Cross-layer Monitoring in Transparent Optical Networks

Yvan Pointurier, Member, IEEE, Mark Coates, Senior Member, IEEE, and Michael Rabbat, Member, IEEE

Abstract—In transparent optical networks, signals propagate over all-optical lightpaths. The absence of regenerating devices that act in the electrical domain renders end-to-end monitoring difficult. Quality of Transmissions (QoT) metrics quantify the degradation in quality that a signal experiences as it traverses a lightpath. Hardware monitors that can directly measure QoT are expensive, which motivates the development of monitoring schemes that require fewer monitors but can still generate accurate QoT estimates. In this paper we describe a monitoring scheme that estimates the QoT of multiple lightpaths in a network. Our focus is on estimating Bit-Error-Rates (BERs), but the methodology is also applicable for other metrics. One of the primary innovations in this monitoring framework is the establishment of “active lightpaths” — lightpaths that carry no useful data but are instead used as measurement probes. We describe a method for choosing where to establish the active lightpaths in order to maximize the information gain. We demonstrate with simulations the possibility to trade-off the amount of costly hardware monitoring equipment with cheaper, temporary active lightpaths, while still achieving accurate monitoring.

Index Terms—Optical networks; monitoring; physical impairments; estimation.

I. INTRODUCTION

Optical signals sustain impairments stemming from their propagation in the fiber medium. In transparent optical networks, the absence of electrical regeneration at every node means that signals can traverse very long distances leading to the accumulation of significant impairments. This can result in the unacceptable degradation of end-to-end transmission performance, as measured by Quality of Transmission (QoT) metrics. A standard QoT metric is the Bit-Error Rate (BER); the Q-factor (defined in Section 2) is another common metric. In transparent optical networks, network operators are interested in monitoring the QoT of each lightpath, which is the combination of a route and a wavelength used to transmit a signal. Inadequate performance of a lightpath in terms of BER can indicate that a component on the lightpath is not performing as originally planned (a case of soft failure, where connectivity is maintained but degraded) or that the Routing and Wavelength Assignment algorithm used to pick the degraded lightpath made a poor decision. Signal monitoring is essential, both for the detection of failures and to make better network decisions in order to satisfy service level agreements. It is a challenging task, however, because the absence of electrical generation and the high cost of monitoring devices mean that signals can be monitored only at selected — and possibly few — locations in the network.

In this paper, we describe a QoT monitoring scheme that strives to reduce the amount of hardware needed to perform monitoring while preserving the accuracy of the performance metric estimates. There are two main novel components of this scheme: (i) an optical network probing technique, which we call active lightpath monitoring; and (ii) an inference technique that exploits topological knowledge and incorporates regularization to permit performance metric estimation for unmeasured lightpaths.

The active lightpath monitoring technique involves the construction of lightpath probes. These are additional lightpaths that are lit only for QoT monitoring purposes; they do not carry any meaningful data (they should carry random bits). They are established for a small period of time (a fraction of a second or a few seconds, depending on the monitored parameters) to make measurements, and the resources they are using can be re-allocated at any time. If in-band probing is undesirable, an out-of-band optical monitoring channel can be used instead.

Automated optical networks employ wavelength-selective switches and reconfigurable optical add/drop multiplexer nodes (so-called “ROADMs”), making it possible to dynamically establish and reconfigure lightpaths. Moreover, such networks must be over-provisioned — at any point in time resources, including transmitters, receivers, and wavelengths, must be ready to accommodate future demands. These resources are natural candidates for use by the active probing framework introduced in this paper since, although they will already be configured and lit, they will not carry customer traffic.

Our proposed inference methodology exploits the spatial correlation between the QoT metrics of all established lightpaths, both the data-carrying lightpaths (we will call these the passive lightpaths) and the lightpath probes (active lightpaths). This spatial correlation is induced because the physical impairments are caused at the link level, so that if two lightpaths (on different wavelengths) share one or more links, their BERs become correlated. It is only possible to directly monitor a lightpath’s performance if it terminates where a hardware monitor is located. This can mean that only a small fraction of the passive (data-carrying) lightpaths are directly measured. The lightpath probes, which are designed to terminate at hardware monitors, provide valuable information (through the correlation) about the performance of the passive lightpaths that cannot be measured.
There is a clear trade-off between the amount of monitoring hardware and the accuracy of QoT estimates. Our monitoring scheme adds another factor into this trade-off — the number of lightpath probes. Our proposed monitoring scheme can preserve the estimation accuracy while reducing the number of hardware monitors by creating lightpath probes. This is an example of “cross-layer” network monitoring, because network layer information (the routing and wavelength assignment state of the network) is exploited to address a physical layer task (lightpath QoT monitoring).

