Endhost-Based Shortest Path Routing in Dynamic Networks: Supporting Materials

Ting He*, Dennis Goeckel†, Ramya Raghavendra*, and Don Towsley†
* IBM Research, Yorktown, NY, USA. Email: {the,rraghav}@us.ibm.com
† University of Massachusetts, Amherst, MA, USA.
Email: {goeckel@ecs,towsley@cs}.umass.edu
*

1 Introduction

This report contains supplementary materials in support of [1], including proofs and algorithm pseudo codes. We refer to the original paper for formulation and definitions.

2 Proof of Theorems

Theorem 2.1. *For mutually independent and i.i.d. link weights, the average regret of any $O(\log T)$-regret algorithm under coupled probing and routing satisfies*

$$\mathbb{E}[\mathcal{R}_T] \geq \min_{\Phi} \sum_{\mathbf{p}\in\mathcal{P}'} \Delta_{\mathbf{p}} u_T(\mathbf{p}) \tag{1}$$

for all sufficiently large T, where $\Delta_{\mathbf{p}} \triangleq \mathbb{E}[W_{\mathbf{p}} - W_{\mathbf{p}^*}]$ is the suboptimality of path $\mathbf{p}$, and $u_T(\mathbf{p})$ the average number of times $\mathbf{p}$ is used up to time T, constrained by

$$\begin{aligned} \Phi \triangleq \{(u_T(\mathbf{p}))_{\mathbf{p}\in\mathcal{P}'} :\ & u_T(\mathbf{p}) \geq 0,\ \forall \mathbf{p}\in\mathcal{P}'; \\ & \sum_{\mathbf{p}\in\mathcal{P}'} p_i u_T(\mathbf{p}) \geq \frac{\log T}{D_i},\ \forall i \in E'\}. \end{aligned} \tag{2}$$

Proof. The proof is based on a similar idea as in the proof of Theorem 3.1 in [2]: if we modify the link weights so that a competitive link becomes part of the optimal path and thus must be used most of the time by any good algorithm, then this link must be used sufficiently often ($\Omega(\log T)$ in T steps) by the same algorithm under the original link weights because we do not know which weight configuration is in effect, thus introducing a minimum learning cost.

Specifically, consider a competitive link $i \in E'$. For any $\epsilon > 0$, select a new link weight distribution L_i' such that a suboptimal path $\mathbf{p} \in \mathcal{P}'$ containing link i becomes optimal and $|D(L_i||L_i') - D_i| \leq \epsilon D_i$.

Consider any algorithm achieving $O(\log T)$ regret under arbitrary link configurations. Let $U_T(i)$ ($U_T'(i)$) denote the number of times link i is used up to time T under the original (new) configuration. The $O(\log T)$

*Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-06-3-0001. The views and conclusions are those of the authors and do not represent the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding copyright notation.

regret implies that $\mathbb{E}[T - U'_T(i)] = O(\log T)$ since link i is part of the optimal path under the new configuration. From the proof of Theorem 3.1 in [2], we have that under the original configuration,

$$\lim_{T\to\infty} \Pr\{U_T(i) < \frac{\log T}{(1+2\epsilon)(1+\epsilon)D_i}\} = 0, \tag{3}$$

i.e., $\liminf_{T\to\infty} \mathbb{E}[U_T(i)]/\log T \geq 1/D_i$. That is, for all sufficiently large T, each competitive link $i \in E'$ must be used at least $\log T/D_i$ times on the average.

