorithm 8: This algorithm iteratively updates $\lambda^{(k)},\xi^{(k)}_{out},\xi^{(k)}_{relay}$ after each relay is placed. Let $(\lambda^{(k)},\xi^{(k)}_{out},\xi^{(k)}_{relay})$ be the iterates after placing the k-th relay (the sink is called node 0), and let $(\lambda^{(0)},\xi^{(0)}_{out},\xi^{(0)}_{relay})$ be the initial estimates. In the process of deploying the k-th relay, if the shadowing (which is measured indirectly only via $Q_{out}(u,\gamma,w_u)$ for $A+1\leq u\leq A+B$ and $\gamma\in\mathcal{S}$) is $\underline{w}=\{w_{A+1},\cdots,w_{A+B}\}$, then place the k-th relay according to the following policy:

$$(u_k,\gamma_k)=\underset{u,\gamma}{\operatorname{argmin}}\left(\gamma+\xi^{(k-1)}_{out}Q_{out}(u,\gamma,w_u)+\xi^{(k-1)}_{relay}-\lambda^{(k-1)}u\right) \quad (12)$$

After placing the k-th relay, let us denote the transmit power, distance (in steps) and outage probability from relay k to relay $(k-1)$ by γ_k, u_k and $Q_{out}(u_k,\gamma_k,w_{u_k})$. After placing the k-

th relay, make the following updates (using the measurements made in the process of placing the k-th relay):

$$\begin{aligned}\lambda^{(k)} &= \lambda^{(k-1)} + a_k \min_{u,\gamma}\Big(\gamma + \xi_{out}^{(k-1)} Q_{out}(u,\gamma,w_u) \\ &\quad + \xi_{relay}^{(k-1)} - \lambda^{(k-1)} u\Big) \\ \xi_{out}^{(k)} &= \Lambda_{[0,A_2]}\Big(\xi_{out}^{(k-1)} + b_k(Q_{out}(u_k,\gamma_k,w_{u_k}) - \overline{q}u_k)\Big) \\ \xi_{relay}^{(k)} &= \Lambda_{[0,A_3]}\Big(\xi_{relay}^{(k-1)} + b_k(1 - \overline{N}u_k)\Big) \qquad (13)\end{aligned}$$

where $\Lambda_{[0,A_2]}(x)$ denotes the projection of x on the interval $[0, A_2]$. A_2 and A_3 need to be chosen carefully; the reason is explained in the discussion later in this section (along with a brief discussion on how A_2 and A_3 have to be chosen), and a detailed method of choosing A_2 and A_3 has been described in Appendix E, Section C5.

$\{a_k\}_{k\geq 1}$ and $\{b_k\}_{k\geq 1}$ are two decreasing sequences such that $a_k, b_k > 0, \forall k \geq 1$, $\sum_k a_k = \infty$, $\sum_k a_k^2 < \infty$, $\sum_k b_k = \infty$, $\sum_k b_k^2 < \infty$ and $\lim_{k\to\infty} \frac{b_k}{a_k} = 0$. In particular, we can use $a_k = C_1 k^{-n_1}$ and $b_k = C_2 k^{-n_2}$ where $C_1 > 0$, $C_2 > 0$, $\frac{1}{2} < n_1 < n_2 \leq 1$. □

Note that, for $(\xi_{out}, \xi_{relay}) \in [0, A_2] \times [0, A_3]$, we have $0 < \lambda^*(\xi_{out}, \xi_{relay}) \leq (P_M + A_2 + A_3)$. Let us define the set $\hat{\mathcal{K}}(\overline{q}, \overline{N}) := \mathcal{K}(\overline{q}, \overline{N}) \cap ([0, (P_M + A_2 + A_3)] \times [0, A_2] \times [0, A_3])$ which is a subset of $\mathcal{K}(\overline{q}, \overline{N})$.

Theorem 12: Under Assumption 1, Assumption 2 and under proper choice of A_2 and A_3, the iterates $(\lambda^{(k)}, \xi_{out}^{(k)}, \xi_{relay}^{(k)})$ in Algorithm 8 converge almost surely to $\hat{\mathcal{K}}(\overline{q}, \overline{N})$ as $k \to \infty$.

Proof: See Appendix E, Section C for a detailed proof. ■

Discussion of Theorem 12:

(i) *Two timescales:* In Appendix E, Section C, we rewrite the update scheme (13) as a two-timescale stochastic approximation (see [21], Chapter 6). Note that, $\lim_{k\to\infty} \frac{b_k}{a_k} = 0$, i.e., ξ_{out} and ξ_{relay} are adapted in a *slower* timescale compared to λ (which is adapted in the *faster* timescale). The dynamics behaves as if ξ_{out} and ξ_{relay} are updated simultaneously in a slow outer loop, and, between two successive updates of ξ_{out} and ξ_{relay}, we update λ in an inner loop for a long time. Thus, the λ update equation views ξ_{out} and ξ_{relay} as quasi-static, while the ξ_{out} and ξ_{relay} update equations view the λ update equation as almost equilibrated. See [21] (Chapter 6) for reference.

(ii) *Structure of the iteration:* Note that, $(Q_{out}(u_k, \gamma_k, w_{u_k}) - \overline{q}u_k)$ is the excess outage compared to the allowed outage $\overline{q}u_k$ for the k-th link. If this quantity is positive (resp., negative), the algorithm increases (resp., decreases) ξ_{out} in order to reduce (resp., increase) the outage probability in subsequent steps. Similarly, if $u_k < \frac{1}{\overline{N}}$, the algorithm increases ξ_{relay} in order to reduce the relay placement rate. The objective is to ensure $\lim_{k\to\infty}(\overline{Q}^*_{out}(\xi_{out}^{(k)}, \xi_{relay}^{(k)}) - \overline{q}\overline{U}^*(\xi_{out}^{(k)}, \xi_{relay}^{(k)})) = 0$ and $\lim_{k\to\infty}(1 - \overline{N}\overline{U}^*(\xi_{out}^{(k)}, \xi_{relay}^{(k)})) = 0$; we will see later (in Theorem 13) that this iteration will ensure that the constraints in (4) are met. In the faster timescale, our aim is to ensure that $\lim_{k\to\infty} \mathbb{E}_W \min_{u,\gamma}(\gamma + \xi_{out}^{(k)} Q_{out}(u, \gamma, W_u) + \xi_{relay}^{(k)} - \lambda^{(k)} u) = 0$.

(iii) *Outline of the proof:* We will present the proof of Theorem 12 in five subsections in Appendix E, Section C.
We first prove the almost sure boundedness of the $\lambda^{(k)}$ iterates in Subsection C1.
Next, we prove in Subsection C2 that the difference between the sequences $\lambda^{(k)}$ and $\lambda^*(\xi_{out}^{(k)}, \xi_{relay}^{(k)})$ converges to 0 almost surely; this will be required to prove the desired convergence in the faster timescale. This result has been proved using the theory in [21] (Chapter 6) and Theorem 9.
In order to ensure almost sure boundedness of the slower timescale iterates, we have used the projection operation in the slower timescale. In Subsection C3 we pose the slower timescale iteration in the same form as a projected stochastic approximation iteration (see [23], Equation 5.3.1*).*
In order to prove the desired convergence of the projected stochastic approximation, in Subsection C4, we will show that our iteration satisfies certain conditions given in [23] (see Theorem 5.3.1 in [23]).
In Subsection C5, we argue (using Theorem 5.3.1 of [23]) that the slower timescale iterates converge to the set of stationary points of a suitable ordinary differential equation (o.d.e.). But, in general, a stationary point on the boundary of the closed set $[0, A_2] \times [0, A_3]$ in the (ξ_{out}, ξ_{relay}) plane may not correspond to a point in $\mathcal{K}(\overline{q}, \overline{N})$. *Hence, we will need to ensure that if $(\xi'_{out}, \xi'_{relay})$ is a stationary point of the o.d.e., then $(\lambda^*(\xi'_{out}, \xi'_{relay}), \xi'_{out}, \xi'_{relay}) \in \mathcal{K}(\overline{q}, \overline{N})$. In order to ensure this, we need to choose A_2 and A_3 properly. The choice of A_2 and A_3 is rather technical, and is explained in detail in Appendix E, Section C5.* Here we will just provide a brief description of their choices, without any explanation. The number A_2 has to be chosen so large that under $\xi_{out} = A_2$ and for all $A + 1 \leq u \leq A + B$, we will have $\mathbb{P}(\text{argmin}_{\gamma\in\mathcal{S}}(\gamma + A_2 Q_{out}(u, \gamma, W)) = P_M) > 1 - \kappa$ for some small enough $\kappa > 0$. We must also have $\frac{\overline{Q}^*_{out}(A_2,0)}{\overline{U}^*(A_2,0)} \leq \overline{q}$. The number A_3 has to be chosen so large that for any $\xi_{out} \in [0, A_2]$, we will have $\overline{U}^*(\xi_{out}, A_3) > \frac{1}{\overline{N}}$ (provided that $\frac{1}{\overline{N}} < A + B$). The numbers A_2 and A_3 have to be chosen so large that there exists at least one $(\xi'_{out}, \xi'_{relay}) \in [0, A_2] \times [0, A_3]$ such that $(\lambda^*(\xi'_{out}, \xi'_{relay}), \xi'_{out}, \xi'_{relay}) \in \mathcal{K}(\overline{q}, \overline{N})$.

(iv) *Asymptotic behaviour of the iterates:* If the pair $(\overline{q}, \overline{N})$ is such that one can be met with strict inequality and the other can be met with equality while using the optimal mean power per step for this pair $(\overline{q}, \overline{N})$, then one Lagrange multiplier will converge to 0. This will happen if $\overline{q} > \frac{\mathbb{E}_W Q_{out}(A+B, P_1, W)}{A+B}$; we will have $\xi_{out}^{(k)} \to 0$ (obvious from OptExploreLim with $\xi_{out} = 0$) in this case. Here we will place all the relays at the $(A + B)$-th step and use the smallest power level at each node. On the other hand, if the constraints are not feasible, then either $\xi_{out}^{(k)} \to A_2$ or $\xi_{relay}^{(k)} \to A_3$ (since convergence to ∞ is not possible due to the projection operation) or both will

happen.

Note that, if $\mathcal{K}(\overline{q},\overline{N})$ is nonempty, then there might possibly exist multiple pairs $\xi^*_{out}\geq 0, \xi^*_{relay}\geq 0$ such that $(\lambda^*(\xi^*_{out},\xi^*_{relay}),\xi^*_{out},\xi^*_{relay})\in\mathcal{K}(\overline{q},\overline{N})$. So, in the rest of the paper, we will assume that $\mathcal{K}(\overline{q},\overline{N})$ may possibly have multiple tuples. However, we strongly believe that if $\mathcal{K}(\overline{q},\overline{N})$ is nonempty, then the iterates will always (for all sample paths) converge to the same limit (probably the only tuple in $\mathcal{K}(\overline{q},\overline{N})$); we found the evidence of this assertion by extensive simulation. □

C. Asymptotic Performance of OptExploreLimAdaptiveLearning

Let us denote by π_{oelal} the (nonstationary) deployment policy induced by the OptExploreLimAdaptiveLearning algorithm (i.e., Algorithm 8). We will now show that π_{oelal} is an optimal policy for the constrained problem (4).

Theorem 13: Suppose that Assumption 1 and Assumption 2 hold. Then, under proper choice of A_2 and A_3, the policy π_{oelal} solves the problem (4); i.e., we have:

$$\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}\Gamma_i}{x}=\gamma^*$$
$$\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}Q_{out}^{(i,i-1)}}{x}\leq\overline{q},\ \limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}N_x}{x}\leq\overline{N} \tag{14}$$

Proof: See Appendix E, Section D. ■

Theorem 14: Suppose that Assumption 1 and Assumption 2 hold. Then, under proper choice of A_2 and A_3, we have:

$$\limsup_{n\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{n}\Gamma_i}{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{n}U_i}=\gamma^*$$
$$\limsup_{n\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{n}Q_{out}^{(i,i-1)}}{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{n}U_i}\leq\overline{q},$$
$$\limsup_{n\to\infty}\frac{n}{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{n}U_i}\leq\overline{N} \tag{15}$$

Proof: The proof is similar to the proof of Theorem 13, and hence has been skipped. ■

VIII. Convergence Speed of Learning Algorithms: A Simulation Study

In this section, we provide a simulation study to demonstrate the convergence rate of the OptExploreLimLearning algorithm (Algorithm 7) and the OptExploreLimAdaptiveLearning algorithm (Algorithm 8).

The simulations are provided for $\eta=4.7$, $\sigma=7.7$ dB, $\delta=20$ m, $A=0$, $B=5$, $c=10^{0.17}$, $P_{rcv-min}=-97$ dBm, $\mathcal{S}=\{-18,-7,-4,0,5\}$ dBm (see Section II to recall the notation and Section V-A to recall the reason behind the choice of these parameter values).

A. OptExploreLimLearning for Given ξ_{out} and ξ_{relay}

Let us choose $\xi_{out}=100$ and $\xi_{relay}=1$. The optimal average cost per step, for this choice of parameters and under the propagation environment as in Section V-A, is 0.8312 (computed numerically using policy iteration).

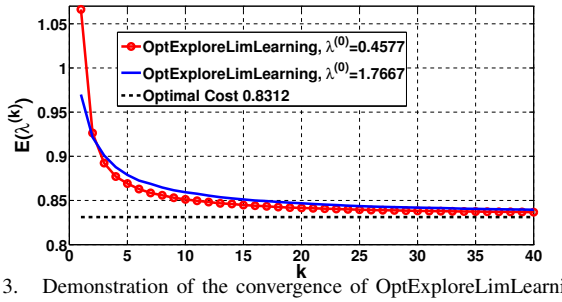

Fig. 3. Demonstration of the convergence of OptExploreLimLearning (Algorithm 7) as deployment progresses; details are provided in Section VIII-A. $\lambda^{(0)}$ has not been included in the plot.

On the other hand, for $\eta=4$, $\sigma=7$ dB, $\xi_{out}=100$ and $\xi_{relay}=1$, the optimal average cost per step (keeping the other parameters unchanged) is 0.4577, and it is 1.7667 for $\eta=5.5$, $\sigma=9$ dB.

Suppose that the actual $\eta=4.7$, $\sigma=7.7$ dB, but at the time of deployment we have an initial estimate that $\eta=4$, $\sigma=7$ dB; thus, we start with $\lambda^{(0)}=0.4577$. After placing the k-th relay, the actual average cost per step of the relay network connecting the k-th relay to the sink is $\lambda^{(k)}$; this quantity is a random variable whose realization depends on the shadowing realizations over the links measured in the process of deployment up to the k-th relay. We ran 10000 simulations of Algorithm 7 starting with different seeds for the shadowing random process (using the MATLAB package), and estimating $\mathbb{E}(\lambda^{(k)})$ as the average of the samples of $\lambda^{(k)}$ over these 10000 simulations. We also do the same for $\lambda^{(0)}=1.7667$, which is the optimal cost for $\eta=5.5$, $\sigma=9$ dB.

The estimates of $\mathbb{E}(\lambda^{(k)}), k\geq 1$ as a function of k, for the two initial values of $\lambda^{(0)}$, are shown in Figure 3. Also shown, in Figure 3, is the optimal value $\lambda^*=0.8312$ for the true propagation parameters (i.e., $\eta=4.7$, $\sigma=7.7$ dB). From Figure 3, we observe that $\mathbb{E}(\lambda^{(k)})$ approaches the optimal cost 0.8312 for the actual propagation parameters, as the number of deployed relays increases, and gets to within 10% of the optimal cost by the time that 4 or 5 relays are placed, starting with two widely different initial guesses of the propagation parameters. Thus, OptExploreLimLearning could be useful even in situations where the distance from sink to the source can be covered in as few as 4 to 5 relays.

Note that, each simulation yields one sample path of the deployment process. We ran 10000 simulations in order to obtain the estimates of $\mathbb{E}(\lambda^{(k)})$ as a function of k (by averaging over 10000 sample paths); the convergence speed will vary from one sample path to another even though $\lambda^{(k)}\to 0.8312$ almost surely as $k\to\infty$.

B. OptExploreLimAdaptiveLearning

In this section, we will discuss how OptExploreLimAdaptiveLearning (Algorithm 8) performs for deployment over a finite distance under an unknown propagation environment. We assume that the true propagation parameters are given in Section V-A (e.g., $\eta=4.7$, $\sigma=7.7$ dB). If we know the true propagation environment, then, under the choice

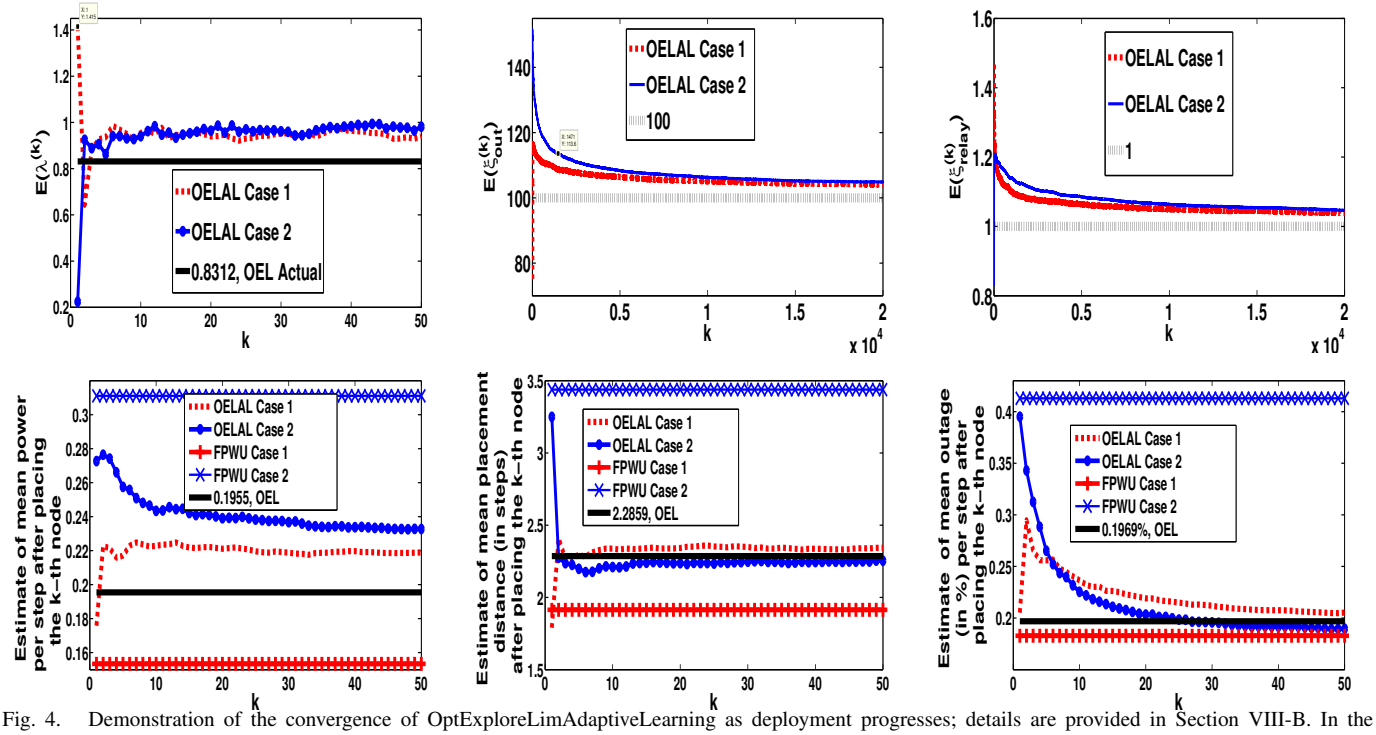

Fig. 4. Demonstration of the convergence of OptExploreLimAdaptiveLearning as deployment progresses; details are provided in Section VIII-B. In the legends, "OEL" refers to the values that are obtained if OptExploreLim is used; these are the target values for OptExploreLimAdaptiveLearning. Note that, we have used line styles for ξ_{out} and ξ_{relay} updates, that are different from the line styles of other four plots. Also note that, outage probabilities are shown in percentage and not in decimal.