A. Applications

We now highlight a few potential applications of the proposed monitoring framework:

**Application 1:** users of a core network infrastructure expect that the QoT of their lightpaths is guaranteed to be above a certain level as defined in a Service Level Agreement (SLA) between customer and operator. Enforcing the SLA requires that the core network operator should be able to monitor the QoT of all lightpaths at all times. An operator can use the framework outlined in this work to reduce the amount of monitoring equipment required to achieve this task.

**Application 2:** Hard failures such as link cuts can be detected and located relatively easily using power monitors, which are deployed in most networking equipment. In contrast, soft failures are caused by the degradation rather than the complete failure of a transmission device. By correlating the alarms generated by the proposed monitoring scheme when monitored quantities become unacceptable, it is possible to detect where a soft failure has occurred using failure localization algorithms such as those in [2], [3]. In a more preventive context, operators can use the proposed monitoring framework to anticipate soft failures.

**Application 3:** Instead of monitoring the lightpaths that are already established, a network manager may want to predict the QoT of a potential lightpath, before deciding whether to actually establish it. Using monitoring information from lightpaths already established and the estimation framework presented here, QoT prediction before establishment becomes possible (see [4] for one approach).

B. Related Work

There has been a significant amount of work directly addressing the problem of monitoring in transparent optical networks [5]. Typically, in optical networks, monitoring encompasses both impairment monitoring, where the focus is a single (type of) impairment (e.g., noise/OSNR or chromatic dispersion) [6], and performance monitoring, where the effects of all impairments are evaluated as a whole using a metric that reflects the end-to-end QoT of a lightpath (e.g., BER or Q-factor) [5]. Although we focus in this paper on performance monitoring, the monitoring technique described herein can also be applied to impairment monitoring. Research in this direction has been presented in [4], [7].

The monitoring framework we propose is distinct from existing strategies because it addresses the network-wide estimation of end-to-end (lightpath-level) QoT performance metrics and it involves the establishment of active lightpath probes. There have been other efforts to exploit spatial and temporal correlation in order to perform network-wide path-level performance estimation [8]–[11]. These techniques focus primarily on network performance metrics such as path delay and loss rather than the physical layer QoT metrics. The monitoring framework we propose builds upon the passive monitoring scheme (without any active probing) for transparent optical networks that we presented in [11]. As indicated by the simulation results presented in Section IV, the use of active lightpath monitoring makes a major difference in terms of the number of monitoring devices required to achieve a specified estimation accuracy. We presented a preliminary method and associated results in [1].

C. Contributions and Organization

The contributions of this paper are three-fold. First, we adapt estimation techniques to the problem of network-wide monitoring of transparent optical networks, refining them to take advantage of prior knowledge about the physical layer. This allows performance estimation for lightpaths that are not directly measured; the regularization procedure we employ leads to significantly more accurate estimates. Second, we propose the procedure of establishing active lightpaths (lightpath probes) to gain additional information and we provide an algorithm for selecting the routes. The algorithm strives to select the most informative paths to maximize the effective coverage of an existing monitoring hardware deployment. Third, we propose an algorithm for monitor placement to address the scenario when network designers must make decisions about where to deploy equipment to maximize estimation coverage and performance. We present simulation results that indicate that the proposed monitoring scheme provides much better estimates and monitoring coverage than the passive monitoring approach of [11]. The simulations also clearly illustrate the trade-offs between number of monitors, number of lightpath probes, and estimation accuracy.

This paper is organized as follows. In Section II, we state our assumptions concerning the network physical layer and hardware monitors. Section III gives an overview of the estimators used in in this work and presents our active monitoring scheme. We present simulation results in Section IV and we conclude the paper in Section V.