Let $u_T(\mathbf{p}) \triangleq \mathbb{E}[U_T(\mathbf{p})]$ for each $\mathbf{p} \in \mathcal{P}'$ denote the average number of times that a suboptimal path $\mathbf{p}$ is used up to time T. The above result implies that the usage of all suboptimal paths, denoted by $u_T(\mathbf{p})$ ($\mathbf{p} \in \mathcal{P}'$), must result in a per-link usage of at least $\log T/D_i$ for each competitive link i, i.e., $(u_T(\mathbf{p}))_{\mathbf{p}\in\mathcal{P}'} \in \Phi$ for Φ defined in (2). This constraint holds for any algorithm with $O(\log T)$ regret. Since for a fixed algorithm, the average regret is given by $\sum_{\mathbf{p}\in\mathcal{P}'} \Delta_{\mathbf{p}} u_T(\mathbf{p})$, the minimum value of this function over all possible $(u_T(\mathbf{p}))_{\mathbf{p}\in\mathcal{P}'} \in \Phi$ thus bounds the minimum average regret. That is, the average regret of an arbitrary $O(\log T)$-regret algorithm is lower bounded as

$$\mathbb{E}[\mathcal{R}_T] \geq \min_{\Phi} \sum_{\mathbf{p}\in\mathcal{P}'} \Delta_{\mathbf{p}} u_T(\mathbf{p}). \tag{4}$$

□

Corollary 2.2. *The lower bound in (1) is $\Omega(\log T)$ and $O(\min(N, M)\log T)$. Specifically, for $D_{min} \triangleq \min_{i\in E'} D_i$,*

$$\frac{\Delta_{min}}{D_{min}} \log T \leq \min_{\Phi} \mathbf{\Delta}^T \mathbf{u}_T \leq \frac{\Delta_{max}}{D_{min}} \min(M,\ N) \log T.$$

Proof. The lower bound is simply because $\mathbf{\Delta}^T\mathbf{u}_T \geq \Delta_{\min} \sum_{\mathbf{p}\in\mathcal{P}'} u_T(\mathbf{p})$ and $\sum_{\mathbf{p}\in\mathcal{P}'} u_T(\mathbf{p}) \geq \sum_{\mathbf{p}\in\mathcal{P}'} p_i u_T(\mathbf{p}) \geq \log T/D_i$ for all $i \in E'$.

For the upper bound, first note that $u_T(\mathbf{p}) = \max_{i\in E': p_i=1} \log T/D_i$ is a feasible solution. Thus,

$$\min_{\Phi} \mathbf{\Delta}^T \mathbf{u}_T \leq \sum_{\mathbf{p}\in\mathcal{P}'} \frac{\Delta_{\mathbf{p}} \log T}{\min_{i\in E':p_i=1} D_i} \leq \frac{\Delta_{\max}}{D_{\min}} M \log T. \tag{5}$$

Moreover, since each path contains at least one link, the total number of times of using suboptimal paths does not need to exceed the total number of times of using competitive links, i.e., $\min_\Phi \sum_{\mathbf{p}\in\mathcal{P}'} u_T(\mathbf{p}) \leq \sum_{i\in E'} \log T/D_i$. Thus,

$$\min_{\Phi} \mathbf{\Delta}^T \mathbf{u}_T \leq \Delta_{\max} \sum_{i\in E'} \frac{\log T}{D_i} \leq \frac{\Delta_{\max}}{D_{\min}} N \log T. \tag{6}$$

□

Theorem 2.3. *The average regret of OSPR is $O(N^4)$ and specifically,*

$$\mathbb{E}[\mathcal{R}_T] \leq \Delta_{max} N \left(\frac{2c}{\Delta_{min}^2} H^2 N + \frac{4}{\Delta_{min}^2} H^3 + 1 \right), \tag{7}$$

where c is a constant satisfying $c \geq \log M/N$.

Proof. Define $U(t)$ as the number of times suboptimal paths are used for routing in the first t steps. Since we always probe the least measured link, we have that $\min_i m_i(t) \geq \lfloor t/N \rfloor$. For any $n \geq 0$,

$$\mathbb{E}[U(T)] \leq nN + \sum_{t=nN+1}^{T} \Pr\{\exists \mathbf{p} \in \mathcal{P}' : \sum_i p_i \hat{l}_i(t-1) \leq \sum_i p_i^* \hat{l}_i(t-1)\} \tag{8}$$

$$\leq nN + \sum_{t=nN+1}^{T} \sum_{\mathbf{p}\in\mathcal{P}'} \Pr\{\sum_i p_i \hat{l}_i(t-1) \leq \sum_i p_i^* \hat{l}_i(t-1)\}, \tag{9}$$

where (8) is because we have chosen $\hat{\mathbf{p}}(t)$ to be the empirically shortest path based on $\hat{\mathbf{l}}(t-1)$. We leverage the Chernoff-Hoeffding bound to bound (9).