$\xi_{relay} = 1$ and $\xi_{out} = 100$, the optimal average cost per step will be 0.8312, and this can be achieved by OptExploreLim (Algorithm 3). The corresponding mean outage per step will be $\frac{0.0045}{2.2859} = 0.001969$ (i.e., 0.1969%) and the mean number of relays per step will be $\frac{1}{2.2859}$ (see Figure 7 in Appendix C).

Now, suppose that we wish to solve the constrained problem in (4) with the targets $\overline{q} = 0.001969$ (i.e., 0.1969%) and $\overline{N} = \frac{1}{2.2859}$, but we do not know the true propagation environment. Hence, the deployment will use OptExploreLimAdaptiveLearning with some initial choice of $\xi_{out}^{(0)}$, $\xi_{relay}^{(0)}$ and $\lambda^{(0)}$.

In order to make a fair comparison, we seek to compare among the following three scenarios: (i) η and σ are completely known (we use OptExploreLim with $\xi_{relay} = 1$ and $\xi_{out} = 100$ in this case), (ii) imperfect estimates of η and σ are available prior to deployment, and OptExploreLimAdaptiveLearning is used to learn the optimal policy, and (iii) imperfect estimates of η and σ are available prior to deployment, but a corresponding suboptimal policy is used throughout the deployment without any update. For convenience in writing, we introduce the abbreviations OELAL and OEL for OptExploreLimAdaptiveLearning and OptExploreLim, respectively. We also use the abbreviation FPWU for "Fixed Policy without Update." Now, we formally introduce the following cases that we consider in our simulations:

(i) OEL: OEL corresponds to the case where we know $\eta = 4.7$, $\sigma = 7.7$ dB, and use OptExploreLim (Algorithm 3) with $\xi_{out} = 100$, $\xi_{relay} = 1$, $\lambda^* = 0.8312$. OEL will meet both the constraints with equality, and at the same time will minimize the mean power per step.

(ii) OELAL Case 1: OELAL Case 1 is the case where the true η and σ (which are unknown to the deployment agent) are specified by Section V-A, but we use OptExploreLimAdaptiveLearning with $\xi_{out}^{(0)} = 75$, $\xi_{relay}^{(0)} = 1.25$ and $\lambda^{(0)} = 0.5007$, in order to meet the constraints specified earlier in this subsection. Note that, under $\xi_{out} = 75$ and $\xi_{relay} = 1.25$, the optimal mean cost per step is 0.5007 for $\eta = 4$, $\sigma = 7$ dB. Hence, we start with a wrong choice of Lagrange multipliers, a wrong estimate of η and σ, and an estimate of the optimal average cost per step which corresponds to these wrong choices. The goal is to see how fast the variables $\lambda^{(k)}$, $\xi_{out}^{(k)}$ and $\xi_{relay}^{(k)}$ converge to the desired target 0.8312, 100 and 1, respectively. We also seek to study how close to the desired target values are the quantities such as mean power per step, mean outage per step and mean placement distance for the relay network between k-th relay and the sink node (for any $k \geq 1$).

(iii) OELAL Case 2: OELAL Case 2 is different from OELAL Case 1 only in the aspect that $\lambda^{(0)} = 1.7679$ is used in OELAL Case 2. Note that, under $\xi_{out} = 75$ and $\xi_{relay} = 1.25$, the optimal mean cost per step is 1.7679 for $\eta = 5.5$, $\sigma = 9$ dB.

(iv) FPWU Case 1: In this case, the true η and σ are unknown to the deployment agent. The deployment agent uses $\xi_{out} = 75$, $\xi_{relay} = 1.25$ and $\lambda^* = 0.5007$ throughout the deployment process under the algorithm specified by (7). Clearly, he chooses a wrong set of Lagrange multipliers $\xi_{out} = 75$, $\xi_{relay} = 1.25$, and he has a wrong estimate $\eta = 4$, $\sigma = 7$ dB. The optimal

average cost per step $\lambda^* = 0.5007$ is computed for these wrong choice of parameters, and the corresponding policy is used throughout the deployment process without any update. This case is simulated to see what is the gain in performance by updating the policy under OptExploreLimAdaptiveLearning, w.r.t. the case where a suboptimal policy driven by the initial imperfect estimate of parameters is used without any online update.

(v) FPWU Case 2: It differs from FPWU Case 1 only in the aspect that we use $\lambda^* = 1.7679$ in FPWU Case 2. Recall that, under $\xi_{out} = 75$ and $\xi_{relay} = 1.25$, the optimal mean cost per step is 1.7679 for $\eta = 5.5$, $\sigma = 9$ dB.

For simulation of OELAL, we chose the step sizes as follows. We chose $a_k = \frac{1}{k^{0.55}}$, chose $b_k = \frac{10000}{k^{0.8}}$ for the ξ_{out} update and $b_k = \frac{1}{k^{0.8}}$ for the ξ_{relay} update (note that, both ξ_{out} and ξ_{relay} are updated in the same timescale). We simulated 10000 independent network deployments (i.e., 10000 sample paths of the deployment process) with OptExploreLimAdaptiveLearning in MATLAB, and estimated (by averaging over 10000 deployments) the expectations of $\lambda^{(k)}$, $\xi_{out}^{(k)}$, $\xi_{relay}^{(k)}$, mean power per step $\frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{k} \Gamma_i}{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{k} U_i}$, mean outage per step $\frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{k} Q_{out}^{(i,i-1)}}{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{k} U_i}$ and mean placement distance $\frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{k} U_i}{k}$, from the sink node to the k-th placed node. In each simulated network deployment, we placed 20000 nodes, i.e., k was allowed to go up to 20000. Asymptotically the estimates are supposed to converge to the values provided by OEL (by Theorem 14).

Observations from the Simulations: The results of the simulations are summarized in Figure 4 (see the previous page). From these plots, we make the following important observations.

The estimates of the expectations of $\lambda^{(20000)}$, $\xi_{out}^{(20000)}$, $\xi_{relay}^{(20000)}$, mean power per step up to the 20000-th node, mean outage per step up to the 20000-th node, and mean placement distance (in steps) over 20000 deployed nodes are 0.8551, 104.0606, 1.0385, 0.2005, 0.2% (i.e., 0.002) and 2.2939 for the OELAL Case 1, whereas those quantities are supposed to be equal to 0.8312, 100, 1, 0.1955, 0.1969% (i.e., 0.001969) and 2.2859, respectively. We found similar results for OELAL Case 2 also. Hence, the quantities converge very close to the desired values. *We have shown convergence only up to $k = 50$ deployments in most cases, since the convergence rate of the algorithms in the initial phase are most important in practical deployments.*

All the quantities except expectation of $\xi_{out}^{(k)}$ and $\xi_{relay}^{(k)}$ (which are updated in a slower timescale) converge reasonably close to the desired values by the time the 50-th relay is placed, which will cover a distance of roughly $2 - 3.5$ km. distance.

FPWU Case 1 and FPWU Case 2 either violate some constraint or uses significantly higher per-step power compared to OEL. But, by using the OptExploreLimAdaptiveLearning algorithm, we can achieve per-step power expenditure close to the optimal while (possibly) violating the constraints by small amount; even in case the performance of OELAL is not very close to the optimal performance, it will be significantly better than the performance under FPWU cases (compare OELAL Case 2 and FPWU Case 2 in Figure 4). □

The speed of convergence will depend on the choice of the step sizes. We have shown the numerical results for one particular a_k and b_k sequence; optimizing the rate of convergence by choosing optimal step sizes is left for future endeavours in this direction. Also, note that, the choice of $\xi_{out}^{(0)}$, $\xi_{relay}^{(0)}$ and $\lambda^{(0)}$ will have a significant effect on the performance of the network over a finite length; the more accurate are the estimates of η and σ, and the better are the initial choice of $\xi_{out}^{(0)}$, $\xi_{relay}^{(0)}$ and $\lambda^{(0)}$, the better will be the convergence speed of OptExploreLimAdaptiveLearning.

IX. Conclusion

In this paper, we have developed several approaches for as-you-go deployment of wireless relay networks using online measurements, under very light traffic assumption. Each problem was formulated as an MDP and its optimal policy structure was studied. We also studied a few learning algorithms that will asymptotically converge to the corresponding optimal policies. Numerical results have been provided to illustrate the performance and trade-offs.

This work can be extended or modified in several ways: (i) Networks that are robust to node failures and long term link variations would either require each relay to have multiple neighbours (i.e., the deployment would need to be multi-connected), or each node can use a power level that is higher than the power specified by the deployment algorithm, or the nodes can choose their transmit powers adaptively as the environment changes. (ii) It would be of interest to develop deployment algorithms for 2 and 3 dimensional regions, where a team of agents cooperates to carry out the deployment. (iii) If network lifetime is not a matter of concern, one could use a fixed high transmit power level in all nodes. Then the problem would be to find deployment strategies so that the mean outage per step is minimized subject to a constraint on the mean relay placement rate. The approach taken in our current paper is able to address this new problem as well. (iv) We have assumed very light traffic conditions in our design (what we call "lone packet" traffic). But we have found that these designs can carry a useful amount of positive traffic; our experimental deployment (reported in [9]) of a network using OptExploreLimLearning with suitable parameters demonstrates that a 500 m long network having only 5 relays can carry 4 packets per second, while having end-to-end packet loss probability less than 1%, in a typical forest-like environment. It will be of interest, however, to develop deployment algorithms that can provide theoretical guarantees to achieve desired traffic rates.

References

[1] A. Chattopadhyay, M. Coupechoux, and A. Kumar. Sequential decision algorithms for measurement-based impromptu deployment of a wireless relay network along a line. *http://arxiv.org/abs/1502.06878*.

[2] M. Howard, M.J. Matarić, and S. Sukhat Gaurav. An incremental self-deployment algorithm for mobile sensor networks. *Kluwer Autonomous Robots*, 13(2):113–126, 2002.

[3] M.R. Souryal, J. Geissbuehler, L.E. Miller, and N. Moayeri. Real-time deployment of multihop relays for range extension. In *Proc. of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 85–98. ACM, 2007.

[4] Thorsten Aurisch and Jens Tölle. Relay Placement for Ad-hoc Networks in Crisis and Emergency Scenarios. In *Proc. of the Information Systems and Technology Panel (IST) Symposium*. NATO Science and Technology Organization, 2009.

[5] H. Liu, J. Li, Z. Xie, S. Lin, K. Whitehouse, J. A. Stankovic, and D. Siu. Automatic and robust breadcrumb system deployment for indoor firefighter applications. In *Proc. of the International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2010.

[6] Jeffrey Q. Bao and C. Lee. Rapid deployment of wireless ad hoc backbone networks for public safety incident management. In *Global Telecommunications Conference (GLOBECOM)*, pages 1217–1221. IEEE, 2007.

[7] A. Sinha, A. Chattopadhyay, K.P. Naveen, M. Coupechoux, and A. Kumar. Optimal sequential wireless relay placement on a random lattice path. *Ad Hoc Networks Journal (Elsevier).*, 21:1–17, 2014.

[8] A. Chattopadhyay, M. Coupechoux, and A. Kumar. Measurement based impromptu deployment of a multi-hop wireless relay network. In *Proc. of the 11th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*. IEEE, 2013.

[9] A. Chattopadhyay, A. Ghosh, A.S. Rao, B. Dwivedi, S.V.R. Anand, M. Coupechoux, and Anurag Kumar. Impromptu deployment of wireless relay networks: Experiences along a forest trail. *Proceedings of the IEEE International Conference on Mobile Ad hoc and Sensor Systems (MASS), 2014, a detailed version available in http://arxiv.org/abs/1409.3940*.

[10] Piyush Agrawal and Neal Patwari. Correlated link shadow fading in multi-hop wireless networks. *http://arxiv.org/abs/0804.2708*.

[11] A. Bhattacharya, A. Rao, D. G. Rao Sahib, A. Mallya, S.M. Ladwa, R. Srivastava, S.V.R. Anand, and A. Kumar. Smartconnect: A system for the design and deployment of wireless sensor networks. In *Proc. of the 5th International Conference on Communication Systems and Networks (COMSNETS)*. IEEE, 2013.

[12] A. Bhattacharya and A. Kumar. QoS aware and survivable network design for planned wireless sensor networks. *http://arxiv.org/abs/1110.4746*.

[13] Matthias Vodel and Wolfram Hardt. Energy-efficient communication in distributed, embedded systems. In *Proc. of the 9th International Workshop on Resource Allocation, Cooperation and Competition in Wireless Networks (RAWNET), in conjunction with IEEE WiOpt*. IEEE, 2013.

[14] Frederick J. Beutler and Keith W. Ross. Optimal policies for controlled markov chains with a constraint. *Journal of Mathematical Analysis and Applications*, 112:236–252, 1985.

[15] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific, 2007.

[16] Onesimo Hernandez-Lerma and Jean Bernard Lasserre. *Discrete-Time Markov Control Processes Basic Optimality Criteria*. Springer, 1996.

[17] H.C. Tijms. *A First Course in Stochastic Models*. WILEY, 2003.

[18] Oscar Vega-Amaya and Fernando Luque-Vsquez. Sample-path average cost optimality for semi-Markov control processes on Borel spaces: unbounded costs and mean holding times. *Applicationes Mathematicae*, 27(3):343–367, 2000.

[19] Sheldon M. Ross. *Introduction to Probability Models, Ninth Edition*. Academic Press, 2007.

[20] *http://www.astec.org.in/astec/content/wireless-sensor-network*.

[21] Vivek S. Borkar. *Stochastic approximation: a dynamical systems viewpoint.* Cambridge University Press, 2008.

[22] Dye-Jyun Ma and Armand M. Makowski'. A class of steering policies under a recurrence condition. In *Proc. of the 27th Conference on Decision and Control (CDC)*. IEEE, 1988.

[23] H.J. Kushner and D.S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, 1978.

[24] Vivek S. Borkar. Stochastic approximation with two time scales. *Systems and Control Letters*, 29:291–294, 1997.

[25] Walter Rudin. *Principles of Mathematical Analysis, Third Edition*. McGraw-Hill International Editions, 1976.

[26] Tom M. Apostol. *Mathematical Analysis, Second Edition*. Addison-wesley Publishing Company, 1981.

.

Supplementary Material

Appendix A
Pure As-You-Go Deployment

Proof of Lemma 1 Note that the function $J^{(0)}(\cdot) := 0$ satisfies all the assertions. Let us assume, as our induction hypothesis, that $J^{(k)}(\cdot)$ satisfies all the assertions. Now $Q_{out}(r,\gamma,w)$ is increasing in r and decreasing in w (by our channel modeling assumptions in Section II-A), and the single stage costs are linear (hence concave) increasing in ξ_{relay}, ξ_{out}. Then from the value iteration, $J^{(k+1)}(r,w)$ is pointwise minimum of functions which are increasing in r, ξ_{out} and ξ_{relay}, decreasing in w, and jointly concave in ξ_{out} and ξ_{relay}. Hence, the assertions hold for $J^{(k+1)}(r,w)$. Similarly, we can show that the assertions hold for $J^{(k+1)}(\mathbf{0})$. Since $J^{(k)}(\cdot) \uparrow J(\cdot)$, the results follow.

Proof of Theorem 2 Consider the Bellman equation (6). We will place a relay at state (r,w) iff the cost of placing a relay, i.e., $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w))+\xi_{relay}+J(\mathbf{0})$ is less than or equal to the cost of not placing, i.e., $\theta\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma + \xi_{out}Q_{out}(r+1,\gamma,W)) + (1-\theta)\mathbb{E}_W J(r+1,W)$. Hence, it is obvious that we will place a relay at state (r,w) iff $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w)) \le c_{th}(r)$ where the threshold $c_{th}(r)$ is given by:

$$\begin{aligned} c_{th}(r) &= \theta\mathbb{E}_W \min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r+1,\gamma,W)) \\ &\quad +(1-\theta)\mathbb{E}_W J(r+1,W) - (\xi_{relay}+J(\mathbf{0})) \end{aligned} \tag{16}$$

By Proposition 3.1.3 of [15], if there exists a stationary policy $\{\mu,\mu,\cdots\}$ such that for each state, the action chosen by the policy is the action that achieves the minimum in the Bellman equation, then that stationary policy will be an optimal policy, i.e., the minimizer in Bellman equation gives the optimal action. Hence, if the decision is to place a relay at state (r,w), then the power has to be chosen as $\text{argmin}_{\gamma\in\mathcal{S}}\left(\gamma+\xi_{out}Q_{out}(r,\gamma,w)\right)$.