II. Modeling

Our key assumptions are that the physical-layer metrics of interest are (i) measurable by a hardware device; and (ii) approximately linear, in the sense that the end-to-end (path-level) metric associated with a lightpath is approximately equal to the sum, over the traversed links, of the corresponding link-level metrics. We will provide more details and strive to justify these assumptions, but first we outline the physical-layer properties of the networks under study. We assume that transparent optical regeneration, except for standard optical amplification, is not available. We further assume that wavelength converters are not available and that the wavelength continuity constraint holds: lightpaths are established over the same wavelength...
from end to end. The latter assumption is not essential (our monitoring framework could readily be adapted to scenarios where it does not hold), but it serves to make our analysis concrete. We also assume that a routing and wavelength assignment (RWA) algorithm is in charge of selecting lightpaths on call arrivals. This algorithm enforces both the wavelength continuity constraint and the QoT constraint. The monitoring technique described in this paper makes no assumption about the nature of the RWA algorithm.

The impairments we consider are amplifier noise and inter-symbol interference (ISI). We estimate the BER of a lightpath through its so-called Q-factor: \( Q = \frac{\mu_0 - \mu_1}{\sigma_0 + \sigma_1} \) (where \( \mu_0 \) and \( \mu_1 \) are the means of the distributions of the “0” and “1” symbols at the photo-detection stage, and \( \sigma_0 \) and \( \sigma_1 \) are their respective standard deviations). The BER is related to the Q-factor as follows: \( BER = \frac{1}{2} \text{erfc}(Q/\sqrt{2}) \). We model the impact of amplifier noise and ISI as additive variances (\( \sigma_0^2 \) and \( \sigma_1^2 \), respectively) in \( \sigma_1^2 \) such that \( \sigma_1^2 = \sigma_n^2 + \sigma_{isi}^2 \) [12]. In addition, we model single channel nonlinear effects, namely, the interaction between Self-Phase Modulation (SPM) and chromatic dispersion. We ignore other nonlinear effects such as Cross-Phase Modulation (XPM) and Four-Wave Mixing (FWM)\(^1\). In general, the methods presented in this paper can be used to monitor impairments that either do not depend on the wavelength or have a weak dependence—i.e., impairments that are flat across the transmission spectrum.

We make no assumption about the exact nature of the hardware monitors deployed in the network, but we do require that they can measure electrical power and noise, either directly or indirectly. The hardware monitors can then determine estimates of \( \mu_0, \mu_1, \sigma_0, \) and \( \sigma_1 \). This type of monitoring falls under the category of histogram-based performance monitoring [13]. We model a hardware monitor as a device located at the extremity of a link, after photo-detection and inside the receiver modules (the alternative of placing a monitor directly on a transmission line requires the undesirable diversion of signal power). A consequence is that we can only measure lightpaths that terminate at a link where a monitor is located. However, our monitoring framework would change only moderately if power diversion were available. Using the notation introduced in the next section, the availability of intermediate measurement points for already established lightpaths can easily be incorporated as additional rows in the monitoring matrix \( G_m \); regarding active lightpaths, the proposed framework can easily be adapted by reinterpreting the number \( n_a \) of active lightpaths that can be established in the network as the number of additional monitoring resources used to perform active monitoring.

Although the current work is applied specifically to intensity-modulated networks, it could be generalized to coherent systems, which are currently being deployed in very high capacity backbones. The detection algorithm in coherent receivers (such as DP-QPSK receivers) tracks many physical parameters of lightpaths, such as polarization mode dispersion (PMD), the nonlinear phase, and the chromatic dispersion of the signals, which are responsible for various impairments. In networks where some lightpaths are intensity-modulated (e.g., 10 Gb/s on-off keying) and others are coherent (e.g., 40 Gb/s or 100 Gb/s DP-QPSK), monitoring data available from the coherent receivers terminating established or active lightpaths can be used to estimate the QoT of lightpaths terminated at non-coherent receivers using the estimation framework proposed here and the QoT estimator presented in [14]. In networks equipped with coherent devices only, the coherent receivers can be used in lieu of additional hardware monitors, and the monitoring framework presented in this work can also be leveraged, in conjunction with a QoT estimator adapted to coherent systems.

III. ACTIVE MONITORING FRAMEWORK

The active monitoring technique consists of two components. First, we select the set of active lightpaths, and then we estimate the BERs of the unobserved lightpaths using both passive and active measurements. In Section III-A, we describe our inference methodology. In Section III-B, we present the active monitoring technique that establishes additional monitoring lightpaths to improve estimation accuracy. In Section III-C, we propose a hardware monitor placement algorithm.