Chernoff-Hoeffding bound [3]: Let $X_1, \dots, X_n$ be random variables with common range [0, 1] such that $\mathbb{E}[X_t|X_1, \dots, X_{t-1}] = \mu$, $\forall 1 \le t \le n$. Let $S_n = \sum_{i=1}^n X_i$. Then for all $a \ge 0$,

$$\Pr\{S_n \ge n\mu + a\} \le e^{-2a^2/n}, \quad \Pr\{S_n \le n\mu - a\} \le e^{-2a^2/n}. \tag{10}$$

Using this result, we will show that if $m_i \ge k$ for all i, then

$$\Pr\{\sum_i p_i \hat{l}_i \le \sum_i p_i^* \hat{l}_i\} \le 2He^{-2k(\Delta_{\min}/2H)^2}. \tag{11}$$

This is because by union bound, we have

$$\begin{aligned}\Pr\{\sum_i p_i \hat{l}_i \le \sum_i p_i^* \hat{l}_i\} &\le \Pr\{\sum_i p_i \hat{l}_i \le \sum_i p_i l_i - \frac{\Delta_{\min}}{2}\} + \Pr\{\sum_i p_i^* \hat{l}_i \ge \sum_i p_i^* l_i + \frac{\Delta_{\min}}{2}\} \\ &\le \sum_{i:p_i=1} \Pr\{\hat{l}_i - l_i \le -\frac{\Delta_{\min}}{2\sum_i p_i}\} + \sum_{i:p_i^*=1} \Pr\{\hat{l}_i - l_i \ge \frac{\Delta_{\min}}{2\sum_i p_i^*}\} \\ &\le H \Pr\{\hat{l}_i - l_i \le -\frac{\Delta_{\min}}{2H}\} + H \Pr\{\hat{l}_i - l_i \ge \frac{\Delta_{\min}}{2H}\}.\end{aligned}$$

Then applying the Chernoff-Hoeffding bound to $\hat{l}_i$ and relaxing it using $m_i \ge k$ give (11).

Applying (11) to (9), noting $m_i(t-1) \ge \lfloor (t-1)/N \rfloor \ge \frac{t-1}{N} - 1$, gives

$$\begin{aligned}\mathbb{E}[U(T)] &\le nN + \sum_{t=nN+1}^{T} 2MHe^{(1+N)\Delta_{\min}^2/(2NH^2)} \cdot e^{-t\Delta_{\min}^2/(2NH^2)} \\ &\le nN + \frac{2MHe^{-(n-1)N\Delta_{\min}^2/(2NH^2)}}{1 - e^{-\Delta_{\min}^2/(2NH^2)}}.\end{aligned} \tag{12}$$

This bound can grow exponentially in N because of the M-factor. We can, however, select n such that the product $Me^{-(n-1)N\Delta_{\min}^2/(2NH^2)} = O(1)$. Specifically, if $M \le e^{cN}$, setting $n = 1 + 2cNH^2/\Delta_{\min}^2$ gives

$$\mathbb{E}[U(T)] \le \frac{2c}{\Delta_{\min}^2} N^2H^2 + N + \frac{2H}{1 - e^{-\Delta_{\min}^2/(2NH^2)}} \approx \frac{2c}{\Delta_{\min}^2} N^2H^2 + N + \frac{4}{\Delta_{\min}^2} NH^3 \tag{13}$$

for $NH^2 \gg \Delta_{\min}^2$.