Since $Q_{out}(r,\gamma,w)$ and $J(r,w)$ is increasing in r for each γ,w, it is easy to see that $c_{th}(r)$ is increasing in r.

Appendix B
Explore Forward Deployment

Proof of Theorem 4 Let us recall the definition of the functions $\mu^{(1)}$ and $\mu^{(2)}$. Now, $\lambda_\mu := \frac{\xi_{relay}+\sum_{\underline{w}} g(\underline{w})\left(\mu^{(2)}(\underline{w})+\xi_{out}Q_{out}(\mu^{(1)}(\underline{w}),\mu^{(2)}(\underline{w}),w_{\mu^{(1)}(\underline{w})})\right)}{\sum_{\underline{w}} g(\underline{w})\mu^{(1)}(\underline{w})}$ is the average cost of a specific stationary deterministic policy μ (by the Renewal Reward Theorem, since the placement process regenerates at each placement point). For each policy $(\mu^{(1)},\mu^{(2)})$, the numerator is linear, increasing in ξ_{out} and ξ_{relay} and the denominator is independent of ξ_{out} and ξ_{relay}. Now, $\lambda^*(\xi_{out},\xi_{relay}) = \inf_\mu \lambda_\mu$. Hence, the proof follows immediately since the pointwise infimum of increasing linear functions of ξ_{out} and ξ_{relay} is increasing and jointly concave in ξ_{out} and ξ_{relay}, and since any increasing, concave function is continuous.

Proof of Theorem 5: We will prove only the second statement of the theorem since the proof of the first statement is similar.

Consider any $\kappa > 0$.

Now, since the mean cost per step is a linear combination of the mean power per step, mean outage per step and the mean number of relays per step, we can write:

$$\begin{aligned} &\lambda^*(\xi_{out},\xi_{relay}) \\ = &\frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})+\xi_{out}\overline{Q}^*_{out}(\xi_{out},\xi_{relay})+\xi_{relay}}{\overline{U}^*(\xi_{out},\xi_{relay})} \\ \le &\frac{\overline{\Gamma}^*(\xi_{out}+\kappa,\xi_{relay})+\xi_{out}\overline{Q}^*_{out}(\xi_{out}+\kappa,\xi_{relay})+\xi_{relay}}{\overline{U}^*(\xi_{out}+\kappa,\xi_{relay})} \end{aligned} \tag{17}$$

and

$$\begin{aligned} &\lambda^*(\xi_{out}+\kappa,\xi_{relay}) \\ = &\frac{\overline{\Gamma}^*(\xi_{out}+\kappa,\xi_{relay})+(\xi_{out}+\kappa)\overline{Q}^*_{out}(\xi_{out}+\kappa,\xi_{relay})+\xi_{relay}}{\overline{U}^*(\xi_{out}+\kappa,\xi_{relay})} \\ \le &\frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})+(\xi_{out}+\kappa)\overline{Q}^*_{out}(\xi_{out},\xi_{relay})+\xi_{relay}}{\overline{U}^*(\xi_{out},\xi_{relay})} \end{aligned} \tag{18}$$

where the inequality in (17) follows from the fact that $\pi^*(\xi_{out},\xi_{relay})$ is an optimal policy for (ξ_{out},ξ_{relay}), and the inequality in (18) follows from the fact that $\pi^*(\xi_{out}+\kappa,\xi_{relay})$ is an optimal policy for $(\xi_{out}+\kappa,\xi_{relay})$.

Adding the inequalities (17) and (18) and cancelling the common terms, we obtain that $\frac{\overline{Q}^*_{out}(\xi_{out}+\kappa,\xi_{relay})}{\overline{U}^*(\xi_{out}+\kappa,\xi_{relay})} \le \frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}$. □

Proof of Theorem 6: From (8), we can write:

$$\begin{aligned} \sum_{\underline{w}} g(\underline{w})v^*(\underline{w}) &= \sum_{\underline{w}} g(\underline{w})\left(\min_{u,\gamma}\left\{\gamma+\xi_{out}Q_{out}(u,\gamma,w_u)\right.\right. \\ &\quad \left.\left.+\xi_{relay}-\lambda^* u\right\}\right) + \sum_{\underline{w}'\in\mathcal{W}^B} g(\underline{w}')v^*(\underline{w}') \end{aligned}$$

Cancelling $\sum_{\underline{w}} g(\underline{w})v^*(\underline{w})$ from both sides, we obtain the desired result.

Proof of Theorem 7: Note that in (10), if the minimum is achieved by more than one pair of (u,γ), then any one of them can be considered to be the optimal action. Let us use the convention that among all minimizers the pair (u,γ) with minimum u will be considered as the optimal action, and if there are more than one such minimizing pair with same values of u, then the pair with smallest value of γ will be considered. We recall that $\mathcal{S} = \{P_1,P_2,\cdots,P_M\}$. Let us denote, under policy μ_{k+1}, the probability that the optimal control is (u,γ)

and the shadowing is w at the u-th location, by $b_k(u,\gamma,w)$. Then,

$$\begin{aligned} b_k(u,\gamma,w) &= \Pi_{r=A+1}^{u-1}\mathbb{P}\Big(\min_{\gamma^{'}\in\mathcal{S}}(\gamma^{'}+\xi_{out}Q_{out}(r,\gamma^{'},W_r))-\lambda_k r \\ &\quad > \gamma+\xi_{out}Q_{out}(u,\gamma,w)-\lambda_k u\Big)\times p_W(w) \\ &\quad \times\Pi_{r=u+1}^{A+B}\mathbb{P}\Big(\min_{\gamma^{'}\in\mathcal{S}}(\gamma^{'}+\xi_{out}Q_{out}(r,\gamma^{'},W_r))-\lambda_k r \\ &\quad \geq \gamma+\xi_{out}Q_{out}(u,\gamma,w)-\lambda_k u\Big) \\ &\quad \times\mathbb{I}\Big\{\gamma=\text{argmin}\{P_1,P_2,\cdots,P_M\}: \\ &\quad \gamma+Q_{out}(u,\gamma,w) \\ &\quad = \min_{\gamma^{'}}(\gamma^{'}+\xi_{out}Q_{out}(u,\gamma^{'},w))\Big\} \end{aligned} \tag{19}$$

Now, we can write,

$$\begin{aligned} &\sum_{\underline{w}} g(\underline{w})\Big(\mu_k^{(2)}(\underline{w})+\xi_{out}Q_{out}(\mu_k^{(1)}(\underline{w}),\mu_k^{(2)}(\underline{w}),w_{\mu_k^{(1)}(\underline{w})})\Big) \\ &= \sum_{u=A+1}^{A+B}\sum_{j=1}^{M}\sum_{w\in\mathcal{W}} b_{k-1}(u,P_j,w)\Big(P_j+\xi_{out}Q_{out}(u,P_j,w)\Big) \end{aligned} \tag{20}$$

and

$$\begin{aligned} \sum_{\underline{w}} g(\underline{w})\mu_k^{(1)}(\underline{w}) &= \sum_{u=A+1}^{A+B}\sum_{j=1}^{M}\sum_{w\in\mathcal{W}} b_{k-1}(u,P_j,w)u \\ &= \sum_{u=A+1}^{A+B} u\sum_{j=1}^{M}\sum_{w\in\mathcal{W}} b_{k-1}(u,P_j,w) \end{aligned} \tag{21}$$

Now, for each (u,γ,w), $b_{k-1}(u,\gamma,w)$ (in (19)) can be computed in $O(BM|\mathcal{W}|)$ operations. Hence, total number of operations required to compute $b_{k-1}(u,\gamma,w)$ for all u,γ,w is $O(B^2M^2|\mathcal{W}|^2)$. Now, only $O(BM|\mathcal{W}|)$ operations are required in (20) and (21). Hence, the number of computations required in each iteration is $O(B^2M^2|\mathcal{W}|^2)$.

Note that, the policy improvement step is not explicitly required in the policy iteration. This is because in the policy evaluation step, λ_k is sufficient to compute $b_k(u,\gamma,w)$ for all u,γ,w and thereby to compute λ_{k+1}. Hence, we need not store the policy in each iteration. □

Proof of Lemma 2: Let us denote the HeuExploreLim policy by μ_h and any other stationary, deterministic policy by μ. Let us denote the sequence of link costs incurred in the deployment process (for a semi-infinite line with given shadowing over all possible links) under policy μ_h by $c_{\mu_h,1},c_{\mu_h,2},\cdots$ and the corresponding link lengths by $u_{\mu_h,1},u_{\mu_h,2},\cdots$. Let us denote, under policy μ_h, the shadowing observed at the i-th location (where $A+1\leq i\leq A+B$) in the measurement process for the placement of the l-th node, by $w_{i,l}$. Now, let us couple the deployment processes under policies μ and μ_h in the following way. Suppose that, under policy μ, the shadowing observed at the i-th location for the placement of the l-th node is again $w_{i,l}$ (this is valid since shadowing is i.i.d across links). Clearly, $\frac{c_{\mu_h,j}}{u_{\mu_h,j}}\leq\frac{c_{\mu,j}}{u_{\mu,j}}$. Hence, by the strong law of large numbers, $\mathbb{E}_{\mu_h}\left(\frac{C_{\mu_h}}{U_{\mu_h}}\right)\leq\mathbb{E}_{\mu}\left(\frac{C_{\mu}}{U_{\mu}}\right)$, since $\left(\frac{C_{\mu_h,j}}{U_{\mu_h,j}}\right)$ is i.i.d. across j due to i.i.d. shadowing across links.

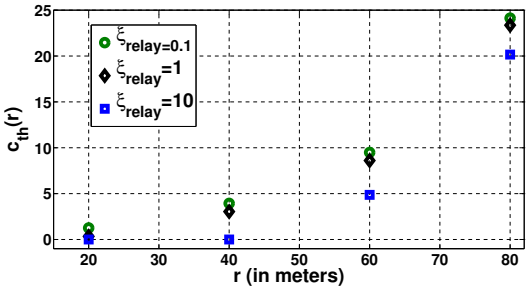

Fig. 5. Pure as-you-go deployment; variation of $c_{th}(r)$ with r for $\xi_{out}=100$ and various values of ξ_{relay}.

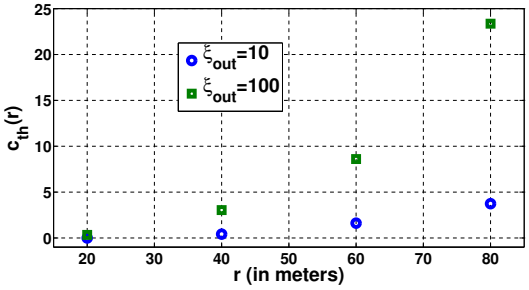

Fig. 6. Pure as-you-go deployment; variation of $c_{th}(r)$ with r for $\xi_{relay}=1$ and various values of ξ_{out}.

APPENDIX C
COMPARISON BETWEEN EXPLORE-FORWARD AND PURE AS-YOU-GO APPROACHES

Proof of Theorem 8 Note that for the average cost problem with pure as-you-go, there exists an optimal threshold policy (similar to Theorem 2), since the optimal policy for problem (5) achieves λ_{ayg}^* average cost per step for θ sufficiently close to 0. So, let one such optimal policy be given by the set of thresholds $\{c_{th}(r)\}_{A+1\leq r\leq A+B-1}$.

Now, let us consider the average cost minimization problem with explore-forward. Consider the policy where we first measure $w_{A+1},w_{A+2},\cdots,w_{A+B}$ and decide to place a relay u steps away from the previous relay (where $A+1\leq u\leq A+B-1$) if $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(r,\gamma,w_r))>c_{th}(r)$ for all $r\leq(u-1)$ and $\min_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(u,\gamma,w_u))\leq c_{th}(u)$. We must place if we reach at a distance $(A+B)$ from the previous relay. But this is a particular policy for the problem where we gather $w_{A+1},w_{A+2},\cdots,w_{A+B}$ and then decide where to place the relay, and clearly the average cost per step for this policy is λ_{ayg}^* which cannot be less than the optimal average cost λ_{ef}^*. □

A. Optimal Policy Structure for the Pure As-You-Go Approach

The variation of $c_{th}(r)$ (see Section III-E and Section III-G, we have taken θ sufficiently close to 0) with r, for various values of the relay cost ξ_{relay} and the cost of outage ξ_{out}, has been shown in Figure 5 and Figure 6. For a fixed ξ_{out}, $c_{th}(r)$ decreases with ξ_{relay}; i.e., as the cost of placing a relay increases, we place relays less frequently. On the other hand,

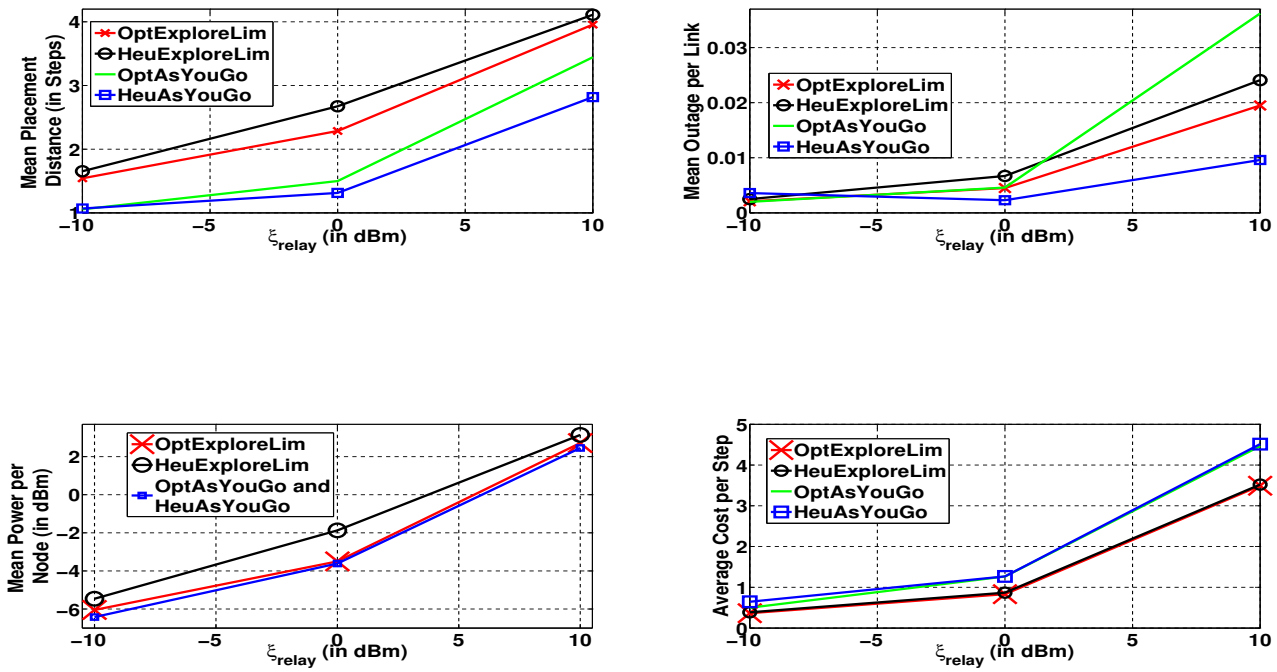

Fig. 7. Results for $\xi_{out} = 100$: mean cost per step, mean power per link, mean outage per link and mean placement distance (steps) vs. ξ_{relay} for the four algorithms: OptExploreLim, OptAsYouGo, HeuExploreLim, and HeuAsYouGo. Unit of ξ_{relay} is actually mW, but in this figure it is shown in dBm. ξ_{relay}, when expressed in dBm, is equal to $10\log_{10}(\xi_{relay})$. In the Power plot, the HeuAsYouGo plot overlaps the OptAsYouGo plot, since the node power in the HeuAsYouGo algorithm was taken to be the same as the mean node power with the OptAsYouGo algorithm.

for a fixed ξ_{relay}, $c_{th}(r)$ increases with ξ_{out}. This happens because if the cost of outage increases, we cannot tolerate outage and place the relays close to each other. Note also that, $c_{th}(r)$ increases in r as stated in Algorithm 1.

B. Comparison Among Various Deployment Algorithms

Next, assuming a system model as described in Section II and assuming the parameter values as in Section V-A, we computed the mean cost per step, mean power per node, mean outage per link and mean placement distance (between successive relays) for four deployment algorithms presented so far[5]. Some of the results are shown in Figure 7. In order to make a fair comparison, *we used the mean power per node for OptAsYouGo as the fixed node transmit power for HeuAsYouGo, and the mean outage per link of OptAsYouGo as the pre-fixed target outage for HeuAsYouGo.* The following observations are from the plots in Figure 7.

1) Mean Placement Distance (see the top left panel of Figure 7): Pure as-you-go algorithms (OptAsYouGo, HeuAsYouGo) place relays sooner than the algorithms that explore forward (OptExploreLim, HeuExploreLim) before placing a relay (see Figure 7). This is as expected, since pure as-you-go algorithms do not have the advantage of exploring over several locations and then picking the best. A pure as-you-go approach tends to be cautious, and therefore tries to avoid a high outage by placing relays frequently. As ξ_{relay} (cost of a relay) increases, relays will be placed less frequently (according to Theorem 5).

2) Mean Outage per Link (see the top right panel of Figure 7): As ξ_{relay} increases, the mean outage per link increases because we will place fewer relays with higher inter-relay distances. Pure as-you-go algorithms have link outage probability comparable to explore-forward algorithms, but they place relays too frequently. We observe that the per-link outage of HeuAsYouGo is different from that of OptAsYouGo. This happens because whenever we place a node using HeuAsYouGo, the exact outage target is never met with equality. Also, the per-link outage may decrease with ξ_{relay} for HeuAsYouGo. As ξ_{relay} increases, the node power and the target outage (chosen from OptAsYouGo) increases in such a way that the per-link outage for HeuAsYouGo behaves in this fashion.

We have also observed that, as ξ_{out}, the penalty for outage, increases, the mean outage per link decreases. But that result has not been shown here.