A. Estimation methodology

We adapt two distinct estimation techniques to the problem at hand: the network kriging procedure of [8] and least-squares minimization with \( \ell_2 \)-norm regularization. The latter technique was formerly employed for more general network performance

---

\(^1\)In existing networks (and envisioned optical networks of the future), the network generally operates in a fully-loaded steady state. Although all wavelengths are not constantly lit, they are all consistently used. In this case, XPM and FWM effects depend only on network topology, in which case they can also be addressed using the framework described in this paper. It is possible that unused wavelengths remain consistently dark, making XPM and FWM effects state- and wavelength-dependent; such dependencies are much more difficult to handle and are out of the scope of this paper.
metric estimation in [9]. Both estimation techniques rely on the assumption of a linear relationship between link-level metrics and the path-level metrics. In transparent optical networks, neither BERs nor Q-factors are linear with respect to the number of spans over which a signal is transmitted, and the required linear relationship does not hold. However, the linearity assumption holds, to within an acceptable approximation error, for each of the four quantities $\mu_0$, $\mu_1$, $\sigma_0$, and $\sigma_1^2$, as indicated by the linear fits in Fig. 1. We can thus apply these link-level estimation strategies for each of these four quantities and combine the estimates to form an estimate of BER (note that we in fact form estimates of $\log(\text{BER})$, which is a more meaningful QoT metric for network operators).

Consider a transparent optical network of $n_L$ links and $N$ nodes where $n_p$ lightpaths are established. We denote by $y_p \in \mathbb{R}^{n_r}$ a column vector containing path-level QoT metrics ($\mu_0$, $\mu_1$, $\sigma_0$, or $\sigma_1^2$ for some lightpaths) and by $x \in \mathbb{R}^{n_r}$ the corresponding link-level metrics. Denote by $G_p$ the routing matrix that describes the network state, that is, $G_p \in \{0, 1\}^{n_p \times n_L}$, where $(G_p)_{i,j} = 1$ when lightpath $i$ traverses link $j$. We partition $y$ and $G_p$ into two components. For the $n_m$ lightpaths that terminate at one of the hardware monitor locations, we denote by $y_m$ and $G_m \in \{0, 1\}^{n_m \times n_L}$ the observed (directly measured) path metrics and corresponding routing matrix. For the other unobserved lightpaths, those that do not terminate at a hardware monitor location, we use the notation $y_n$ and $G_n \in \{0, 1\}^{(n_p-n_m) \times n_L}$. Thus we have $y_m = G_m x$ and $y_n = G_n x$.

We also establish a limited number $n_0$ of lightpath probes (active lightpaths) which terminate at nodes with hardware monitors. We defer the discussion of how to choose the routes of these paths until Section III-B. Denote by $G_a \in \{0, 1\}^{n_a \times n_L}$ the routing matrix of the active lightpaths. Let $y_a = G_a x$ be the column vector for metrics corresponding to these active lightpaths. Call $G_A = [G_m, G_a]^T$ (where $T$ denotes transpose) the routing matrix corresponding to all observed lightpaths and $y_A = [y_m^T, y_a^T]^T$ the column vector containing the metrics for all observed paths. The general estimation problem can be expressed as: given a routing matrix $G_p = [G_m^T, G_n^T]^T$, the end-to-end observations $y_A = G_A x$ where the link-level metrics $x$ are unknown, estimate all unobserved end-to-end metrics $y_n = G_n x$.

Denote the estimator $\hat{y}_n$. If we employ mean-squared error $E[||\hat{y}_n - y_n||_2^2]$ as the quality of the estimator, then the best estimator is given by the conditional expectation $E[y_n|y_A]$. This cannot be computed unless we have knowledge of the joint distributional structure of $y_n$ and $y_A$. We can make the task of finding a good estimator more tractable by focusing on linear estimators (where the estimator can be expressed in the form $B y_A$ for some matrix $B$). Deriving an expression for the best linear estimator does not require knowledge of the joint distribution of $y_n$ and $y_A$, but it does require knowledge of the mean and covariance of $x$. A natural solution is to estimate this mean from the available data, using for example generalized least squares. If enough data is available then the covariance matrix could also be estimated, but we assume a diagonal covariance matrix for simplicity. This leads to the following estimator:

$$\hat{y}_n = G_n G_A^T (G_A G_A^T)^+ y_A.$$  

Here $(\cdot)^+$ denotes a pseudo-inverse such as the Moore-Penrose inverse. A more general version of this estimator is presented in a paper called “network kriging” [8], and we will retain that name here. The complexity of the kriging estimator is dominated by the complexity of computing the pseudo-inverse in (1), that is, $O((n_m + n_a)^3)$.