Finally, since we use suboptimal paths $U(T)$ times in the first T steps, each incurring at most $\Delta_{\max}$ extra weight, the average regret is bounded as

$$\mathbb{E}[\mathcal{R}_T] \le \Delta_{\max} \mathbb{E}[U(T)] \le \Delta_{\max} N \left(\frac{2c}{\Delta_{\min}^2} H^2 N + \frac{4}{\Delta_{\min}^2} H^3 + 1 \right). \tag{14}$$

□

Theorem 2.4. *For mutually independent and temporally i.i.d. link weights, the average regret of any constant-regret (in T) algorithm under decoupled probing and routing satisfies*

$$\mathbb{E}[\mathcal{R}_T] \ge \frac{\Delta_{min}(1 - e^{-\delta_{min}T})}{1 - e^{-\delta_{min}}} \tag{15}$$

for all sufficiently large T.

Proof. For any given algorithm with constant regret in T, let $U(T)$ denote the number of times suboptimal paths are used up to time T. We have

$$\mathbb{E}[U(T)] = \sum_{t=1}^{T} \Pr\{\exists \mathbf{p} \in \mathcal{P}' \text{ appearing optimal at time } t\}. \tag{16}$$

Consider the event that the empirical link weight distributions $\mathbf{L}'$ make a suboptimal path $\mathbf{p} \in \mathcal{P}'$ appear better than the optimal path, i.e., $\sum_i p_i \mathbb{E}[L'_i] < \sum_i p^*_i \mathbb{E}[L'_i]$. By Sanov's theorem [4], we know that this happens with probability $e^{-\sum_i m_i D(L'_i||L_i)}$ for all m_i's large enough, and this approximation is accurate to the first order in the exponent. At time t, since $m_i(t-1) \le t-1$, this probability is lower bounded by $e^{-(t-1)\sum_i D(L'_i||L_i)} = e^{-(t-1)D(\mathbf{L}'||\mathbf{L})}$.

Let $\Psi_{\mathbf{p}} \triangleq \{\mathbf{L}' : \sum_i p_i \mathbb{E}[L'_i] < \sum_i p^*_i \mathbb{E}[L'_i]\}$ for each $\mathbf{p} \in \mathcal{P}'$. Then $\Pr\{\exists \mathbf{p} \in \mathcal{P}'$ appearing optimal at $t\}$ is lower bounded by

$$\max_{\mathbf{p}\in\mathcal{P}'} \max_{\mathbf{L}'\in\Psi_{\mathbf{p}}} \Pr\{\text{empirically, links} \sim \mathbf{L}'\} \ge e^{-(t-1)\min_{\mathcal{P}'}\min_{\Psi_{\mathbf{p}}} D(\mathbf{L}'||\mathbf{L})} = e^{-(t-1)\delta_{\min}}. \tag{17}$$

Applying this to (16) gives

$$\mathbb{E}[U(T)] \ge \frac{1 - e^{-\delta_{\min}T}}{1 - e^{-\delta_{\min}}}. \tag{18}$$

Then substituting this bound into $\mathbb{E}[\mathcal{R}_T] \ge \Delta_{\min}\mathbb{E}[U(T)]$ yields the final result.

□

Corollary 2.5. *The average regret of OSPR with S-path probing is $O(N^4/S)$ and specifically,*

$$\mathbb{E}[\mathcal{R}_T] \le \frac{\Delta_{max} N}{S}\left(\frac{2cNH^2}{\Delta^2_{min}} + \frac{4H^3}{\Delta^2_{min}} + 1\right), \tag{19}$$

assuming $c \ge \log M/N$.

Proof. The proof is analogous to that of Theorem 2.3. Still let $U(t)$ denote the number of times of using suboptimal paths in the first t steps. With multi-path probing, the number of measurements per link satisfies $\min_i m_i(t) \ge \lfloor St/N \rfloor \ge St/N - 1$. Applying this to (11) gives

$$\Pr\{\sum_i p_i \hat{l}_i(t-1) \le \sum_i p^*_i \hat{l}_i(t-1)\} \le 2H \exp\left(-2\left(\frac{S(t-1)}{N} - 1\right)\left(\frac{\Delta_{\min}}{2H}\right)^2\right). \tag{20}$$