3) Mean Power per Link (see the bottom left panel of Figure 7): Increasing ξ_{relay} will place relays less frequently, hence the transmit power increases. OptAsYouGo has smaller placement distance compared to OptExploreLim and HeuExploreLim, and hence it uses less power at each hop; we note,

[5]Note that, these computations were done on MATLAB; they did not involve any field deployment. Field experimentations were done only to validate the assumptions (such as independent shadowing assumption) and to compute the values of the parameters such as η and σ.

however, that OptAsYouGo places more relays, and, hence, could still end up using more power per step.

In the power plot, the HeuAsYouGo plot overlaps the OptAsYouGo plot, since the node power in the HeuAsYouGo algorithm was taken to be the same as the mean node power with the OptAsYouGo algorithm.

We have also seen that increasing ξ_{out} (the cost per unit outage) will lower outage and hence the per-node transmit power increases.

4) Network Cost Per Step (see the bottom right panel of Figure 7): The network cost per step is the optimal average cost per step; see (3). Cost increases with ξ_{relay} (see Figure 7) and ξ_{out}. OptAsYouGo has a larger cost than OptExploreLim and HeuExploreLim, owing to shorter links. *The average cost per step of HeuExploreLim is very close to OptExploreLim and cost of HeuAsYouGo is close to OptAsYouGo, even though the heuristic policies are not optimal.* However, we observed that this does not always happen. For example, for $\xi_{relay} = 0.1$ and $\xi_{out} = 1000$, we found that the average cost per step for OptAsYouGo and HeuAsYouGo are 1.3485 and 1.9581 respectively, and the average cost per step for OptExploreLim and HeuExploreLim are 0.9810 and 1.0537 respectively.

Discussion:

(i) HeuExploreLim and HeuAsYouGo appear to be attractive at the first sight because they are intuitive, easy to implement, and they do not require any channel model for given ξ_{out} and ξ_{relay}. But, they are suboptimal, and we do not have any performance guarantee (e.g., the optimality gap w.r.t. the optimal algorithms OptExploreLim and OptAsYouGo). Hence, if we know the radio propagation model (e.g., η and σ) exactly, and if ξ_{out} and ξ_{relay} are given, it is better to compute the optimal policies and then deploy according to them.

(ii) Note that, the mean number of measurements made per step for the pure as-you-go approach is 1, whereas it is $\frac{B}{\mathbb{E}(U)}$ under the explore-forward approach, where $\mathbb{E}(U)$ is the mean distance between successive relays. From the numerical results presented in this section, we find that, under the explore-forward approach, the mean number of measurements required will be at most 3, and can be even less than 2 depending on the situation. For applications that do not require rapid deployment, such as deployment in a large forest for monitoring purpose, this many measurements is affordable. *Hence, for the learning algorithms presented later in the paper, we will consider only explore-forward approach.*

(iii) More importantly, in practice the propagation environment will not be known, and, in order to solve the problem defined in (4), we need to choose ξ^*_{out} and ξ^*_{relay} while deploying (as explained in Theorem 1), if possible. But we cannot choose this pair if we do not have a prior knowledge of the propagation environment. Poor choice of ξ_{out} and ξ_{relay} might lead to violation of the constraints in the constrained problem defined in (4), or might result in a higher mean power per step compared to the optimal mean power per step under the constraints. Hence, we need to adapt ξ_{out} and ξ_{relay} as deployment progresses; this has been explained later in this paper. The adaptive algorithms use the structure of the optimal policy OptExploreLim.

APPENDIX D
OPTEXPLORELIMLEARNING: LEARNING WITH EXPLORE-FORWARD, FOR GIVEN ξ_{out} AND ξ_{relay}

Proof of Theorem 9:

Let us denote the shadowing random variable in the link between the potential locations located at distances $i\delta$ and $j\delta$ from the sink node by $W_{i,j}$. The sample space Ω associated with the deployment process is the collection of all ω (each ω corresponds to a fixed realization $\{w_{i,j} : i \geq 0, j \geq 0, i > j, A+1 \leq i-j \leq A+B\}$ of all possible shadowing random variables that might be encountered in the measurement process for deployment up to infinity). Let $\mathcal{F}$ be the Borel σ-algebra on Ω. Let $S_k = \sum_{i=1}^{k} U_i$ be the distance (in steps) of the k-th relay from the source, and $\mathcal{F}_k := \sigma\Big(\lambda^{(0)}; W_{i,j} : i \geq 0, j \geq 0, i > j, A+1 \leq i-j \leq A+B, i \leq S_{k-1}+A+B, j \leq S_{k-1}+A+B\Big)$. The sequence of σ-algebras $\mathcal{F}_k$ is increasing in k, and $\mathcal{F}_k$ captures the history of the deployment process up to the deployment of the k-th relay.

Note that, we can rewrite the update equation in Algorithm 6 as follows:

$$\lambda^{(k+1)} = \lambda^{(k)} + a_{k+1}\Big(f(\lambda^{(k)}) + N_{k+1}\Big)$$

where

$$f(\lambda) = \mathbb{E}_{\underline{\mathbb{W}}} \min_{u,\gamma} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda u\Big)$$

and

$$\begin{aligned} N_{k+1} \quad = \quad & \min_{u,\gamma} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda^{(k)} u\Big) - \\ & \mathbb{E}_{\underline{\mathbb{W}}} \min_{u,\gamma} \Big(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda^{(k)} u\Big) \end{aligned}$$

Note that, $(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda u)$ is a linearly decreasing function in λ. Hence, $\min_{u,\gamma}(\gamma + \xi_{out} Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda u)$ is a concave, strictly decreasing function in λ. The function $f(\lambda)$ is a nonnegative linear combination of concave, strictly decreasing functions of λ. Hence, $f(\lambda)$ is strictly decreasing, concave function of λ for $\lambda \in [0,\infty)$. Hence, $f(\lambda)$ is continuous in λ. Now, $f(0) > 0$ and $\lim_{\lambda\to\infty} f(\lambda) = -\infty$. *Hence,* $f(\lambda) = 0$ *will have a unique positive solution.*

Also, if we increase λ by an amount Δ, then we will have $(A+1)\Delta \leq |f(\lambda+\Delta) - f(\lambda)| \leq (A+B)\Delta$. *Hence,* $f(\cdot)$ *is Lipschitz continuous with Lipschitz constant* $(A+B)$.

Let us invoke four conditions from Chapter 2 of [21] as follows:

(i) $f(\cdot)$ is a Lipschitz continuous function.
(ii) $\sum_{k=1}^{\infty} a_k = \infty$, $\sum_{k=1}^{\infty} a_k^2 < \infty$.

(iii) $\{N_k\}_{k\geq 1}$ is a Martingale difference sequence w.r.t the sigma field $\mathcal{F}_k$ and $\mathbb{E}(|N_{k+1}|^2|\mathcal{F}_k) \leq K(1+|\lambda^{(k)}|^2)$ for some $K > 0$.

(iv) $\sup_{k\geq 1} |\lambda^{(k)}| < \infty$ almost surely.

By Theorem 2 (in Chapter 2) of [21], if the four conditions are satisfied, then $\lambda^{(k)}$ will almost surely converge to the unique zero of $f(\cdot)$. But, that unique zero is the optimal average cost per step λ^* which satisfies $f(\lambda^*) = 0$ (by Theorem 6). Hence, the problem reduces to checking the conditions (i)-(iv).

Since $f(\lambda)$ is Lipschitz continuous with Lipschitz constant $(A+B)$, condition (i) is satisfied. Condition (ii) is satisfied by the choice of a_k.

By definition of N_k, we have $\mathbb{E}_{\underline{W}}(N_{k+1}|\mathcal{F}_k) = \mathbb{E}_{\underline{W}}(N_{k+1}|\lambda^{(k)}) = 0$ (since shadowing is i.i.d. across links, the shadowing values encountered in the process of measurement for placing a new node are independent of the shadowing values encountered in the measurement process for deploying the previous nodes) which implies that $\{N_{k+1}\}_{k\geq 1}$ is a Martingale difference sequence w.r.t. $\mathcal{F}_k$. Now, since the conditional second moment is greater than conditional variance almost surely, we have (almost surely):

$$\begin{aligned}\mathbb{E}(|N_{k+1}|^2|\mathcal{F}_k) &\leq \mathbb{E}\Bigg(\Bigg(\min_{u,\gamma}(\gamma+\xi_{out}Q_{out}(u,\gamma,W_u) \\ &\quad +\xi_{relay}-\lambda^{(k)}u)\Bigg)^2|\mathcal{F}_k\Bigg)\end{aligned}$$

Now, we know that $\gamma \leq P_M$, $A+1 \leq u \leq A+B$, outage probability is always in $[0,1]$, and ξ_{out} and ξ_{relay} are fixed. Hence, $\mathbb{E}(|N_{k+1}|^2|\mathcal{F}_k)$ can be upper bounded by $K(1+|\lambda^{(k)}|^2)$ for some $K > 0$. Hence, condition (iii) is also satisfied. Condition (iv) is satisfied by the following lemma.

Lemma 3: For the iterates $\{\lambda^{(k)}\}_{k\geq 1}$ in (11), $\sup_{k\geq 1}|\lambda^{(k)}| < \infty$ almost surely.

Proof: Let us define K_0 to be the smallest integer such that $a_k(A+B) < 1$ for all $k \geq K_0$ (K_0 exists since $a_k \downarrow 0$). For any starting value $\lambda^{(0)}$, it is easy to find a positive real number d (depending on the value of $\lambda^{(0)}$) such that $\lambda^{(k)} \in [-d,d]$ for all $k \leq K_0$; this is easy to see because the node transmit power, node outage probability and placement distance for each node are bounded quantities.

Without loss of generality, we can take $d > P_M + \xi_{out} + \xi_{relay}$ where P_M is the maximum transmit power level of a node. We already have that $\lambda^{(k)} \in [-d,d]$ for all $k \leq K_0$. Now we will show that $\lambda^{(k)} \in [-d,d]$ for all $k \geq K_0$. To this end, let us assume, as our induction hypothesis, that $\lambda^{(k)} \in [-d,d]$ for some $k \geq K_0$. If we can show that $\lambda^{(k+1)} \in [-d,d]$, we will be done with the proof.

From the update equation (11), we can write that (using $(A+B) \geq u_{k+1} \geq 1$ and $0 \leq a_{k+1}u_{k+1} < 1$):

$$\begin{aligned}\lambda^{(k+1)} &\leq \lambda^{(k)} + a_{k+1}(P_M+\xi_{out}+\xi_{relay}-\lambda^{(k)}u_{k+1}) \\ &= (1-a_{k+1}u_{k+1})\lambda^{(k)} + a_{k+1}(P_M+\xi_{out}+\xi_{relay}) \\ &\leq (1-a_{k+1}u_{k+1})\lambda^{(k)} + a_{k+1}u_{k+1}(P_M+\xi_{out}+\xi_{relay}) \\ &\leq \max\{\lambda^{(k)}, P_M+\xi_{out}+\xi_{relay}\} \\ &\leq d\end{aligned}$$

On the other hand:

$$\begin{aligned}\lambda^{(k+1)} &\geq \lambda^{(k)} + a_{k+1}(0-\lambda^{(k)}u_{k+1}) \\ &= (1-a_{k+1}u_{k+1})\lambda^{(k)} \\ &\geq -(1-a_{k+1}u_{k+1})d \\ &\geq -d\end{aligned}$$

Hence, $\lambda^{(k+1)} \in [-d,d]$ and the lemma is proved. ■

Now, since conditions (i)-(iv) are satisfied, by Theorem 2, Chapter 2 of [21], $\lambda^{(k)} \rightarrow \lambda^*$ almost surely. □

Proof of Corollary 1: Suppose that, we choose $a_k = \frac{1}{\sum_{i=1}^k u_i}$ in Algorithm 6. Then, $\sum_{k=1}^{\infty} a_k \geq \sum_{k=1}^{\infty} \frac{1}{k(A+B)} = \infty$ almost surely and $\sum_{k=1}^{\infty} a_k^2 \leq \sum_{k=1}^{\infty} \frac{1}{k^2(A+1)^2} < \infty$ almost surely.

Now, with this step size,

$$\begin{aligned}\lambda_1 &= \lambda^{(0)} + a_1(\gamma_1 + \xi_{out}Q_{out}^{(1,0)} + \xi_{relay} - \lambda^{(0)}u_1) \\ &= \lambda^{(0)} + \frac{1}{u_1}(\gamma_1 + \xi_{out}Q_{out}^{(1,0)} + \xi_{relay} - \lambda^{(0)}u_1) \\ &= \frac{\gamma_1 + \xi_{out}Q_{out}^{(1,0)} + \xi_{relay}}{u_1}\end{aligned}$$

and, in general,

$$\begin{aligned}\lambda^{(k+1)} &= \lambda^{(k)} + a_{k+1}(\gamma_{k+1} + \xi_{out}Q_{out}^{(k+1,k)} + \xi_{relay} - \lambda^{(k)}u_{k+1}) \\ &= \lambda^{(k)} + \frac{(\gamma_{k+1} + \xi_{out}Q_{out}^{(k+1,k)} + \xi_{relay} - \lambda^{(k)}u_{k+1})}{\sum_{i=1}^{k+1} u_i} \\ &= \frac{\lambda^{(k)}\sum_{i=1}^{k} u_i + (\gamma_{k+1} + \xi_{out}Q_{out}^{(k+1,k)} + \xi_{relay})}{\sum_{i=1}^{k+1} u_i} \\ &= \frac{\sum_{i=1}^{k+1}(\gamma_i + \xi_{out}Q_{out}^{(i,i-1)} + \xi_{relay})}{\sum_{i=1}^{k+1} u_i}\end{aligned}$$

Hence, in (11) of Algorithm 6, we can replace $\lambda^{(k)} = \frac{\sum_{i=1}^{k}(\gamma_i + \xi_{out}Q_{out}^{(i,i-1)} + \xi_{relay})}{\sum_{i=1}^{k} u_i}$, and this proves the theorem.

APPENDIX E
OPTEXPLORELIMADAPTIVELEARNING WITH CONSTRAINT ON OUTAGE PROBABILITY AND RELAY PLACEMENT RATE

A. *Proof of Theorem 10*

Proof of the first statement: Let us assume that $\pi^*(\xi_{out}^*, \xi_{relay}^*)$ satisfies both constraints in (4) with equality for some $\xi_{out}^* > 0$, $\xi_{relay}^* > 0$, i.e., $\pi^*(\xi_{out}^*, \xi_{relay}^*)$ is an optimal policy for problem (4). Now, let us assume that there exists $\xi_{out}' \geq 0, \xi_{relay}' \geq 0$ satisfying (i) $(\lambda^*(\xi_{out}', \xi_{relay}'), \xi_{out}', \xi_{relay}') \in \mathcal{K}(\overline{q}, \overline{N})$, and (ii) $\frac{\overline{Q}_{out}^*(\xi_{out}', \xi_{relay}')}{\overline{U}^*(\xi_{out}', \xi_{relay}')} < \overline{q}$. We will show that this leads to a contradiction.

Let us consider the problem of minimizing the mean outage per step subject to a constraint $\frac{\overline{\Gamma}^*(\xi_{out}^*, \xi_{relay}^*)}{\overline{U}^*(\xi_{out}^*, \xi_{relay}^*)}$ on the mean power per step and a constraint $\frac{1}{\overline{U}^*(\xi_{out}^*, \xi_{relay}^*)} = \overline{N}$ on the mean number of relays per step. Clearly, by Theorem 1, $\pi^*(\xi_{out}^*, \xi_{relay}^*)$ is an optimal policy for this problem since it satisfies both constraints with equality. Note that, $\pi^*(\xi_{out}^*, \xi_{relay}^*)$ has a mean outage per step $\overline{q}$. But, we also see

that the policy $\pi^*(\xi'_{out}, \xi'_{relay})$ has the same mean power per step and a smaller mean number of relays per step compared to $\pi^*(\xi^*_{out}, \xi^*_{relay})$ (since $(\lambda^*(\xi'_{out}, \xi'_{relay}), \xi'_{out}, \xi'_{relay}) \in \mathcal{K}(\overline{q}, \overline{N})$), and has a *strictly* smaller mean outage per step compared to $\pi^*(\xi^*_{out}, \xi^*_{relay})$. This leads to a contradiction since $\pi^*(\xi^*_{out}, \xi^*_{relay})$ is an optimal policy for the problem of minimizing the mean outage per step subject to a constraint $\frac{\overline{\Gamma}^*(\xi^*_{out}, \xi^*_{relay})}{\overline{U}^*(\xi^*_{out}, \xi^*_{relay})}$ on the mean power per step and a constraint $\frac{1}{\overline{U}^*(\xi^*_{out}, \xi^*_{relay})} = \overline{N}$ on the mean number of relays per step.

Similarly, we can show a contradiction if, instead of assuming $\frac{\overline{Q}^*_{out}(\xi'_{out}, \xi'_{relay})}{\overline{U}(\xi'_{out}, \xi'_{relay})} < \overline{q}$, we had assumed $\frac{1}{\overline{U}(\xi'_{out}, \xi'_{relay})} < \overline{N}$.

Hence, the first statement is proved.

Proof of the second statement: This statement follows from the fact that for any $\xi_{relay} \geq 0$, $\pi^*(0, \xi_{relay})$ always places at a distance $(A + B)$ and uses the smallest power P_1, thereby incurring a mean outage per step equal to $\frac{\mathbb{E}_W Q_{out}(A+B, P_1, W)}{A+B}$. □

B. *Proof of Theorem 11*

Denote by $g(r, \gamma), r \in \{A + 1, A + 2, \cdots, A + B\}, \gamma \in \mathcal{S}$ the joint distribution of (U_k, Γ_k) when λ in (9) is replaced by $\lambda^*(\xi_{out}, \xi_{relay})$, i.e., when deployment is done using the OptExploreLim algorithm (Algorithm 3).

Let us assume that $g(r, \gamma)$ is continuous in both ξ_{out} and ξ_{relay} (we will prove this assertion in Lemma 4 at the end of the proof of the theorem). By Lemma 4, the mean placement distance $\overline{U}^*(\xi_{out}, \xi_{relay}) = \sum_{r=A+1}^{A+B} \sum_{\gamma \in \mathcal{S}} r g(r, \gamma)$ is continuous in ξ_{out} and ξ_{relay}. Similarly, the mean power per link $\overline{\Gamma}^*(\xi_{out}, \xi_{relay}) = \sum_{r=A+1}^{A+B} \sum_{\gamma \in \mathcal{S}} \gamma g(r, \gamma)$ is continuous in ξ_{out} and ξ_{relay}.