The main problem with the network kriging estimator is that there is no positivity constraint (we know that the four metrics of interest are always positive quantities). This can be remedied by projecting onto the viable space of solutions or by incorporating the non-negativity constraint directly in the optimization. In formulating the second estimator, we add the constraint, but we also choose to incorporate an additional prior belief about the “best” explanation for the observed data.\footnote{Note that the pseudo-inverse in the network kriging estimator does select one of (infinitely) many possible values of $x$ that can explain the observed data. Thus in employing the kriging estimator we are implicitly specifying a definition of the “best” explanation. In our second estimator, “best” is defined in terms of an attribute of $x$ (minimum $\ell_2$-norm). In the kriging estimator, “best” is specified by the choice of the linear mapping function, $G_A^T (G_A G_A^T)^+$ that takes us from $y_A$ to $\hat{x}$.}

The second estimator we formulate in this paper is based on $\ell_2$-norm minimization. The associated optimization problem is:

$$\min_{x \in \mathbb{R}^{n_r}} ||r||_2^2 + ||x||_2^2$$  

subject to $G_A x + D_2 r = y_A, x \geq 0$.

In (2), $r$ is a regularization parameter and $D_2$ is a diagonal matrix specifying the error tolerance (how closely the path-level metric estimates must match the associated measurements). Problem (2) can be written as a convex quadratic program. Indeed, define vector $z = [r^T, x^T]^T$; then the objective function in (2) is simply $z^T I z = z^T z$, with linear constraints on $z$. In (2), we have $c = 0$, and the problem is convex since $G = I$ is positive definite. In that case, there are polynomial-time algorithms for solving (2) as well. In particular, there exist efficient solved software packages such as PDCO [16] that can solve problem (2).

The assumption of the (approximate) linear relationships $y_m = G_m x$, $y_n = G_n x$ and $y_a = G_a x$ allows us to use the network kriging and $\ell_2$-norm minimization procedures to estimate $y_n$ given $G_n, G_m, G_a$ and $y_m$. This estimation procedure is run in turn for $y_m, y_n, y_a \in \{\mu_0, \mu_1, \sigma_0, \sigma_1^2\}$ to return estimates for each of these four quantities for the unobserved lightpaths. Based on these, we can construct estimates for the $Q$-factors and BERs. Since the four metrics $\mu_0$, $\mu_1$, $\sigma_0$, and $\sigma_1^2$ are only approximately linear with respect to distance, additional estimation error will accrue from the linearity approximation.

B. Active monitoring

In this section, we address the task of choosing which lightpaths to activate for monitoring. We assume that hardware
monitors have been placed at fixed locations and that we have a limited budget of \( n_a \) active lightpaths. The task is to design an active routing matrix \( G_a \in \{0, 1\}^{n_a \times n_c} \), where each row corresponds to an active lightpath.

Our goal is to choose a set of active paths \( G_a \) in order to maximize the "energy" of the unobserved, passive lightpaths (described by \( G_p \)) captured by the measurements; this is equivalent to selecting paths to measure (i.e., rows of \( G_a \)) that span the space of \( G_p \) not captured by \( G_m \) and are as linearly independent from \( G_m \) as possible. We can construct such lightpaths by appending to each unobserved active lightpath whose performance is highly correlated with each monitor, but the size of this candidate set would quickly shrink and potentially provide information about multiple paths.

We initialize \( G_a \) as an empty matrix. At each step of the heuristic, we evaluate, for each candidate monitor location \( i \), the rowspace \( R_{G_m}^{(i)} \) (and a corresponding orthonormal basis \( B_{G_m}^{(i)} \)) that would result if we added to \( G_m \) all of the lightpaths that terminate on link \( i \). We then choose the monitor location \( i \) that maximizes the energy of the unobserved lightpaths \( ||G_a B_{G_p,G_m}^{(i)}||_F \). The algorithm terminates when we have selected \( M \) monitor locations.