Substituting (20) into (9) gives

$$\begin{aligned}\mathbb{E}[U(T)] &\le nN + \sum_{t=nN+1}^{T} 2MHe^{(S+N)\Delta^2_{\min}/(2NH^2)} \cdot e^{-tS\Delta^2_{\min}/(2NH^2)} \\ &\le nN + \frac{2MHe^{-(Sn-1)\Delta^2_{\min}/(2H^2)}}{1 - e^{-S\Delta^2_{\min}/(2NH^2)}}.\end{aligned}$$

If $M \le e^{cN}$, then setting $n = \frac{1}{S}(1 + 2cNH^2/\Delta^2_{\min})$ yields

$$\mathbb{E}[U(T)] \le \frac{N}{S}\left(\frac{2cNH^2}{\Delta^2_{\min}} + 1\right) + \frac{2H}{1 - e^{-S\Delta^2_{\min}/(2NH^2)}} \approx \frac{N}{S}\left(\frac{2cNH^2}{\Delta^2_{\min}} + \frac{4H^3}{\Delta^2_{\min}} + 1\right) \tag{21}$$

for $NH^2 \gg S\Delta^2_{\min}$. The final result is then obtained by $\mathbb{E}[\mathcal{R}_T] \le \Delta_{\max}\mathbb{E}[U(T)]$.

□

Corollary 2.6. *The average regret of OSPR-UP for source-destination pair f is bounded as*

$$\mathbb{E}[\mathcal{R}_{T,f}] \leq \max_{i\in E_f} \frac{\Delta_{max,f} N_f \left(c_f N_f + \frac{\Delta^2_{min,f}}{2H_f^2}\sum_s C_{is}\right)}{\frac{N_f \Delta^2_{min,f}}{2H_f^2}\sum_s \frac{C_{is}}{N_s}} + \frac{4\Delta_{max,f} H_f^3}{\Delta^2_{min,f}\min_{i\in E_f}\sum_s \frac{C_{is}}{N_s}} \tag{22}$$

for $c_f \geq \log M_f / N_f$.

Proof. Consider a source-destination pair f; subscript f will be omitted if there is no ambiguity. The key is to bound the minimum link sample size $\min_{i\in E_f} m_i(t)$ for general C_{is}'s. By definition, $\sum_i m_{is}(t) \geq t$ for all $s = 1, \ldots, S$. In the worst case $\sum_i m_{is}(t) = t$, step 5 in Algorithm 4 iteratively solves the following integer linear programming (ILP)[1]:

$$\begin{aligned} &\max\ k \\ &\text{s.t.} \sum_{s=1}^{S} m_{is}(t) C_{is} \geq k, \quad i = 1, \ldots, N, \\ &\qquad \sum_{i=1}^{N} m_{is}(t) \leq t, \quad s = 1, \ldots, S, \\ &\qquad m_{is}(t) \in \mathbb{N}, \quad \forall i,\ \forall s. \end{aligned} \tag{23}$$

That is, distributing the t probing paths of each source-destination pair s to evenly cover the links, such that the minimum link sample size $\min_i \sum_{s=1}^{S} m_{is}(t) C_{is}$ is maximized. In particular, consider $\min_{i\in E_f}\sum_{s=1}^{S} m_{is}(t) C_{is}$, the minimum link sample size for source-destination pair f. Denote its maximum value under the constraints of (23) by k^*. Moreover, note that $m_{is}(t) = \lfloor t/N_s \rfloor C_{is}$ is a feasible solution to (23). Thus, $\min_{i\in E_f} m_i(t) \geq k^* \geq \min_{i\in E_f}\sum_{s=1}^{S}\lfloor t/N_s\rfloor C_{is} \triangleq \kappa(t)$.