Let us denote by $\lambda^*(\xi_{out}, \xi_{relay})$ the optimal average cost per step for the problem in (3), for given ξ_{out} and ξ_{relay}. By Renewal-Reward Theorem,

$$\lambda^*(\xi_{out}, \xi_{relay}) = \frac{\overline{\Gamma}^*(\xi_{out}, \xi_{relay}) + \xi_{out} \overline{Q}^*_{out}(\xi_{out}, \xi_{relay}) + \xi_{relay}}{\overline{U}^*(\xi_{out}, \xi_{relay})}$$

Since $\lambda^*(\xi_{out}, \xi_{relay})$ is continuous in ξ_{out} and ξ_{relay} (by Theorem 4), we conclude that $\overline{Q}^*_{out}(\xi_{out}, \xi_{relay})$ is continuous in ξ_{out} and ξ_{relay}. Hence, $\frac{\overline{\Gamma}^*(\xi_{out}, \xi_{relay})}{\overline{U}^*(\xi_{out}, \xi_{relay})}$, $\frac{\overline{Q}^*_{out}(\xi_{out}, \xi_{relay})}{\overline{U}^*(\xi_{out}, \xi_{relay})}$ and $\frac{1}{\overline{U}^*(\xi_{out}, \xi_{relay})}$ are continuous in ξ_{out} and ξ_{relay}. Hence, the theorem is proved. □

Lemma 4: Under Assumption 2, $g(r, \gamma)$ is continuous in ξ_{out} and ξ_{relay}.

Proof: Let us fix any $r \in \{A + 1, \cdots, A + B\}$ and any $\gamma \in \mathcal{S}$. We will show that $g(r, \gamma)$ is continuous in ξ_{out}. The continuity of $g(r, \gamma)$ w.r.t. ξ_{relay} will follow the same line of arguments.

Consider any sequence $\{\xi_n\}_{n \geq 1}$ such that $\xi_n \to \xi_{out}$. Let us denote the joint probability distribution of placement distance and node transmit power by $g_n(r, \gamma)$, if the cost per unit outage is ξ_n and if OptExploreLim is used in the deployment process. We will show that $g_n(r, \gamma) \to g(r, \gamma)$ as $n \to \infty$.

Let us define the sets $\mathcal{E}_{\gamma'} = \Big\{ \underline{w} : \gamma + \xi_{out} Q_{out}(r, \gamma, w_r) < \gamma' + \xi_{out} Q_{out}(r, \gamma', w_r) \Big\}$ and $\mathcal{E}_{u,\gamma'} = \Big\{ \underline{w} : \gamma + \xi_{out} Q_{out}(r, \gamma, w_r) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) r < \gamma' + \xi_{out} Q_{out}(u, \gamma', w_u) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) u \Big\}$.

In state $\underline{w}$, the OptExploreLim algorithm (Algorithm 3) will place the next relay at distance r and decide power level γ if $\underline{w} \in \mathcal{E}_{\gamma'}$ for all $\gamma' \neq \gamma, \gamma' \in \mathcal{S}$ and if $\underline{w} \in \mathcal{E}_{u,\gamma'}$ for all $u \neq r, \gamma' \in \mathcal{S}$.

Let us define $\mathcal{E} = \cap_{\gamma' \neq \gamma} \mathcal{E}_{\gamma'} \cap_{u \neq r, \gamma' \in \mathcal{S}} \mathcal{E}_{u,\gamma'}$.

Note that, $g(r, \gamma) = \mathbb{P}(\mathcal{E}) = \mathbb{E}(\mathbb{I}_{\mathcal{E}})$, where $\mathbb{I}$ denotes the indicator function, and the expectation is over the joint distribution of the shadowing vector $\underline{W}$ (shadowing random variables from B locations).

Now, for any $\gamma' \neq \gamma$, we have $\mathbb{P}\Big(\gamma + \xi_{out} Q_{out}(r, \gamma, W_r) = \gamma' + \xi_{out} Q_{out}(r, \gamma', W_r)\Big) = 0$. Also, $\mathbb{P}\Big(\gamma + \xi_{out} Q_{out}(r, \gamma, W_r) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) r = \gamma' + \xi_{out} Q_{out}(u, \gamma', W_u) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) u\Big) = 0$ if $\gamma' \in \mathcal{S}$, $u \neq r$. These two assertions follow from Assumption 2 and the fact that $Q_{out}(r, \gamma, w)$ is continuous in w. Hence, we discard these zero probability events in our analysis and safely assume that:

- For $\gamma' \neq \gamma$, the complement $\overline{\mathcal{E}_{\gamma'}}$ has the same expression as $\mathcal{E}_{\gamma'}$ except that the $<$ sign is replaced by $>$ sign.
- For $\gamma' \in \mathcal{S}$, $u \neq r$, $\overline{\mathcal{E}_{u,\gamma'}}$ has the same expression as $\mathcal{E}_{u,\gamma'}$ except that the $<$ sign is replaced by $>$ sign.

Now, consider any sequence $\{\xi_n\}_{n \geq 1}$ such that $\xi_n \to \xi_{out}$. Let $\mathcal{E}^{(n)}_{\gamma'}$, $\mathcal{E}^{(n)}_{u,\gamma'}$ and $\mathcal{E}^{(n)}$ be the sets obtained where we replace ξ_{out} by ξ_n in the expressions of the sets $\mathcal{E}_{\gamma'}$, $\mathcal{E}_{u,\gamma'}$ and $\mathcal{E}$ respectively. Clearly, we can make similar claims for $\mathcal{E}^{(n)}_{\gamma'}$, $\mathcal{E}^{(n)}_{u,\gamma'}$ for any $n \geq 1$.

Recall that, $g(r, \gamma) = \mathbb{P}(\mathcal{E}) = \mathbb{E}(\mathbb{I}_{\mathcal{E}})$. Clearly, if we can show that $\mathbb{E}(\mathbb{I}_{\mathcal{E}^{(n)}}) \to \mathbb{E}(\mathbb{I}_{\mathcal{E}})$, the lemma will be proved.

Claim 1: $\mathbb{I}_{\mathcal{E}^{(n)}_{u,\gamma'}} \to \mathbb{I}_{\mathcal{E}_{u,\gamma'}}$ almost surely as $n \to \infty$, for $u \neq r, \gamma' \in \mathcal{S}$. Also, $\mathbb{I}_{\mathcal{E}^{(n)}_{\gamma'}} \to \mathbb{I}_{\mathcal{E}_{\gamma'}}$ almost surely as $n \to \infty$, for $\gamma' \neq \gamma$.

Proof: Suppose that, for some value of $\underline{w}$, $\mathbb{I}_{\mathcal{E}_{u,\gamma'}}(\underline{w}) = 1$, i.e., $\gamma + \xi_{out} Q_{out}(r, \gamma, w_r) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) r < \gamma' + \xi_{out} Q_{out}(u, \gamma', w_u) + \xi_{relay} - \lambda^*(\xi_{out}, \xi_{relay}) u$. Now, by Theorem 4, $\lambda^*(\xi_{out}, \xi_{relay})$ is continuous in ξ_{out} and ξ_{relay}. Hence, there exists an integer n_0 sufficiently large such that for all $n > n_0$, we have $\gamma + \xi_n Q_{out}(r, \gamma, w_r) + \xi_{relay} - \lambda^*(\xi_n, \xi_{relay}) r < \gamma' + \xi_n Q_{out}(u, \gamma', w_u) + \xi_{relay} - \lambda^*(\xi_n, \xi_{relay}) u$, i.e., $\mathbb{I}_{\mathcal{E}^{(n)}_{u,\gamma'}}(\underline{w}) = 1$ for all $n > n_0$. Hence, $\mathbb{I}_{\mathcal{E}^{(n)}_{u,\gamma'}}(\underline{w}) \to \mathbb{I}_{\mathcal{E}_{u,\gamma'}}(\underline{w})$ if $\mathbb{I}_{\mathcal{E}_{u,\gamma'}}(\underline{w}) = 1$. Similar argument works when $\mathbb{I}_{\mathcal{E}_{u,\gamma'}}(\underline{w}) = 0$. Hence, the first part of the claim is proved.

The second part of the claim is proved in a similar way. ■

Note that, $\mathbb{I}_{\mathcal{E}^{(n)}} = \prod_{\gamma' \neq \gamma} \mathbb{I}_{\mathcal{E}^{(n)}_{\gamma'}} \prod_{u \neq r, \gamma' \in \mathcal{S}} \mathbb{I}_{\mathcal{E}^{(n)}_{u,\gamma'}}$. By Claim 1, $\mathbb{I}_{\mathcal{E}^{(n)}} \to \mathbb{I}_{\mathcal{E}}$ almost surely. Since indicator functions always take values in the set $\{0, 1\}$, we have $\mathbb{E}(\mathbb{I}_{\mathcal{E}^{(n)}}) \to \mathbb{E}(\mathbb{I}_{\mathcal{E}})$

$$
\begin{aligned}
\lambda^{(k)} &= \lambda^{(k-1)} + a_k \min_{u,\gamma}\Big(\gamma + \xi_{out}^{(k-1)} Q_{out}(u,\gamma,w_u) + \xi_{relay}^{(k-1)} - \lambda^{(k-1)} u\Big) \\
\xi_{out}^{(k)} &= \xi_{out}^{(k-1)} + b_k \lim_{\beta\downarrow 0} \frac{\Lambda_{[0,A_2]}\Big(\xi_{out}^{(k-1)} + \beta(Q_{out}(u_k,\gamma_k,w_{u_k}) - \overline{q}u_k)\Big) - \xi_{out}^{(k-1)}}{\beta} + o(b_k) \\
&= \xi_{out}^{(k-1)} + a_k\Big(\frac{b_k}{a_k}\Big(\lim_{\beta\downarrow 0} \frac{\Lambda_{[0,A_2]}\Big(\xi_{out}^{(k-1)} + \beta(Q_{out}(u_k,\gamma_k,w_{u_k}) - \overline{q}u_k)\Big) - \xi_{out}^{(k-1)}}{\beta} + \frac{o(b_k)}{b_k}\Big)\Big) \\
\xi_{relay}^{(k)} &= \xi_{relay}^{(k-1)} + b_k \lim_{\beta\downarrow 0} \frac{\Lambda_{[0,A_3]}\Big(\xi_{relay}^{(k-1)} + \beta(1 - \overline{N}u_k)\Big) - \xi_{relay}^{(k-1)}}{\beta} + o(b_k) \\
&= \xi_{relay}^{(k-1)} + a_k\Big(\frac{b_k}{a_k}\Big(\lim_{\beta\downarrow 0} \frac{\Lambda_{[0,A_3]}\Big(\xi_{relay}^{(k-1)} + \beta(1 - \overline{N}u_k)\Big) - \xi_{relay}^{(k-1)}}{\beta} + \frac{o(b_k)}{b_k}\Big)\Big) \qquad (22)
\end{aligned}
$$

by Dominated Convergence Theorem.

Hence, the lemma is proved. ∎

C. *Proof of Theorem 12*

Let us denote the shadowing random variable in the link between the potential locations located at distances $i\delta$ and $j\delta$ from the sink node by $W_{i,j}$. The sample space Ω associated with the deployment process is the collection of all ω (each ω corresponds to a fixed realization $\{w_{i,j} : i \geq 0, j \geq 0, i > j, A+1 \leq i-j \leq A+B\}$ of all possible shadowing random variables that might be encountered in the measurement process for deployment up to infinity). Let $\mathcal{F}$ be the Borel σ-algebra on Ω. Let $S_k = \sum_{i=1}^{k} U_i$ be the distance (in steps) of the k-th relay from the source, and $\mathcal{F}_k := \sigma\Big((\lambda^{(0)}, \xi_{out}^{(0)}, \xi_{relay}^{(0)}); W_{i,j} : i \geq 0, j \geq 0, i > j, A+1 \leq i-j \leq A+B, i \leq S_{k-1}+A+B, j \leq S_{k-1}+A+B\Big)$. The sequence of σ-algebras $\mathcal{F}_k$ is increasing in k, and $\mathcal{F}_k$ captures the history of the deployment process up to the deployment of the k-th relay.

Let us recall the outline of the proof of Theorem 12 in Section VII-B.

*1) **Almost sure boundedness of the** $\lambda^{(k)}$ **iterates**:*

Lemma 5: The iterates $\{\lambda^{(k)}\}_{k\geq 1}$ in (13) are bounded almost surely.

Proof: Let us define K_0 to be the smallest integer such that $a_k(A+B) < 1$ for all $k \geq K_0$ (K_0 exists since $a_k \downarrow 0$). For any starting value $\lambda^{(0)}$, it is easy to find a positive real number d (depending on the value of $\lambda^{(0)}$) such that $\lambda^{(k)} \in [-d, d]$ for all $k \leq K_0$; this is easy to see because $\xi_{out}^{(k)} \in [0, A_2]$, $\xi_{relay}^{(k)} \in [0, A_3]$ for all k, and the node transmit power, node outage probability and placement distance for each node are bounded quantities.

Without loss of generality, we can take $d > P_M + A_2 + A_3$ where P_M is the maximum transmit power level of a node. We already have that $\lambda^{(k)} \in [-d, d]$ for all $k \leq K_0$. Now we will show that $\lambda^{(k)} \in [-d, d]$ for all $k \geq K_0$. To this end, let us assume, as our induction hypothesis, that $\lambda^{(k)} \in [-d, d]$ for some $k \geq K_0$. If we can show that $\lambda^{(k+1)} \in [-d, d]$, we will be done with the proof.

From the update equation (13), we can write that (using $(A+B) \geq u_{k+1} \geq 1$ and $0 \leq a_{k+1}u_{k+1} < 1$):

$$
\begin{aligned}
\lambda^{(k+1)} &\leq \lambda^{(k)} + a_{k+1}(P_M + A_2 + A_3 - \lambda^{(k)} u_{k+1}) \\
&= (1 - a_{k+1}u_{k+1})\lambda^{(k)} + a_{k+1}(P_M + A_2 + A_3) \\
&\leq (1 - a_{k+1}u_{k+1})\lambda^{(k)} + a_{k+1}u_{k+1}(P_M + A_2 + A_3) \\
&\leq \max\{\lambda^{(k)}, P_M + A_2 + A_3\} \\
&\leq d
\end{aligned}
$$

On the other hand:

$$
\begin{aligned}
\lambda^{(k+1)} &\geq \lambda^{(k)} + a_{k+1}(0 - \lambda^{(k)} u_{k+1}) \\
&= (1 - a_{k+1}u_{k+1})\lambda^{(k)} \\
&\geq -(1 - a_{k+1}u_{k+1})d \\
&\geq -d
\end{aligned}
$$

Hence, $\lambda^{(k+1)} \in [-d, d]$ and the lemma is proved. ∎

*2) **Analyzing the Faster Time-Scale Iteration of** $\lambda^{(k)}$:* Let us denote by $\lambda^*(\xi_{out}, \xi_{relay})$ the optimal average cost per step for the problem in (3), for given ξ_{out} and ξ_{relay}.

Lemma 6: For Algorithm 8, we have $(\lambda^{(k)}, \xi_{out}^{(k)}, \xi_{relay}^{(k)}) \to \{(\lambda^*(\xi_{out}, \xi_{relay}), \xi_{out}, \xi_{relay}) : (\xi_{out}, \xi_{relay}) \in [0, A_2] \times [0, A_3]\}$ and $\lim_{k\to\infty} |\lambda^{(k)} - \lambda^*(\xi_{out}^{(k)}, \xi_{relay}^{(k)})| = 0$ almost surely.

Proof: We follow the proof of Lemma 1, Chapter 6 of [21].

Using the first order Taylor series expansion of the function $\Lambda_{[0,A_2]}(\cdot)$, and using the fact that $\Lambda_{[0,A_2]}(\xi_{out}^{(k-1)}) = \xi_{out}^{(k-1)}$ (since $\xi_{out}^{(k-1)} \in [0, A_2]$), the update equation (13) can be rewritten as (22).