IV. SIMULATIONS

We simulate the operation of a scaled-down version of the NSFNET topology (Fig. 2) with standard physical parameters (10 Gbps NRZ signals traveling over 70 km spans of 100% post-dispersion compensated SMF, noise figure of 6 dB, 8 wavelengths); The topology has a diameter close to 800 km, ensuring that end-to-end transparency is feasible even without FEC and with standard, non-advanced dispersion maps and modulation formats. Although the actual physical reach of signals with no regeneration in real networks can be much longer than 800 km thanks for instance to the utilization of FEC, advanced dispersion maps and modulation formats, our
models do not depend on those parameters. Only the details but not the essence of the conclusions would change. The network contains 42 unidirectional links and hence would require 42 monitors to gather a complete observation set about the QoT in the network.

Arriving traffic demands are drawn from a uniform distribution over all (source, destination) pairs. Demands are assumed to arrive according to a Poisson process with rate $\lambda$ and demand durations are exponentially distributed with mean $1/\mu$; numerical results are given for a network load of $\lambda/\mu = 50$ Erlang, which is a usual load for a network of this size. Note, however, that the techniques demonstrated in this paper do not depend on the demand arrival/duration processes.

The route taken by a new demand is determined by an adaptive route-wavelength assignment (RWA) algorithm. The selected route depends on the network state at the time a demand arrives; routes are not necessarily shortest paths, and two demands with the same source and destination do not necessarily take the same route. In each network state, the number of concurrent lightpaths in the network is approximately 50 (varying between 40 and 54 due to the randomness in the demand generation). The results presented in this section are averages over 500 network states.

Monitors are deployed at the locations determined by the monitor placement algorithm described in Section III-C. In Fig. 2, we indicate the locations of the top 15 monitors returned by the placement algorithm described in Section III-C. We use different symbols for the top 5 monitors (monitors 1-5), the next 5 monitors (monitors 6-10) and the again the next 5 monitors (monitors 11-15). Note that the proposed placement algorithm returns nested placements: if we run the placement algorithm to select, independently, $M$ and $M'$ locations such that $M < M'$, then the $M$ locations form a subset of the $M'$ locations returned by the placement algorithm. This property is useful in the context of incremental monitoring hardware deployment.

We first examine the accuracy of the estimators in terms of the relative mean squared error (RMSE) for $\log(\text{BER})$. Fig. 3 compares the accuracy of the two estimators presented in Section III-A for a fixed number of monitors (ten), varying the number of active lightpaths in the network. The kriging estimator generates negative estimates for the BER for 5-10 percent of the lightpaths, even when the number of active lightpaths is relatively high, i.e. 30. These are not physically meaningful values, so we map the $\log(\text{BER})$ to zero for these estimates. The accuracies of both estimators improve as more active lightpaths are injected. The estimator based on $\ell_2$-norm minimization outperforms the kriging-based estimator when the number of active lightpaths is relatively small (less than 15), but when more lightpath probes are employed the accuracies are comparable. The remainder of our performance analysis focuses on the $\ell_2$-norm minimization estimator.

Fig. 4(a) displays how the RMSE for the $\ell_2$-norm minimization estimator is affected by changes in both the number of active lightpaths and the number of monitors. Values for $n_a = 0$ active lightpaths denote passive monitoring results. Establishing even a relatively small number of active lightpaths (e.g., 15) results in a sharp decrease in the estimation error, especially when fewer monitoring devices are installed. There exists a trade-off between the number of physical devices and the number of active lightpaths required to achieve a specified estimation accuracy. For example, to achieve an estimation error of approximately 7%, one can either deploy 20 monitors and use passive monitoring, or deploy 5 monitors and use active monitoring with 25 active lightpaths. Note that this trade-off could be captured by a single cost function combining both estimation accuracy and number of established active lightpaths $n_a$ (for instance, a linear combination of those two metrics). Such a cost function would likely vary depending on the operator, hence, we are not proposing any such cost function but report detailed results for both estimation accuracy and $n_a$.