The rest of the proof follows that of Theorem 2.3. Specifically, applying

$$\min_{i\in E_f} m_i(t-1) \geq \kappa(t-1) \geq \min_{i\in E_f}\left(\sum_s \frac{tC_{is}}{N_s} - \sum_s \frac{(1+N_s)C_{is}}{N_s}\right)$$

into (11) and then into (9) yields

$$\begin{aligned} \mathbb{E}[U(T)] &\leq \max_{i\in E_f} nN_f + \sum_{t=nN_f+1}^{T} 2MH\exp\left(-\frac{\Delta^2_{\min}}{2H^2}\left(t\sum_s \frac{C_{is}}{N_s} - \sum_s \frac{(1+N_s)C_{is}}{N_s}\right)\right) \\ &\leq \max_{i\in E_f} nN_f + \frac{2HM\exp\left(\frac{\Delta^2_{\min}}{2H^2}\sum_s \frac{(N_s - nN_f)C_{is}}{N_s}\right)}{1-\exp\left(-\frac{\Delta^2_{\min}}{2H^2}\sum_s\frac{C_{is}}{N_s}\right)}. \end{aligned} \tag{24}$$

For $M \leq e^{cN_f}$, setting

$$n = \max_{i\in E_f}\frac{cN_f + \frac{\Delta^2_{\min}}{2H^2}\sum_s C_{is}}{\frac{N_f\Delta^2_{\min}}{2H^2}\sum_s \frac{C_{is}}{N_s}} \tag{25}$$

[1] Here $\mathbb{N}$ denotes the set of natural numbers.

makes $M \exp\left(\frac{\Delta^2_{\min}}{2H^2}\sum_s \frac{(N_s - nN_f)C_{is}}{N_s}\right) \leq 1$ for all $i \in E_f$. Therefore, (24) becomes

$$\mathbb{E}[U(T)] \leq \max_{i\in E_f} nN_f + \frac{2H}{1-\exp\left(-\frac{\Delta^2_{\min}}{2H^2}\sum_s \frac{C_{is}}{N_s}\right)} \approx nN_f + \frac{4H^3}{\Delta^2_{\min} \min_{i\in E_f}\sum_s \frac{C_{is}}{N_s}} \tag{26}$$

for $S\Delta^2_{\min} \ll H^2 \min_s N_s$. Finally, combining (26) with $\mathbb{E}[\mathcal{R}(T)] \leq \Delta_{\max}\mathbb{E}[U(T)]$ and substituting (25) gives the result.

□

Corollary 2.7. *Consider a variant of OSPR where a probed link is excluded from probing for K slots*[2]*. If link weights $(L_i(t))_{t=1}^{\infty}$ are K-dependent and $K < N$, then the regret of this variant of OSPR has the same upper bound as in (7).*

Proof. The modified probing process guarantees that samples used to estimate mean link weights are mutually independent. Thus, the same proof in Theorem 2.3 applies, except that the minimum link sample size at slot t is now bounded by $\min_i m_i(t) \geq \lfloor t/\max(N, K+1)\rfloor$. As long as $K < N$, this bound is still $\lfloor t/N\rfloor$, the same as that in the proof of Theorem 2.3. Therefore, the upper bound in Theorem 2.3 still holds.

□

Corollary 2.8. *If probes are delayed by no more than $(\beta-1)N$ slots ($\beta \geq 1$), the average regret of OSPR remains $O(N^4)$ and specifically,*

$$\mathbb{E}[\mathcal{R}_T] \leq \Delta_{max} N\left(\frac{2c}{\Delta^2_{min}}H^2N + \frac{4}{\Delta^2_{min}}H^3 + \beta\right), \tag{27}$$

where the terms are defined the same as in (7).

Proof. The proof follows the same arguments as in the proof of Theorem 2.3, except that the minimum link sample size at slot t is now bounded by $\min_i m_i(t) \geq \lfloor \frac{t}{N} - \beta + 1\rfloor \geq \frac{t}{N} - \beta$. Applying this bound to (11) and then to (9) gives

$$\begin{aligned}\mathbb{E}[U(T)] &\leq nN + \sum_{t=nN+1}^{T} 2MHe^{-2(\frac{t-1}{N}-\beta)(\Delta_{\min}/2H)^2} \\ &\leq nN + \frac{2MHe^{-(n-\beta)\Delta^2_{\min}/(2H^2)}}{1-e^{-\Delta^2_{\min}/(2NH^2)}}.\end{aligned} \tag{28}$$