Consider the update equation for ξ_{relay} in (22). Note that:

$$
\begin{aligned}
&\lim_{\beta\downarrow 0} \frac{\Lambda_{[0,A_3]}\Big(\xi_{relay}^{(k-1)} + \beta(1 - \overline{N}u_k)\Big) - \xi_{relay}^{(k-1)}}{\beta} \\
&= (1 - \overline{N}u_k)\mathbb{I}\{0 < \xi_{relay}^{(k-1)} < A_3\} \\
&+ (1 - \overline{N}u_k)^{+}\mathbb{I}\{\xi_{relay}^{(k-1)} = 0\} \\
&- (1 - \overline{N}u_k)^{-}\mathbb{I}\{\xi_{relay}^{(k-1)} = A_3\}
\end{aligned}
$$

$$
\begin{aligned}
\xi_{out}^{(k)} &= \Lambda_{\mathcal{G}}\left(\xi_{out}^{(k-1)} + b_k\left(Q_{out}(U_k,\Gamma_k,W_{U_k}) - \overline{q}U_k\right)\right) \\
&= \Lambda_{\mathcal{G}}\Bigg(\xi_{out}^{(k-1)} + b_k\Bigg(\underbrace{\overline{Q}^*_{out}(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) - \overline{q}\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}_{:=f_1(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})} \\
&\quad + \underbrace{\overline{Q}_{out}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) - \overline{q}\overline{U}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) - f_1(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}_{:=g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})} \\
&\quad + \underbrace{Q_{out}(U_k,\Gamma_k,W_{U_k}) - \overline{q}U_k - \left(\overline{Q}_{out}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) - \overline{q}\overline{U}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\right)}_{:=M_1^{(k)}}\Bigg)\Bigg) \\
&= \Lambda_{\mathcal{G}}\left(\xi_{out}^{(k-1)} + b_k\left(f_1(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + M_1^{(k)}\right)\right) \\
\xi_{relay}^{(k)} &= \Lambda_{\mathcal{G}}\left(\xi_{out}^{(k-1)} + b_k\left(1 - \overline{N}U_k\right)\right) \\
&= \Lambda_{\mathcal{G}}\Bigg(\xi_{relay}^{(k-1)} + b_k\Bigg(\underbrace{1 - \overline{NU}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}_{:=f_2(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})} \\
&\quad + \underbrace{1 - \overline{NU}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) - f_2(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}_{:=g_2(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})} \\
&\quad + \underbrace{1 - \overline{N}U_k - \left(1 - \overline{NU}(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\right)}_{:=M_2^{(k)}}\Bigg)\Bigg) \\
&= \Lambda_{\mathcal{G}}\left(\xi_{relay}^{(k-1)} + b_k\left(f_2(\xi_{relay}^{(k-1)},\xi_{relay}^{(k-1)}) + g_2(\lambda^{(k-1)},\xi_{relay}^{(k-1)},\xi_{relay}^{(k-1)}) + M_2^{(k)}\right)\right) \qquad (23)
\end{aligned}
$$

where $x^+ = \max\{x,0\}$ and $x^- = -\min\{x,0\}$. A similar expression holds for the $\xi_{out}^{(k)}$ update. Since $Q_{out}(\cdot,\cdot,\cdot)$ and u_k are bounded quantities, and since $\lim_{k\to 0}\frac{b_k}{a_k} = 0$, we have:

$$\lim_{k\to\infty}\left(\frac{b_k}{a_k}\left(\lim_{\beta\downarrow 0}\left(\Lambda_{[0,A_2]}\left(\xi_{out}^{(k-1)} + \beta(Q_{out}(u_k,\gamma_k,w_{u_k}) - \overline{q}u_k)\right) - \xi_{out}^{(k-1)}\right)/\beta + \frac{o(b_k)}{b_k}\right)\right) = 0$$

and

$$\lim_{k\to\infty}\left(\frac{b_k}{a_k}\left(\lim_{\beta\downarrow 0}\frac{\Lambda_{[0,A_3]}\left(\xi_{relay}^{(k-1)} + \beta(1-\overline{N}u_k)\right) - \xi_{relay}^{(k-1)}}{\beta} + \frac{o(b_k)}{b_k}\right)\right) = 0$$

Now, note that, the function $f(\lambda,\xi_{out},\xi_{relay}) = \mathbb{E}_{\underline{\mathbb{W}}}\min_{u,\gamma}\left(\gamma + \xi_{out}Q_{out}(u,\gamma,W_u) + \xi_{relay} - \lambda u\right)$ is Lipschitz continuous in all arguments, and the o.d.e. $\dot{\lambda}(t) = f(\lambda(t),\xi_{out},\xi_{relay})$ has a unique globally asymptotically stable equilibrium $\lambda^*(\xi_{out},\xi_{relay})$ for any $\xi_{out}\geq 0$, $\xi_{relay}\geq 0$ (see the proof of Theorem 9). The quantity $\lambda^*(\xi_{out},\xi_{relay})$ is Lipschitz continuous in ξ_{out} and ξ_{relay}. Also by Lemma 5 and the projection operation in the slower timescale, the iterates are bounded almost surely.

Hence, by a similar argument as in the proof of Lemma 1, Chapter 6 of [21] (or by Corollary 2.1 of [24]), and by using Theorem 6 and Theorem 9, $(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)})$ converges to the internally chain transitive invariant sets of the o.d.e. $\dot{\lambda}(t) = f(\lambda(t),\xi_{out}(t),\xi_{relay}(t))$, $\dot{\xi}_{out}(t) = 0$, $\dot{\xi}_{relay}(t) = 0$. Hence, $(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)}) \to \{(\lambda^*(\xi_{out},\xi_{relay}),\xi_{out},\xi_{relay}) : (\xi_{out},\xi_{relay}) \in [0,A_2]\times[0,A_3]\}$ and $\lim_{k\to\infty}|\lambda^{(k)} - \lambda^*(\xi_{out}^{(k)},\xi_{relay}^{(k)})| = 0$. Hence, the lemma is proved. ■

Remark: Lemma 6 tells us that the faster time-scale iterate $\lambda^{(k)}$ closely tracks $\lambda^*(\xi_{out}^{(k)},\xi_{relay}^{(k)})$. But it is important to note that this lemma does not guarantee the convergence of the slower timescale iterates to a single point in the two-dimensional Euclidean plane.

*3) **The slower timescale iteration**:* Let us recall the notation $\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})$, $\overline{U}(\lambda,\xi_{out},\xi_{relay})$, $\overline{Q}^*_{out}(\xi_{out},\xi_{relay})$ and $\overline{U}^*(\xi_{out},\xi_{relay})$ as defined in Section IV-B. Let us also recall the update equation (13) in Algorithm 8. We will analyze the slower timescale update equations as a projected stochastic approximation (see Equation 5.3.1 of [23]).

Let us denote by $\mathcal{G}$ the compact subset $[0,A_2]\times[0,A_3]$ of the Euclidean space. Clearly, the set $\mathcal{G}$ can be defined by the following set of constraints on the variables ξ_{out} and ξ_{relay}:

$$-\xi_{out}\leq 0, \xi_{out}\leq A_2, -\xi_{relay}\leq 0, \xi_{relay}\leq A_3 \qquad (24)$$

We rewrite the slower timescale update equations in (13) as (23). Note that, the functions $f_1(\xi_{out},\xi_{relay})$, $f_2(\xi_{out},\xi_{relay})$, $g_1(\lambda,\xi_{out},\xi_{relay})$, and $g_2(\lambda,\xi_{out},\xi_{relay})$ have been defined in (23). The quantities $M_1^{(k)}$ and $M_2^{(k)}$ are two zero mean Martingale difference noise sequences w.r.t. $\mathcal{F}_{k-1}$; this can be seen as follows. Since shadowing

is i.i.d. across links, the shadowing values encountered in the process of measurement for placing the k-th node are independent of the history of the process up to the placement of node $(k-1)$. Hence, $\mathbb{E}_{\underline{W}}\left(M_1^{(k)}|\mathcal{F}_{k-1}\right) = \mathbb{E}_{\underline{W}}\left(M_1^{(k)}|(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\right) = 0$ and $\mathbb{E}_{\underline{W}}\left(M_2^{(k)}|\mathcal{F}_{k-1}\right) = \mathbb{E}_{\underline{W}}\left(M_2^{(k)}|(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\right) = 0$.

The update is done as follows. We compute $\tilde{\xi}_{relay}^{(k)} = \xi_{relay}^{(k-1)} + b_k\Big(f_2(\xi_{relay}^{(k-1)},\xi_{relay}^{(k-1)}) + g_2(\lambda^{(k-1)},\xi_{relay}^{(k-1)},\xi_{relay}^{(k-1)}) + M_2^{(k)}\Big)$ and compute $\xi_{relay}^{(k)} = \Lambda_{[0,A_3]}(\tilde{\xi}_{relay}^{(k)})$. We compute $\xi_{out}^{(k)}$ in a similar fashion. *Hence, projection onto the set $\mathcal{G}$ is nothing but coordinatewise projection.*

Note that, (23) is in the same form as the standard projected stochastic approximation (Equation 5.3.1 *of [23]). In order to show that the iterates in (23) converge to the right set, we will make use of Theorem* 5.3.1 *from [23]. To use this theorem, we need to check whether (23) satisfies five conditions from [23], i.e.,* A5.1.3*,* A5.1.4*,* A5.1.5*,* A5.3.1. *and* A5.3.2*. This is done in the next subsection.* □

4) ***Checking the five conditions from [23]*****:** Before checking the five conditions, we will present a lemma that will be useful for checking one condition.

Lemma 7: Suppose that Assumption 2 holds. Under the decision rule given by (9), the mean power per step $\frac{\overline{\Gamma}(\lambda,\xi_{out},\xi_{relay})}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$, mean number of relays per step $\frac{1}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$ and mean outage per step $\frac{\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$ are continuous in λ, ξ_{out} and ξ_{relay}.

Proof: The proof is similar to that of Theorem 11. ■

Now, we will check that conditions $A5.1.3$, $A5.1.4$, $A5.1.5$, $A5.3.1$. and $A5.3.2$ from [23] are satisfied.

Checking Condition A5.1.3*:* This condition requires that $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ are continuous functions. This condition is satisfied as a consequence of Theorem 11. □

Checking Condition A5.1.4*:* This condition is satisfied since $b_k > 0$, $b_k \to 0$ as $k \to \infty$ and $\sum_{k=1}^{\infty} b_k = \infty$. □

Checking Condition A5.1.5*:* This condition requires that $\lim_{k\to\infty} g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) = 0$ and $\lim_{k\to\infty} g_2(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) = 0$ almost surely, and that the sequences $g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})$ and $g_2(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})$ are bounded almost surely.

By Lemma 5, we can find an interval $[-d,d]$ such that $(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)})$ lies inside the compact set $[-d,d]\times[0,A_2]\times[0,A_3]$ for all $k \geq 1$ almost surely.

Note that, $\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})$ is continuous in each argument (by Lemma 7). Hence, $\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})$ is uniformly continuous over the compact set $[-d,d]\times[0,A_2]\times[0,A_3]$ and similarly $\overline{U}(\lambda,\xi_{out},\xi_{relay})$ is uniformly continuous over the compact set $[-d,d]\times[0,A_2]\times[0,A_3]$.

Now, by Lemma 6, the Euclidean distance between $(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)})$ and $(\lambda^*(\xi_{out}^{(k)},\xi_{relay}^{(k)}),\xi_{out}^{(k)},\xi_{relay}^{(k)})$ converges to 0 almost surely as $k\to\infty$. Hence, by uniform continuity, we conclude that $\lim_{k\to\infty}|\overline{Q}_{out}(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)}) - \overline{Q}_{out}(\lambda^*(\xi_{out}^{(k)},\xi_{relay}^{(k)}),\xi_{out}^{(k)},\xi_{relay}^{(k)})| = 0$ and $\lim_{k\to\infty}|\overline{U}(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)}) - \overline{U}(\lambda^*(\xi_{out}^{(k)},\xi_{relay}^{(k)}),\xi_{out}^{(k)},\xi_{relay}^{(k)})| = 0$ almost surely. Hence, $\lim_{k\to\infty} g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) = 0$ and $\lim_{k\to\infty} g_2(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) = 0$ almost surely.

Also, $g_1(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)})$ and $g_2(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)})$ are uniformly bounded across $k \geq 1$, since the outage probabilities and placement distances are bounded quantities.

Checking Condition A5.3.1*:* This condition requires that $\mathcal{G} = [0,A_2]\times[0,A_3]$ is the closure of its interior, which is true in our problem. It also requires that the L.H.S. of each constraint inequality in (24) is continuously differentiable, which is also true in our problem.

Note that, the L.H.S. of each constraint inequality in (24) is a function of ξ_{out} and ξ_{relay}. Condition $A5.3.1$ of [23] needs that for each point on the boundary of $\mathcal{G}$, the gradients of the functions (in the L.H.S. of (24)) corresponding to the active constraints are linearly independent. Note that on each point of the boundary of $\mathcal{G}$, at most two constraints can be simultaneously active (see (24)). If there are exactly two active constraints, one will be for ξ_{out} and the other one will be for ξ_{relay}. Clearly, the gradients (with respect to the tuple (ξ_{out},ξ_{relay})) of the active constraint(s) at any boundary point of $\mathcal{G}$ are orthogonal, and hence linearly independent. □

Checking Condition A5.3.2*:* Let $m(t) := \sup\{n \geq 1 : \sum_{i=1}^{n} b_i \leq t\}$. We have to show that, there exists a $T > 0$ such that for any $\epsilon > 0$, we have $\lim_{n\to\infty}\mathbb{P}\Big(\sup_{j\geq n}\max_{t\leq T}|\sum_{i=m(jT)}^{m(jT+t)-1} b_i M_1^{(i)}| > \epsilon\Big) = 0$ and $\lim_{n\to\infty}\mathbb{P}\Big(\sup_{j\geq n}\max_{t\leq T}|\sum_{i=m(jT)}^{m(jT+t)-1} b_i M_2^{(i)}| > \epsilon\Big) = 0$. We will prove the first result, for any $T > 0$ and any $\epsilon > 0$. Let us define the event $E_j := \{\max_{t\leq T}|\sum_{i=m(jT)}^{m(jT+t)-1} b_i M_1^{(i)}| > \epsilon\}$. Hence,

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}\Big(\sup_{j\geq n}\max_{t\leq T}|\sum_{i=m(jT)}^{m(jT+t)-1} b_i M_1^{(i)}| > \epsilon\Big)\\
&= \lim_{n\to\infty}\mathbb{P}\Big(\cup_{j\geq n} E_j\Big)\\
&= \mathbb{P}\Big(\cap_{n\geq 1}\cup_{j\geq n} E_j\Big)\\
&= \mathbb{P}\Big(\limsup_{n\to\infty} E_n\Big)
\end{aligned}$$

where the second equality follows from the continuity of probability.

Now, $\mathbb{P}(E_n) \leq \frac{\mathbb{E}|\sum_{i=m(nT)}^{m(nT+T)-1} b_i M_1^{(i)}|^2}{\epsilon^2}$ (by Doob's inequality for Martingales). Since, $|M_1^{(i)}| \leq C$ for some $C > 0$ (since outage probability and placement distance are two bounded quantities; see the expression for $M_1^{(i)}$ in (23)),

$$\xi_{out}^{(k)} = \Lambda_{\mathcal{G}}\left(\xi_{out}^{(k-1)} + b_k\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\frac{\left(f_1(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + g_1(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + M_1^{(k)}\right)}{\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}\right)$$

$$\xi_{relay}^{(k)} = \Lambda_{\mathcal{G}}\left(\xi_{relay}^{(k-1)} + b_k\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})\frac{\left(f_2(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + g_2(\lambda^{(k-1)},\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) + M_2^{(k)}\right)}{\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}\right) \quad (25)$$

the above quantity can be upper-bounded by $\mathbb{P}(E_n) \leq \frac{2C^2\sum_{i=m(nT)}^{m(nT+T)-1} b_i^2}{\epsilon^2}$. Hence, $\sum_{n=1}^{\infty}\mathbb{P}(E_n) \leq \frac{2C^2\sum_{i=1}^{\infty} b_i^2}{\epsilon^2} < \infty$. Hence, by Borel-Cantelli lemma, $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0$, which completes checking Condition $A5.3.2$ of [23]. □

5) **Finishing the Proof of Theorem 12***:* Now, we will invoke Theorem 5.3.1 from [23] to complete the proof.

Let us rewrite (23) as (25). Note that, $A+1 \leq \overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)}) \leq A+B$. Hence, if we use step size $b_k\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})$ and use the modified functions like $\frac{f_1(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}{\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}$ as in (25), the conditions checked in the previous subsection will still hold. This is evident from the fact that, once we know $\xi_{out}^{(k-1)}$ and $\xi_{relay}^{(k-1)}$, $\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})$ becomes a deterministic quantity, and that the randomness in the computation of the new iterates $\xi_{out}^{(k)}$ and $\xi_{relay}^{(k)}$ comes from the random shadowing in the links measured in the process of deploying the k-th node. Hence, $\frac{M_1^{(k)}}{\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}$ and $\frac{M_2^{(k)}}{\overline{U}^*(\xi_{out}^{(k-1)},\xi_{relay}^{(k-1)})}$ are also Martingale difference sequences. It is easy to check conditions $A5.1.3$, $A5.1.5$ and $A5.3.2$ for (25), and the condition in $A5.1.4$ is satisfied almost surely.

Hence, from now on, let us consider the slower timescale iteration (25).

For the function $h(\xi_{out},\xi_{relay}) := \left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right) = \left(\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})} - \overline{q}, \frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})} - \overline{N}\right)$, let us define the map:

$$\overline{\Lambda}_{\mathcal{G}}(h(\xi_{out},\xi_{relay})) = \lim_{0<\beta\to 0}\frac{\Lambda_{\mathcal{G}}\left((\xi_{out},\xi_{relay}) + \beta h(\xi_{out},\xi_{relay}))\right) - (\xi_{out},\xi_{relay})}{\beta} \quad (26)$$

We want to show that the iterates $(\xi_{out}^{(k)},\xi_{relay}^{(k)})$ will converge almost surely to the set of stationary points of the o.d.e. $(\dot{\xi}_{out}(t),\dot{\xi}_{relay}(t)) = \overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out}(t),\xi_{relay}(t))}{\overline{U}^*(\xi_{out}(t),\xi_{relay}(t))}, \frac{f_2(\xi_{out}(t),\xi_{relay}(t))}{\overline{U}^*(\xi_{out}(t),\xi_{relay}(t))}\right)$. This will follow from Theorem 5.3.1 from [23], if we can show that $\left(-\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, -\frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ is the gradient of a continuously differentiable function.

Let us denote, by $\overline{\Gamma}_\pi$, $\overline{U}_\pi$ and $\overline{Q}_{out,\pi}$, the mean power per link, mean placement distance per link and mean outage per link respectively, under any given stationary deployment policy π. Let us define the function

$$G(\xi_{out},\xi_{relay}) := \inf_{\pi}\left(\frac{\overline{\Gamma}_\pi}{\overline{U}_\pi} + \xi_{out}\left(\frac{\overline{Q}_{out,\pi}}{\overline{U}_\pi} - \overline{q}\right) + \xi_{relay}\left(\frac{1}{\overline{U}_\pi} - \overline{N}\right)\right) \quad (27)$$

Lemma 8: $G(\xi_{out},\xi_{relay})$ is continuously differentiable and its gradient is $\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$.

Proof: The proof of Lemma 8 will be provided later in this section. ■

Now, $\left(-\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, -\frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ is the gradient of a continuously differentiable function $-G(\xi_{out},\xi_{relay})$. Hence, by Theorem 5.3.1 from [23], the iterates $(\xi_{out}^{(k)},\xi_{relay}^{(k)})$ will almost surely converge to the set of stationary points of the o.d.e. $(\dot{\xi}_{out}(t),\dot{\xi}_{relay}(t)) = \overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out}(t),\xi_{relay}(t))}{\overline{U}^*(\xi_{out}(t),\xi_{relay}(t))}, \frac{f_2(\xi_{out}(t),\xi_{relay}(t))}{\overline{U}^*(\xi_{out}(t),\xi_{relay}(t))}\right)$.