The estimation error does not converge to 0 as $n_a$ increases. It exhibits an error floor that decreases as the number of monitors increases. This can be attributed to the approximation error incurred through the linearization step. Indeed, Fig. 4(b) confirms this assessment. It shows the RMSE in the case where all four metrics $\mu_0$, $\mu_1$, $\sigma_0$, and $\sigma_1^2$ are artificially
When estimated metrics are artificially linearized, the error floor for large $n_a$ disappears.

The benefits of active monitoring are illustrated in Fig. 5 using a more concrete example. At the depicted point in time, 53 lightpaths are established in the network and only 9 of these are directly observed by 5 hardware monitors. We show the BER and the absolute error for the 44 unobserved lightpaths (in arbitrary order, labeled from 1 to 44). The plots compare passive monitoring with active monitoring employing $n_a = 20$ active lightpaths. In the case of passive monitoring, the estimation technique cannot generate estimates for 25 lightpaths, whereas with active monitoring, the BERs for only 2 lightpaths are not estimated. The technique cannot generate a meaningful estimate for any lightpath that does not share a link with any of the measured lightpaths. For nearly all lightpaths, the estimates obtained using active monitoring case are consistently closer to the real BER values than the passive monitoring estimates. There is no apparent correlation between the values of the BERs and the associated RMSE.

As discussed in Section III-B, the active monitoring strategy strives to decrease the “variance term” in the estimation error, which involves trying to construct an observation subspace

linearized, i.e. they are assigned values that are exactly linear with distance. With genuinely linear metrics, the estimation error converges to 0 as the amount of monitoring hardware or the number of active lightpaths, $n_a$, increases.

The benefits of active monitoring are illustrated in Fig. 5 using a more concrete example. At the depicted point in time, 53 lightpaths are established in the network and only 9 of these are directly observed by 5 hardware monitors. We show the BER and the absolute error for the 44 unobserved lightpaths (in arbitrary order, labeled from 1 to 44). The plots compare passive monitoring with active monitoring employing $n_a = 20$ active lightpaths. In the case of passive monitoring, the estimation technique cannot generate estimates for 25 lightpaths, whereas with active monitoring, the BERs for only 2 lightpaths are not estimated. The technique cannot generate a meaningful estimate for any lightpath that does not share a link with any of the measured lightpaths. For nearly all lightpaths, the estimates obtained using active monitoring case are consistently closer to the real BER values than the passive monitoring estimates. There is no apparent correlation between the values of the BERs and the associated RMSE.

As discussed in Section III-B, the active monitoring strategy strives to decrease the “variance term” in the estimation error, which involves trying to construct an observation subspace
that contains as much of the energy of the unobserved lightpaths as possible. We can measure how successful the algorithm is by assessing the average fraction of energy of each unobserved lightpath that is contained in the constructed space. The rank of the constructed space, \( [G_m^T, G_a^+]^T \), also provides an important indication. If the technique is placing monitors in good locations and activating informative active lightpaths, then the rank should increase rapidly as the number of monitors or active lightpaths is increased (indicating that the addition of resources is providing new information). The top panel of Fig. 6 depicts how the rank of \( [G_m^T, G_a^+]^T \) changes when we vary either the number of monitors or the number of active lightpaths. For low numbers of monitors and active lightpaths, the rank is much less than the number of lit links, making BER estimation inaccurate. However, as \( n_a \) increases, rank \( [G_m^T, G_a^+]^T \) increases with a slope of approximately 1, converging to rank \( [G_m^T, G_n^T, G_a^+]^T \). This latter value (depicted as a bold line with no markers) is the rank of the complete routing matrix, which includes observed, unobserved and active lightpaths. The bottom panel of Fig. 6 shows the average fraction of energy of the unobserved lightpaths that lies in the space spanned by the vectors \( [G_m^T, G_a^+]^T \). As desired, this fraction increases rapidly as the number of monitors is increased and as the number of active lightpaths is increased.

We now analyze the performance of the monitor placement algorithm, comparing it to a naive strategy of randomly placing monitors in the network. Fig. 7 compares the RMSE achieved by using these two placement strategies followed by application of the \( \ell_2 \)-norm minimization estimator. Performance is assessed for the two cases of passive monitoring \( (n_a = 0, \) top panel) and for 25 active lightpaths \( (n_a = 25, \) bottom panel), with the number of hardware monitors ranging from 5 to 35. To obtain the random placement curve, we averaged the RMSE for 20 random placements with 50 different network states. The error bars correspond to one standard deviation. The curve corresponding to our proposed monitoring strategy (which is deterministic) is obtained by averaging over 500 network states. The proposed algorithm achieves a 10-20% improvement in estimation accuracy compared to random placement.