For $M \leq e^{cN}$, setting $n = \beta + 2cNH^2/\Delta^2_{\min}$ yields

$$\mathbb{E}[U(T)] \leq \frac{2c}{\Delta^2_{\min}}N^2H^2 + \beta N + \frac{4}{\Delta^2_{\min}}NH^3. \tag{29}$$

This implies that the regret bound is the same as (14) except that the constant term within the parentheses is changed from 1 to β. □

Proposition 2.9. *If link weights in the control/sliding windows are i.i.d. respectively with the same mean, then*

$$\Pr\left\{|\hat{l}_{i,c} - \hat{l}_{i,s}| \geq \sqrt{\frac{2}{\omega}\log\left(\frac{2}{\alpha}\right)}\right\} \leq \alpha. \tag{30}$$

[2]An excluded link can still be on a probing path, but it cannot be considered as a least measured link, and its measurement will not be used for link quality estimation.

Algorithm 1 Learning with Linear Cost (LLC)

Require: Candidate paths $\mathcal{P}$.

Ensure: Select a path to route and probe at each time step and compute estimated mean link weights $(\hat{l}_i)_{i=1}^N$.

Initialization: select paths $\mathbf{p}(1), \ldots, \mathbf{p}(N)$ s.t. $p_i(i) = 1$ to measure each link at least once, compute $(\hat{l}_i,\, m_i)_{i=1}^N$ accordingly

for $t = N+1, \ldots$ **do**

  Select the path $\mathbf{p}(t)$ such that $\mathbf{p}(t) = \arg\min_{\mathbf{p}\in\mathcal{P}} \sum_i p_i \left(\hat{l}_i - \sqrt{\frac{(H+1)\log t}{m_i}}\right)$

  Update $(\hat{l}_i,\, m_i)_{i=1}^N$ accordingly

Algorithm 2 Online SPR (OSPR)

Require: Candidate paths $\mathcal{P}$.

Ensure: Select a path to route and a path to probe at each time step, and compute estimated mean link weights $(\hat{l}_i)_{i=1}^N$.

Initialization: for $t = 1, \ldots, N$,

Route over a randomly selected path

Probe paths $\mathbf{p}(1), \ldots, \mathbf{p}(N)$ s.t. $p_i(i) = 1$ to measure each link at least once

Compute $(\hat{l}_i,\, m_i)_{i=1}^N$ accordingly

for $t = N+1, \ldots$ **do**

  Route over path $\hat{\mathbf{p}}(t) = \arg\min_{\mathbf{p}\in\mathcal{P}} \sum_i p_i \hat{l}_i$

  Probe the path $\mathbf{p}(t)$ containing the least measured link, i.e., link j s.t. $m_j = \min_i m_i$

  Update $(\hat{l}_i,\, m_i)_{i=1}^N$ accordingly

Proof. We prove the result by bounding the deviation between $\hat{l}_{i,c}$ and $\hat{l}_{i,s}$. For any $\epsilon > 0$, we have

$$\begin{aligned}\Pr\{|\hat{l}_{i,c} - \hat{l}_{i,s}| \geq \epsilon\} &\leq \Pr\{|\hat{l}_{i,c} - l_i| \geq \frac{\epsilon}{2}\} + \Pr\{|\hat{l}_{i,s} - l_i| \geq \frac{\epsilon}{2}\} \\ &\leq 2e^{-\omega\epsilon^2/2}, \end{aligned} \tag{31}$$

where (31) is by applying the Chernoff-Hoeffding bound (10). Then setting $\epsilon = \sqrt{\frac{2}{\omega}\log\left(\frac{2}{\alpha}\right)}$ makes the upper bound equal to α.

□

3 Pseudo Code for Algorithms

Algorithm 1 is the coupled probing/routing algorithm proposed in [5]. Algorithm 2 is the basic version of our decoupled probing and routing algorithm based on single-source learning. Algorithms 3 and 4 are the extended versions of the above for joint learning of multiple sources, with/without coordination in probing path selection. Algorithm 5 is the window-based detection procedure used to trigger re-learning in the case of fundamental changes in link qualities.