Lemma 9: If $(\xi_{out},\xi_{relay}) \in [0,A_2]\times[0,A_3]$ is a stationary point of $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$, then $(\lambda^*(\xi_{out},\xi_{relay}),\xi_{out},\xi_{relay}) \in \mathcal{K}(\overline{q},\overline{N})$, provided that A_2 and A_3 are chosen properly.

Proof: The proof of Lemma 9 will be provided later in this section. One way of choosing A_2 and A_3 has been described before the proof of this lemma. ■

We have already shown that $(\xi_{out}^{(k)},\xi_{relay}^{(k)})$ converges to the set of (ξ_{out},ξ_{relay}) pairs for which $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right) = (0,0)$. Hence, by Lemma 6 and Lemma 9, $(\lambda^{(k)},\xi_{out}^{(k)},\xi_{relay}^{(k)}) \to \mathcal{K}(\overline{q},\overline{N})$ almost surely, which completes the proof of Theorem 12.

Now we will prove Lemma 8 and Lemma 9. Before we prove Lemma 9, we will explain how A_2 and A_3 have to be chosen.

Proof of Lemma 8 Suppose that, for a given (ξ_{out},ξ_{relay}), the partial derivative $\frac{\partial G}{\partial \xi_{out}}$ exists. We will first show that this partial derivative is equal to $\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})} = \frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})} - \overline{q}$. Note that, the right partial derivative w.r.t. ξ_{out} (if it exists) is:

$$\frac{\partial G}{\partial \xi_{out}+} = \lim_{0<\Delta\to 0}\frac{G(\xi_{out}+\Delta,\xi_{relay}) - G(\xi_{out},\xi_{relay})}{\Delta}$$

Now, the optimal policy $\pi^*(\xi_{out},\xi_{relay})$ for the unconstrained problem in (3) will also minimize the expression for

$G(\xi_{out},\xi_{relay})$ in (27). But, the policy $\pi^*(\xi_{out},\xi_{relay})$ will be suboptimal for the pair $(\xi_{out}+\Delta,\xi_{relay})$. Hence, we have:

$$\begin{aligned} & G(\xi_{out}+\Delta,\xi_{relay}) \\ = & \inf_{\pi}\left(\frac{\overline{\Gamma}_{\pi}}{\overline{U}_{\pi}}+(\xi_{out}+\Delta)(\frac{\overline{Q}_{out,\pi}}{\overline{U}_{\pi}}-\overline{q})+\xi_{relay}(\frac{1}{\overline{U}_{\pi}}-\overline{N})\right) \\ \leq & \left(\frac{\overline{\Gamma}^*(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}+(\xi_{out}+\Delta)(\frac{\overline{Q}^*(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{q})\right. \\ & \left.+\xi_{relay}(\frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{N})\right) \\ = & G(\xi_{out},\xi_{relay})+\Delta\left(\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{q}\right) \end{aligned}$$

which implies that,

$$\frac{\partial G}{\partial \xi_{out}+} \leq \left(\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{q}\right)$$

In a similar manner, by using the fact that $\pi^*(\xi_{out},\xi_{relay})$ is suboptimal for the pair $(\xi_{out}-\Delta,\xi_{relay})$, we can claim that

$$\frac{\partial G}{\partial \xi_{out}-} \geq \left(\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{q}\right)$$

Since we have assumed that $\frac{\partial G}{\partial \xi_{out}}$ exists, we must have $\frac{\partial G}{\partial \xi_{out}+}=\frac{\partial G}{\partial \xi_{out}-}$, which proves that the partial derivative w.r.t. ξ_{out} will be equal to $(\frac{\overline{Q}^*_{out}(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{q})$.

We now turn to the existence of $\frac{\partial G}{\partial \xi_{out}}$. Note that, since $G(\xi_{out},\xi_{relay})$ is the minimum of a family of affine functions of ξ_{out} and ξ_{relay}, $G(\xi_{out},\xi_{relay})$ is concave and hence coordinatewise concave. Hence, for any given ξ_{relay}, there are only at most countably many values of ξ_{out} where $\frac{\partial G}{\partial \xi_{out}}$ does not exist. To see this, let us define the function $H(\xi_{out},\xi_{relay})$ to be the supremum of the subgradients of $G(\xi_{out},\xi_{relay})$ with respect to ξ_{out} (keeping ξ_{relay} fixed), at a point (ξ_{out},ξ_{relay}). Since $G(\xi_{out},\xi_{relay})$ is concave, $H(\xi_{out},\xi_{relay})$ will be decreasing in ξ_{out}. But any monotone real-valued function has an at most countable number of discontinuities (see [25], Theorem 4.30). Hence, for a given ξ_{relay}, the function $H(\xi_{out},\xi_{relay})$ is discontinuous for an at most countable number of values of ξ_{out}, and consequently $\frac{\partial G}{\partial \xi_{out}}$ exists everywhere except for an at most countable set of values of ξ_{out}.

For a given ξ_{relay}, let ξ'_{out} be one such value where $\frac{\partial G}{\partial \xi_{out}}$ does not exist. Then, there exists a sequence $\{\zeta_n\}_{n\geq 1}\downarrow 0$ such that $\frac{\partial G}{\partial \xi_{out}}$ exists at each $\xi_{out}=\xi'_{out}+\zeta_n$. This follows from the fact that for any $\zeta>0$, we can find one $\xi_{out}\in(\xi'_{out},\xi'_{out}+\zeta)$ where $\frac{\partial G}{\partial \xi_{out}}$ exists, otherwise the number of points where $\frac{\partial G}{\partial \xi_{out}}$ does not exist will become uncountable. Similarly, there exists a sequence $\{\kappa_n\}_{n\geq 1}\downarrow 0$ such that $\frac{\partial G}{\partial \xi_{out}}$ exists at each $\xi_{out}=\xi'_{out}-\kappa_n$.

Note that, by concavity, $\lim_{n\to\infty}\frac{\partial G}{\partial \xi_{out}}|_{\xi'_{out}-\kappa_n} \geq \frac{\partial G}{\partial \xi_{out}-}|_{\xi'_{out}} \geq \frac{\partial G}{\partial \xi_{out}+}|_{\xi'_{out}} \geq \lim_{n\to\infty}\frac{\partial G}{\partial \xi_{out}}|_{\xi'_{out}+\zeta_n}$. The last term in this chain of inequalities is equal to $\lim_{n\to\infty}(\frac{\overline{Q}^*_{out}(\xi'_{out}+\zeta_n,\xi_{relay})}{\overline{U}^*(\xi'_{out}+\zeta_n,\xi_{relay})}-\overline{q}) = (\frac{\overline{Q}^*_{out}(\xi'_{out},\xi_{relay})}{\overline{U}^*(\xi'_{out},\xi_{relay})}-\overline{q})$ by the arguments in the beginning of this proof and by the continuity results in Theorem 11. Same arguments hold for the first term in the chain of inequalities. Hence, $\frac{\partial G}{\partial \xi_{out}-}|_{\xi'_{out}} = \frac{\partial G}{\partial \xi_{out}+}|_{\xi'_{out}} = \frac{\partial G}{\partial \xi_{out}}|_{\xi'_{out}} = (\frac{\overline{Q}^*_{out}(\xi'_{out},\xi_{relay})}{\overline{U}^*(\xi'_{out},\xi_{relay})}-\overline{q})$.

In a similar way, we can show that $\frac{\partial G}{\partial \xi_{relay}} = (\frac{1}{\overline{U}^*(\xi_{out},\xi_{relay})}-\overline{N})$.

Now we see that both of the partial derivatives of G exist at all points and the partial derivatives are continuous in both ξ_{out} and ξ_{relay} (by Theorem 11). Hence, by Theorem 12.11 of [26], $G(\cdot,\cdot)$ is differentiable. Hence, the lemma is proved. □

Choice of A_2 and A_3: Let us consider the scenario where $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})},\frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ has a stationary point $(\xi'_{out},\xi'_{relay})$ (on the boundary of $\mathcal{G}$) such that $(\lambda^*(\xi'_{out},\xi'_{relay}),\xi'_{out},\xi'_{relay}) \notin \mathcal{K}(\overline{q},\overline{N})$. In this case, if $(\lambda^{(k)},\xi^{(k)}_{out},\xi^{(k)}_{relay}) \to (\lambda^*(\xi'_{out},\xi'_{relay}),\xi'_{out},\xi'_{relay})$ (depending on the sample path of the iterates in the OptExploreLimAdaptiveLearning algorithm), then we cannot expect the desired performance from the OptExploreLimAdaptiveLearning algorithm. To alleviate this problem, we need to choose A_2 and A_3 in a proper way. One method of choosing A_2 and A_3 is given below.

We will first explain how A_2 has to be chosen. Note that, for any given link of length u and shadowing realization w, $\text{argmin}_{\gamma\in\mathcal{S}}(\gamma+\xi_{out}Q_{out}(u,\gamma,w))=P_M$ if we choose ξ_{out} sufficiently large. We use this fact in the choice of A_2. The number A_2 has to be chosen so large that under $\xi_{out}=A_2$ and for all $A+1\leq u\leq A+B$, we will have $\mathbb{P}(\text{argmin}_{\gamma\in\mathcal{S}}(\gamma+A_2Q_{out}(u,\gamma,W))=P_M)>1-\kappa$ for some small enough $\kappa>0$. Such a choice of A_2 ensures that (i) the mean power per link (under policy $\pi^*(A_2,\xi_{relay})$), $\overline{\Gamma}^*(A_2,\xi_{relay})\geq(1-\kappa)P_M+\kappa P_1$ (which is close enough to P_M), which, further, ensures that (ii) $\frac{\overline{\Gamma}^*(A_2,\xi_{relay})}{1/\overline{N}}$ is greater than or equal to the optimal mean power per step for problem (4). The second claim is easy to see, since $\overline{\Gamma}^*(A_2,\xi_{relay}) \geq (1-\kappa)P_M+\kappa P_1 \geq \overline{\Gamma}^*(\xi^*_{out},\xi^*_{relay})$, and since $\overline{U}^*(\xi^*_{out},\xi^*_{relay})\geq\frac{1}{\overline{N}}$ (recall Assumption 1 about the existence of ξ^*_{out} and ξ^*_{relay}). Note that, the choice of κ depends on $(\overline{q},\overline{N})$ and the radio propagation parameters, and, hence, must be made carefully so that the condition is satisfied. In the proof of Lemma 9, we will see that this condition ensures that for any stationary point of the form $\xi_{out}=A_2,\xi_{relay}\in(0,A_3)$, we have $\frac{\overline{Q}^*_{out}(A_2,\xi_{relay})}{\overline{U}^*(A_2,\xi_{relay})}=\overline{q}$ and $\frac{1}{\overline{U}^*(A_2,\xi_{relay})}=\overline{N}$, and, consequently, the point (A_2,ξ_{relay}) will be in $\mathcal{K}(\overline{q},\overline{N})$.

The choice of A_2 must satisfy another condition. We need to choose A_2 so large that $\frac{\overline{Q}^*_{out}(A_2,0)}{\overline{U}^*(A_2,0)}\leq\overline{q}$. Note that, if $(\overline{q},\overline{N})$ is a feasible constraint pair, then a constraint $\overline{q}$ on the mean outage per step alone (if we drop the constraint on the relay placement rate) is also feasible. Let us consider the problem of minimizing the mean power per step subject to a constraint $\overline{q}$ on the mean outage per step. Then, we will choose $\xi_{relay}=0$. The mean outage per step under policy $\pi^*(\xi_{out},0)$ will still

decrease as ξ_{out} increases (by Theorem 5). Hence, we can choose an A_2 which satisfies this condition. This condition will be used in showing that if $(A_2, 0)$ is a stationary point of the o.d.e., then $(\lambda^*(A_2,0), A_2, 0) \in \mathcal{K}(\overline{q}, \overline{N})$.

A_2 has to be chosen (according to the two criteria mentioned above) via prior computation, using the prior knowledge of the propagation environment; if we know the range of values of radio propagation parameters (e.g., η and σ), we can compute what value of A_2 will satisfy the criteria under all possible radio propagation parameters.

Once A_2 is chosen, we need to choose A_3. The number A_3 has to be chosen so large that for any $\xi_{out} \in [0, A_2]$, we will have $\overline{U}^*(\xi_{out}, A_3) > \frac{1}{\overline{N}}$ (provided that $\frac{1}{\overline{N}} < A + B$). This is possible and obvious from the structure of OptExploreLim (Algorithm 3); by choosing ξ_{relay} large enough, we can achieve a mean placement distance equal to $(A + B)$, provided that $\xi_{out} \in [0, A_2]$. For example, if we choose $A_3 = 100(A + B)(P_M + A_2)$, then:

$$\begin{aligned} & \lambda^*(\xi_{out}, A_3) \\ = & \frac{\overline{\Gamma}^*(\xi_{out}, A_3) + \xi_{out}\overline{Q}^*_{out}(\xi_{out}, A_3) + A_3}{\overline{U}^*(\xi_{out}, A_3)} \\ \geq & \frac{A_3}{A+B} = 100(P_M + A_2) \end{aligned}$$

and $\pi^*(\xi_{out}, A_3)$ will always place at a distance of $(A + B)$. This choice of A_3 ensures that the policy $\pi^*(\xi_{out}, A_3)$ satisfies the constraint on the relay placement rate with strict inequality, and hence no point of the form (ξ_{out}, A_3) is a stationary point of the o.d.e.

The numbers A_2 and A_3 have to be chosen so large that there exists at least one $(\xi'_{out}, \xi'_{relay}) \in [0, A_2] \times [0, A_3]$ such that $(\lambda^*(\xi'_{out}, \xi'_{relay}), \xi'_{out}, \xi'_{relay}) \in \mathcal{K}(\overline{q}, \overline{N})$. □

Proof of Lemma 9: Suppose that $(\xi_{out}, \xi_{relay}) \in [0, A_2] \times [0, A_3]$ is a stationary point of $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$.

Note that, $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ is equal to $\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ if (ξ_{out}, ξ_{relay}) lies in the interior of $[0, A_2] \times [0, A_3]$. Thus, for any stationary point $(\xi_{out}, \xi_{relay}) \in (0, A_2) \times (0, A_3)$, the optimal policy $\pi^*(\xi_{out}, \xi_{relay})$ meets both constraints in (4) with R.H.S.=0. For such a stationary point, $(\lambda^*(\xi_{out}, \xi_{relay}), \xi_{out}, \xi_{relay}) \in \mathcal{K}(\overline{q}, \overline{N})$ (by Theorem 1).

A point (ξ_{out}, ξ_{relay}) on the boundary has $\xi_{out} = 0$ or $\xi_{out} = A_2$ or $\xi_{relay} = 0$ or $\xi_{relay} = A_3$.

Let us recall Assumption 1 and the definition of $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ (equation (26)). The first component of this vector-valued function at $\xi_{out} = 0$ is equal to $\frac{f_1(0,\xi_{relay})}{\overline{U}^*(0,\xi_{relay})}$ if $f_1(0, \xi_{relay}) \geq 0$, and 0 otherwise. We can make similar observations at $\xi_{out} = A_2$, $\xi_{relay} = 0$ and $\xi_{relay} = A_3$.

If $\xi_{relay} = A_3$, then by the choice of A_3 as suggested in the OptExploreLimAdaptiveLearning algorithm (Algorithm 8), we will have $f_2(\xi_{out}, A_3) < 0$ (since $\overline{U}^*(\xi_{out}, A_3) > \frac{1}{\overline{N}}$). This implies that no point on $\xi_{relay} = A_3$ can be a stationary point of $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$, since the second component of this function will be $\frac{f_2(\xi_{out},A_3)}{\overline{U}^*(\xi_{out},A_3)}$ which is strictly negative.

Suppose that that there is a stationary point of $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ of the form $\xi_{out} = A_2, \xi_{relay} \in (0, A_3)$. Then $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ will be zero if and only if $f_1(A_2, \xi_{relay}) \geq 0$ and $f_2(A_2, \xi_{relay}) = 0$. If $f_1(A_2, \xi_{relay}) = 0$ and $f_2(A_2, \xi_{relay}) = 0$, then $(\lambda^*(A_2, \xi_{relay}), A_2, \xi_{relay})$ will belong to $\mathcal{K}(\overline{q}, \overline{N})$ (by Theorem 1), since the corresponding optimal policy $\pi^*(A_2, \xi_{relay})$ will satisfy both constraints in (4) with equality. Now, we will show that, if A_2 is chosen appropriately as explained before, the case $f_1(A_2, \xi_{relay}) > 0$ and $f_2(A_2, \xi_{relay}) = 0$ will never arise. Suppose that $f_1(A_2, \xi_{relay}) > 0$ and $f_2(A_2, \xi_{relay}) = 0$ for some $\xi_{relay} \in (0, A_3)$. Consider a new problem of minimizing the mean outage per step, subject to a constraint $\frac{\overline{\Gamma}^*(A_2,\xi_{relay})}{\overline{U}^*(A_2,\xi_{relay})}$ on the mean power per step and a constraint $\frac{1}{\overline{U}^*(A_2,\xi_{relay})} = \overline{N}$ on the mean number of relays per step. By Theorem 1, $\pi^*(A_2, \xi_{relay})$ is the optimal policy for this new problem, since it satisfies both constraints with equality. But the policy $\pi^*(\xi^*_{out}, \xi^*_{relay})$ has the following properties: (i) $\frac{1}{\overline{U}^*(A_2,\xi_{relay})} = \overline{N} \geq \frac{1}{\overline{U}^*(\xi^*_{out},\xi^*_{relay})}$ (see Assumption 1 in Section VII), i.e., $\pi^*(\xi^*_{out}, \xi^*_{relay})$ has a smaller relay placement rate compared to $\pi^*(A_2, \xi_{relay})$ (since $\pi^*(\xi^*_{out}, \xi^*_{relay})$ satisfies the constraint $\overline{N}$ on the mean number of relays per step), (ii) $\frac{\overline{\Gamma}^*(A_2,\xi_{relay})}{\overline{U}^*(A_2,\xi_{relay})} \geq \frac{(1-\kappa)P_M + \kappa P_1}{\frac{1}{\overline{N}}} \geq \frac{\overline{\Gamma}^*(\xi^*_{out},\xi^*_{relay})}{\overline{U}^*(\xi^*_{out},\xi^*_{relay})}$, i.e., $\pi^*(\xi^*_{out}, \xi^*_{relay})$ has a smaller mean power per step compared to $\pi^*(A_2, \xi_{relay})$ (by the choice of A_2, see the previous discussion on the choice of A_2), and (iii) $\frac{\overline{Q}^*_{out}(A_2,\xi_{relay})}{\overline{U}^*(A_2,\xi_{relay})} > \overline{q} \geq \frac{\overline{Q}^*_{out}(\xi^*_{out},\xi^*_{relay})}{\overline{U}^*(\xi^*_{out},\xi^*_{relay})}$, i.e., $\pi^*(\xi^*_{out}, \xi^*_{relay})$ has a strictly smaller mean outage per step compared to $\pi^*(A_2, \xi_{relay})$ (since $\pi^*(\xi^*_{out}, \xi^*_{relay})$ satisfies the constraint $\overline{q}$ on the mean outage per step and since $f_1(A_2, \xi_{relay}) > 0$. This leads to a contradiction since it violates the optimality of the policy $\pi^*(A_2, \xi_{relay})$ for the new problem. Hence, $\overline{\Lambda}_{\mathcal{G}}\left(\frac{f_1(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}, \frac{f_2(\xi_{out},\xi_{relay})}{\overline{U}^*(\xi_{out},\xi_{relay})}\right)$ cannot have a stationary point of the form $\xi_{out} = A_2$, $\xi_{relay} \in (0, A_3)$ such that $f_1(A_2, \xi_{relay}) > 0$ and $f_2(A_2, \xi_{relay}) = 0$.