The top panel of Fig. 8 displays the proportion \( n_m/n_p \) of (passive) lightpaths that are actually observed directly using a monitor when the proposed monitor placement is used. The bold line is the proportion of monitored links (there are 42 links in total). This line also corresponds to the expected fraction of lightpaths that would be observed using a uniform random monitor placement strategy. Using the monitor placement algorithm presented in Section III-C, we monitor more lightpaths (consistently 5-10% more) than would be captured if monitors were randomly placed. Although the number of directly monitored paths has some impact on estimation accuracy, of more interest is how well the monitored lightpaths capture the subspace occupied by the unobserved lightpaths. The bottom panel of Fig. 8 displays the average fraction of energy of the unobserved lightpaths that lies in the monitored subspace, comparing the cases when the proposed placement algorithm is employed to when monitors are placed on random links. It is clear that the greedy placement algorithm results in a monitored subspace that contains much more of the energy of the lightpaths in the network. This leads to the improved estimation performance observed in Fig. 7.
V. CONCLUSIONS

We showed that it is possible to decrease the number of hardware monitors needed to monitor the QoT of lightpaths in a transparent optical network by establishing carefully selected “active lightpaths”, without sacrificing estimation accuracy. In addition, we studied the sparse monitor placement problem and proposed an algorithm to select locations where hardware monitors should be installed in the network in order to improve estimation accuracy.

The estimation technique relies on a linearization of the quantities involved in the QoT computations. Generalizing our technique to more complex cases where linearity does not hold will be the subject of future work. The estimation technique currently assumes that wavelength-dependent impairments such as XPM and FWM are negligible. It also assumes that the physical layer effects of interest are statistically independent and identically distributed across channels. Accounting for channel dependence is a challenging but important topic of future research.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their feedback, which helped improve the manuscript.

Yvan Pointurier (S’02–M’06) received a Diplome d’Ingénieur from Ecole Centrale de Lille, France in 2002, a M.S. in Computer Science in 2002, and a Ph.D. in Electrical Engineering in 2006, both from the University of Virginia, USA. He spent two years at McGill University in Montreal, Canada as a Postdoctoral Fellow and then one year at Athens Information Technology, Greece, as a technical leader for a European Project. In 2009 he joined Alcatel-Lucent Bell Labs, France, as a research engineer. His research interests span design, optimization and monitoring of networks in general, and optical networks in particular.

Dr. Pointurier is a co-recipient of the Best Paper Award at the IEEE ICC 2006 Symposium on Optical Systems and Networks.

Mark J. Coates (M’99–SM’07) received the B.E. degree (first class honors) in computer systems engineering from the University of Adelaide, Australia, in 1995, and the Ph.D. degree in information engineering from the University of Cambridge, U.K., in 1999. He joined McGill University, Montreal, QC, Canada, in 2002, where he is currently an Associate Professor. He was awarded the Texas Instruments Postdoctoral Fellowship in 1999 and was a research associate and lecturer at Rice University, Houston, TX, from 1999 to 2001. His research interests include communication and sensor networks, statistical signal processing, and Bayesian and Monte Carlo inference.

Michael G. Rabbat (S’02–M’07) earned the B.Sc. from the University of Illinois at Urbana-Champaign (2001), the M.Sc. degree from Rice University (2003), and the Ph.D. from the University of Wisconsin–Madison (2006), all in electrical engineering. He is currently an Assistant Professor at McGill University. He was a visiting researcher at Applied Signal Technology, Inc., during the summer of 2003. He received the Best Paper Award (Signal Processing and Information Theory Track) at the 2010 IEEE Conference on Distributed Computing in Sensor Systems, Outstanding Student Paper Honorable Mention at the 2006 Conference on Neural Information Processing Systems, the Best Student Paper award at the 2004 ACM/IEEE Conference on Information Processing in Sensor Networks, and the Harold A. Peterson Thesis Prize. His research interests include network monitoring, network inference, and distributed information processing in sensor networks. He is currently an Associate Editor for the ACM Transactions on Sensor Networks.

REFERENCES