References

[1] T. He, D. Goeckel, R. Raghavendra, and D. Towsley, "Endhost-Based Shortest Path Routing in Dynamic Networks: An Online Learning Approach," 2012, under submission.

[2] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically Efficient Allocation Rules for the Multiarmed Bandit Problem with Multiple Plays—Part I: I.I.D. Rewards," *IEEE Trans. Automatic Control*, 1987.

[3] D. Pollard, *Convergence of Stochastic Processes*, 1984.

Algorithm 3 Online SPR under Coordinated Probing (OSPR-CP)

Require: Candidate paths $\mathcal{P}_f$ for each source-destination pair ($f = 1, \ldots, S$).
Ensure: At each time step, globally select S paths to probe to compute estimated mean link weights $(\hat{l}_i)_{i=1}^N$ and locally select a path to route for each source.

Probing:
Initialization: probe each link at least once, compute $(\hat{l}_i,\, m_i)_{i=1}^N$, broadcast $(\hat{l}_i)_{i=1}^N$ to all the sources
for each time step t **do**
  for $s = 1, \ldots, S$ **do**
    Probe the path $\mathbf{p} \in \mathcal{P} \triangleq \bigcup_{f=1}^S \mathcal{P}_f$ containing the globally least measured link
    Update $(\hat{l}_i,\, m_i)_{i=1}^N$ accordingly
  Broadcast $(\hat{l}_i)_{i=1}^N$ to all the sources

Routing: for a source-destination pair f after receiving initial $(\hat{l}_i)_{i=1}^N$
for each time step t **do**
  Route over path $\hat{\mathbf{p}}_f(t) = \arg\min_{\mathbf{p}\in\mathcal{P}_f} \sum_i p_i \hat{l}_i$
  Receive updated $(\hat{l}_i)_{i=1}^N$

Algorithm 4 Online SPR under Uncoordinated Probing (OSPR-UP)

Require: Candidate paths $\mathcal{P}_f$ for a source-destination pair of interest.
Ensure: Select a path to route and a path to probe at each time step, and compute estimated mean link weights $(\hat{l}_i)_{i=1}^N$.

Initialization: probe each link covered by $\mathcal{P}_f$ at least once, compute $(\hat{l}_{i,f},\, m_{i,f})_{i=1}^N$, share it with the other sources
for each time step t **do**
  Compute the aggregate link estimate: $\hat{l}_i = (\sum_{s=1}^S \hat{l}_{i,s} m_{i,s})/(\sum_{s=1}^S m_{i,s})$ for all i
  Route over path $\hat{\mathbf{p}}_f(t) = \arg\min_{\mathbf{p}\in\mathcal{P}_f} \sum_i p_i \hat{l}_i$
  Probe the path $\mathbf{p}_f(t) \in \mathcal{P}_f$ containing the globally least measured (in terms of $m_i = \sum_{s=1}^S m_{i,s}$) link among those covered by $\mathcal{P}_f$
  Update $(\hat{l}_{i,f},\, m_{i,f})_{i=1}^N$ accordingly and share it with the other sources

Algorithm 5 Window-based Change Detection (WCD)

Require: Window size ω, false alarm constraint α.
Ensure: Detect changes on each link i and restart learning upon detection.

for each new measurement of link i **do**
  Update $\hat{l}_{i,s}$
  if $|\hat{l}_{i,c} - \hat{l}_{i,s}| \geq \sqrt{\frac{2}{\omega}\log\left(\frac{2}{\alpha}\right)}$ **then**
    reset $(\hat{l}_i,\, m_i)$ (and also $\hat{l}_{i,c}$, $\hat{l}_{i,s}$)
  else
    update $(\hat{l}_i,\, m_i)$ according to the measurement

[4] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

[5] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards and Individual Observations," *IEEE/ACM Trans. Networking*, 2012.