Now consider any stationary point of the form $\xi_{out} \in (0, A_2), \xi_{relay} = 0$. Then we must have $f_1(\xi_{out}, 0) = 0$ and $f_2(\xi_{out}, 0) \leq 0$. Now, consider a new problem of minimizing the mean power per step subject to a constraint $\overline{q}$ on the mean outage per step (with no constraint on the relay placement rate); an optimal policy for this problem is $\pi^*(\xi_{out}, 0)$ (by Theorem 1, since $\pi^*(\xi_{out}, 0)$ satisfies the outage constraint with equality). Since (4) has one more constraint, it will have

a larger mean power per step, i.e., $\frac{\overline{\Gamma}^*(\xi_{out},0)}{\overline{U}^*(\xi_{out},0)} \leq \frac{\overline{\Gamma}^*(\xi^*_{out},\xi^*_{relay})}{\overline{U}^*(\xi^*_{out},\xi^*_{relay})}$ (recall Assumption 1 about the existence of ξ^*_{out} and ξ^*_{relay}). If they are equal, then $(\xi_{out}, 0)$ will be an optimal pair for (4) and $(\lambda^*(\xi_{out}, 0), \xi_{out}, 0)$ will be in $\mathcal{K}(\overline{q}, \overline{N})$. If $\frac{\overline{\Gamma}^*(\xi_{out},0)}{\overline{U}^*(\xi_{out},0)} < \frac{\overline{\Gamma}^*(\xi^*_{out},\xi^*_{relay})}{\overline{U}^*(\xi^*_{out},\xi^*_{relay})}$, then the optimality of $\pi^*(\xi^*_{out}, \xi^*_{relay})$ for the problem (4) will be violated, since $\pi^*(A_2, 0)$ will produce a strictly smaller mean power per step while meeting the outage constraint with equality (since $f_1(\xi_{out}, 0) = 0$) and the relay placement rate constraint (since $f_2(\xi_{out}, 0) \leq 0$).

We can take care of stationary points of the form $\xi_{relay} \in (0, A_3), \xi_{out} = 0$ in a similar way.

If $(0, 0)$ is a stationary point, then $\pi^*(0, 0)$ satisfies both constraints. Also, $\pi^*(0, 0)$ places at distance $(A + B)$ steps and uses the minimum power level for all links. Then $\pi^*(0, 0)$ is optimal for our original problem (3).

At $(A_2, 0)$, we will have a stationary point if and only if $f_1(A_2, 0) \geq 0$ and $f_2(A_2, 0) \leq 0$. If $f_1(A_2, 0) \geq 0$ and $f_2(A_2, 0) = 0$, then we can make similar claims as in the $\xi_{out} = A_2$ and $\xi_{relay} \in (0, A_3)$ case. If $f_1(A_2, 0) = 0$ and $f_2(A_2, 0) < 0$, then we can make similar claims as in the $\xi_{out} \in (0, A_2)$ and $\xi_{relay} = 0$ case. By the choice of A_2, $\pi^*(A_2, 0)$ satisfies the outage constraint $\overline{q}$. Hence, the case $f_1(A_2, 0) > 0$ will not arise.

Hence, the lemma is proved. □

D. *Proof of Theorem 13*

We will only prove that $\limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{N_x} \Gamma_i}{x} \leq \gamma^*$ almost surely.

Let us denote the shadowing random variable in the link between the potential locations located at distances $i\delta$ and $j\delta$ from the sink node by $W_{i,j}$. The sample space Ω associated with the deployment process is the collection of all ω (each ω corresponds to a fixed realization $\{w_{i,j} : i \geq 0, j \geq 0, i > j, A + 1 \leq i - j \leq A + B\}$ of all possible shadowing random variables that might be encountered in the measurement process for deployment up to infinity). Let $\mathcal{F}$ be the Borel σ-algebra on Ω. Let $S_k = \sum_{i=1}^{k} U_i$ be the distance (in steps) of the k-th relay from the source, and $\mathcal{F}_k := \sigma\Big((\lambda^{(0)}, \xi^{(0)}_{out}, \xi^{(0)}_{relay}); W_{i,j} : i \geq 0, j \geq 0, i > j, A + 1 \leq i-j \leq A+B, i \leq S_{k-1}+A+B, j \leq S_{k-1}+A+B\Big)$. The sequence of σ-algebras $\mathcal{F}_k$ is increasing in k, and $\mathcal{F}_k$ captures the history of the deployment process up to the deployment of the k-th relay.

Let us fix an $\epsilon > 0$.

Let us recall (from Section VII-B) the definition of the set $\hat{\mathcal{K}}(\overline{q}, \overline{N}) := \mathcal{K}(\overline{q}, \overline{N}) \cap ([0, (P_M + A_2 + A_3)] \times [0, A_2] \times [0, A_3])$.

Now, by Lemma 7 (see Appendix E, Section C4), the quantities $\overline{\Gamma}(\lambda, \xi_{out}, \xi_{relay})$, $\overline{Q}_{out}(\lambda, \xi_{out}, \xi_{relay})$ and $\overline{U}(\lambda, \xi_{out}, \xi_{relay})$ (recall the notation from Section IV-B) are continuous in $(\lambda, \xi_{out}, \xi_{relay})$. Hence, the ratios $\frac{\overline{\Gamma}(\lambda,\xi_{out},\xi_{relay})}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$, $\frac{\overline{Q}_{out}(\lambda,\xi_{out},\xi_{relay})}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$ and $\frac{1}{\overline{U}(\lambda,\xi_{out},\xi_{relay})}$ are uniformly continuous over the compact set $[0, 2(P_M + A_2 + A_3)] \times [0, A_2] \times [0, A_3]$. Hence, for any given $\epsilon > 0$, we can find a $\delta_\epsilon > 0$ such that if $(\lambda, \xi_{out}, \xi_{relay})$ belongs to a δ_ϵ-neighbourhood of $\hat{\mathcal{K}}(\overline{q}, \overline{N})$, then $(\lambda, \xi_{out}, \xi_{relay})$ also belongs to the set $\hat{\mathcal{K}}_\epsilon(\overline{q}, \overline{N})$ where:

$$\begin{aligned}
\hat{\mathcal{K}}_\epsilon(\overline{q}, \overline{N}) \quad = \quad & \Big\{(\lambda, \xi_{out}, \xi_{relay}) : \\
& \frac{\overline{\Gamma}(\lambda, \xi_{out}, \xi_{relay})}{\overline{U}(\lambda, \xi_{out}, \xi_{relay})} \in [\gamma^* - \epsilon, \gamma^* + \epsilon] \\
& \frac{\overline{Q}_{out}(\lambda, \xi_{out}, \xi_{relay})}{\overline{U}(\lambda, \xi_{out}, \xi_{relay})} \leq \overline{q} + \epsilon, \\
& \frac{1}{\overline{U}(\lambda, \xi_{out}, \xi_{relay})} \leq \overline{N} + \epsilon, \\
& 0 \leq \lambda \leq 2(P_M + A_2 + A_3), \\
& 0 \leq \xi_{out} \leq A_2, 0 \leq \xi_{relay} \leq A_3 \Big\}
\end{aligned}$$

But, by Theorem 12, $(\lambda^{(k)}, \xi^{(k)}_{out}, \xi^{(k)}_{relay}) \to \hat{\mathcal{K}}(\overline{q}, \overline{N})$ almost surely. Hence, there exists an integer-valued random variable T such that (i) $(\lambda^{(k)}, \xi^{(k)}_{out}, \xi^{(k)}_{relay})$ belongs to a δ_ϵ neighbourhood of $\hat{\mathcal{K}}(\overline{q}, \overline{N})$ for all $k \geq T$, and (ii) $\mathbb{P}(T < \infty) = 1$. In other words, for a sample path ω (for ω lying in a set of probability 1), there exists $T(\omega) < \infty$ such that $(\lambda^{(k)}, \xi^{(k)}_{out}, \xi^{(k)}_{relay})$ belongs to a δ_ϵ neighbourhood of $\hat{\mathcal{K}}(\overline{q}, \overline{N})$ for all $k \geq T(\omega)$. Hence, for a sample path ω (for ω lying in a set of probability 1), there exists $T(\omega) < \infty$ such that $(\lambda^{(k)}, \xi^{(k)}_{out}, \xi^{(k)}_{relay}) \in \hat{\mathcal{K}}_\epsilon(\overline{q}, \overline{N})$ for all $k \geq T(\omega)$.

Using the boundedness of Γ_i in the first equality, we obtain:

$$\begin{aligned}
& \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{N_x} \Gamma_i}{x} \\
= \quad & \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{N_x+1} \Gamma_i}{x} \\
\leq \quad & \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T < N_x + 1) \sum_{i=1}^{T} \Gamma_i\Big)}{x} \\
& + \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T < N_x + 1) \sum_{i=T+1}^{N_x+1} \Gamma_i\Big)}{x} \\
& + \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T \geq N_x + 1) \sum_{i=1}^{N_x+1} \Gamma_i\Big)}{x} \\
\leq \quad & \mathbb{E}_{\pi_{oelal}} \limsup_{x\to\infty} \frac{\mathbb{I}(T < N_x + 1) \sum_{i=1}^{T} \Gamma_i}{x} \\
& + \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T < N_x + 1) \sum_{i=T+1}^{N_x+1} \Gamma_i\Big)}{x} \\
& + \mathbb{E}_{\pi_{oelal}} \limsup_{x\to\infty} \frac{\mathbb{I}(T \geq N_x + 1) \sum_{i=1}^{N_x+1} \Gamma_i}{x} \\
= \quad & \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T < N_x + 1) \sum_{i=T+1}^{N_x+1} \Gamma_i\Big)}{x} \\
= \quad & \limsup_{x\to\infty} \Bigg(\frac{\mathbb{E}_{\pi_{oelal}} \Big(\mathbb{I}(T < N_x + 1) \sum_{i=T+1}^{N_x+1} \Gamma_i\Big)}{\mathbb{E}_{\pi_{oelal}} \sum_{i=T+1}^{N_x+1} U_i} \\
& \times \mathbb{E}_{\pi_{oelal}} \Big(\frac{\sum_{i=T+1}^{N_x+1} U_i}{x} \Big) \Bigg)
\end{aligned} \tag{28}$$

Here the second inequality follows from Fatou's

lemma. The second equality follows from the facts that $0 \leq \limsup_{x\to\infty} \frac{\sum_{i=1}^{T}\Gamma_i \mathbb{I}(T<N_x+1)}{x} \leq \limsup_{x\to\infty} \frac{\sum_{i=1}^{T}\Gamma_i}{x} \leq \limsup_{x\to\infty} \frac{P_M T}{x} = 0$ almost surely and $0 \leq \limsup_{x\to\infty} \frac{\sum_{i=1}^{N_x+1}\Gamma_i \mathbb{I}(T\geq N_x+1)}{x} \leq \frac{P_M}{A+1}\limsup_{x\to\infty}\mathbb{I}(T \geq N_x+1) = 0$ almost surely (since $\mathbb{P}(T<\infty)=1$ and $\lim_{x\to\infty} N_x = \infty$ almost surely).

Now,

$$\begin{aligned}
& \limsup_{x\to\infty} \mathbb{E}_{\pi_{oelal}}\left(\frac{\sum_{i=T+1}^{N_x+1} U_i}{x}\right) \\
\leq\ & \limsup_{x\to\infty} \mathbb{E}_{\pi_{oelal}} \frac{\sum_{i=1}^{N_x+1} U_i}{x} \\
\leq\ & \mathbb{E}_{\pi_{oelal}} \limsup_{x\to\infty} \frac{\sum_{i=1}^{N_x+1} U_i}{x} \\
=\ & 1
\end{aligned}$$

Here the second inequality follows from Fatou's lemma, and the equality follows from the fact that $\lim_{x\to\infty} \frac{\sum_{i=1}^{N_x+1} U_i}{x} = 1$ almost surely.

Hence, from (28),

$$\begin{aligned}
& \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}} \sum_{i=1}^{N_x}\Gamma_i}{x} \\
\leq\ & \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}}\left(\mathbb{I}(T<N_x+1)\sum_{i=T+1}^{N_x+1}\Gamma_i\right)}{\mathbb{E}_{\pi_{oelal}}\sum_{i=T+1}^{N_x+1} U_i} \\
=\ & \limsup_{x\to\infty} \frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=T+1}^{N_x+1}\Gamma_i}{\mathbb{E}_{\pi_{oelal}}\sum_{i=T+1}^{N_x+1} U_i} \qquad (29)
\end{aligned}$$

Let us denote by $\mathbb{E}_{\pi_{oelal},t}(\cdot)$ the conditional expectation under policy π_{oelal} given that $T=t$. Now,

$$\begin{aligned}
& \mathbb{E}_{\pi_{oelal}} \sum_{i=T+1}^{N_x+1}\Gamma_i \\
=\ & \mathbb{E}_{\pi_{oelal}} \sum_{i=T+1}^{\infty}\Gamma_i \mathbb{I}(i\leq N_x+1) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t) \\
& \times\mathbb{E}_{\pi_{oelal}}\left(\sum_{i=t+1}^{\infty}\Gamma_i\mathbb{I}(i\leq N_x+1)\middle| T=t\right) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t) \\
& \times\mathbb{E}_{\pi_{oelal}}\left(\sum_{i=t+1}^{\infty}\Gamma_i\mathbb{I}(N_x\geq i-1)\middle| T=t\right) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t)\mathbb{E}_{\pi_{oelal},t}\left(\sum_{i=t+1}^{\infty}\Gamma_i\mathbb{I}(N_x\geq i-1)\right) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t)\sum_{i=t+1}^{\infty}\mathbb{E}_{\pi_{oelal},t}\left(\Gamma_i\mathbb{I}(N_x\geq i-1)\right) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t) \\
& \times\sum_{i=t+1}^{\infty}\mathbb{E}_{\pi_{oelal},t}\left(\mathbb{E}_{\pi_{oelal},t}\left(\Gamma_i\mathbb{I}(N_x\geq i-1)\middle|\mathcal{F}_{i-1}\right)\right) \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t) \\
& \times\sum_{i=t+1}^{\infty}\mathbb{E}_{\pi_{oelal},t}\left(\mathbb{I}(N_x\geq i-1)\mathbb{E}_{\pi_{oelal},t}\left(\Gamma_i\middle|\mathcal{F}_{i-1}\right)\right) \\
\leq\ & (\gamma^*+\epsilon)\sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t)\times \\
& \sum_{i=t+1}^{\infty}\mathbb{E}_{\pi_{oelal},t}\left(\mathbb{I}(N_x\geq i-1)\mathbb{E}_{\pi_{oelal},t}\left(U_i\middle|\mathcal{F}_{i-1}\right)\right) \qquad (30)
\end{aligned}$$

where the fifth equality follows from the Monotone Convergence Theorem, and the last equality follows from the fact that the random variable $\mathbb{I}(N_x\geq i-1)=\mathbb{I}(\sum_{k=1}^{i-1}U_k\leq x)$ is measurable with respect to $\mathcal{F}_{i-1}$. The last inequality follows from that fact that $\frac{\mathbb{E}_{\pi_{oelal},t}\left(\Gamma_i\middle|\mathcal{F}_{i-1}\right)}{\mathbb{E}_{\pi_{oelal},t}\left(U_i\middle|\mathcal{F}_{i-1}\right)} \leq \gamma^*+\epsilon$ almost surely for $i>t$, given that $T=t$ (since $(\lambda^{(i-1)},\xi_{out}^{(i-1)},\xi_{relay}^{(i-1)})\in\hat{\mathcal{K}}_\epsilon(\overline{q},\overline{N})$ for all $i-1\geq T$).

On the other hand, we can show that:

$$\begin{aligned}
& \mathbb{E}_{\pi_{oelal}} \sum_{i=T+1}^{N_x+1} U_i \\
=\ & \sum_{t=1}^{\infty}\mathbb{P}_{\pi_{oelal}}(T=t)\times \\
& \sum_{i=t+1}^{\infty}\mathbb{E}_{\pi_{oelal},t}\left(\mathbb{I}(N_x\geq i-1)\mathbb{E}_{\pi_{oelal},t}\left(U_i\middle|\mathcal{F}_{i-1}\right)\right) \qquad (31)
\end{aligned}$$

From (29), (30) and (31), we obtain that $\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}\Gamma_i}{x} \leq \gamma^*+\epsilon$. Since $\epsilon>0$ is arbitrary, we have $\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}\Gamma_i}{x}\leq\gamma^*$. But γ^* is the optimal mean power per step for problem (4). Hence, $\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}\Gamma_i}{x}=\gamma^*$.

In a similar manner, we can show that $\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}\sum_{i=1}^{N_x}Q_{out}^{(i,i-1)}}{x} \leq \overline{q}$ and $\limsup_{x\to\infty}\frac{\mathbb{E}_{\pi_{oelal}}N_x}{x}\leq\overline{N}$. □