Performance Analysis of a Data-Driven Quality-of-Transmission Decision Approach on a Dynamic Multicast-Capable Metro Optical Network

T. Panayiotou, S. P. Chatzis, and G. Ellinas

J. OPT. COMMUN. NETW./VOL. 9, NO. 1/JANUARY 2017

Speaker: Tao Gao

2017-12-1

Group Meeting Presentation

Quality of Transmission

- It is efficient to provision requests with high bandwidth requirement in optical networks.
- **Physical layer impairments** (PLIs) must be considered (e.g. amplified spontaneous emission (ASE) noise, crosstalk, optical filter concatenation, and polarization mode dispersion (PMD)).
- Conventional Q-factor model is based on the true measurement of the PLIs on a transparent optical network (not self-adaptive).
- Utilizing the neural network (NN), QoT data of previously established connections can be analyzed to find a QoT decision with high accuracy, which is **independent from the PLIs and self-adaptive**.

Data-driven QoT Approach

- First proposed in [1]: a static system, a single wavelength (simple model)

Extended in this paper:

- The connections arrive and terminate in a **dynamic** fashion
- On a impairment-aware unicast/multicast routing and wavelength assignment (IA-UMC-RWA) system in which multiple wavelengths are present

More details:

- Training data are generated from the dynamic IA-UMC-RWA algorithm
- Accuracy is evaluated by comparing the QoT decisions of the data-driven model to the QoT decisions of the Q-factor model (IA-UMC-RWA)

[1] T. Panayiotou, G. Ellinas, and S. P. Chatzis, "A data-driven QoT decision approach for multicast connections in metro optical networks," in Proc. Optical Network Design and Modeling (ONDM), Cartagena, Spain, May 2016.

Pattern Features

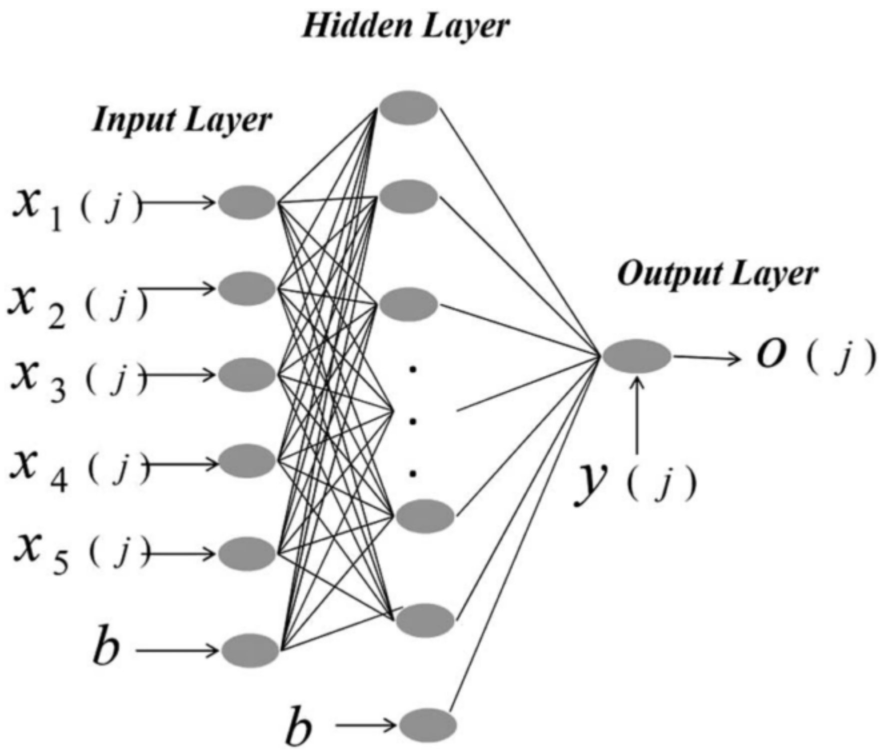
- Input vector **x**, discrete target value y
- Historical data set $\mathcal{D} = \{(\boldsymbol{x}(j), y(j)) | j = 1, \ldots, n\}$

$$\boldsymbol{x}(j)^T = [x_1(j), x_2(j), x_3(j), x_4(j), x_5(j), x_6(j)].$$

- $x_1(j)$ is the nominal path length of j;
- $x_2(j)$ is the number of erbium doped fiber amplifiers (EDFAs) in j;
- $x_3(j)$ is the nominal maximum link length of j;
- $x_4(j)$ is the degree of the destination node in j;
- $x_5(j)$ is the nominal wavelength on which j is established;
- $x_6(j)$ is the bias b of the first layer of the neural network.

Neural Networks with Dropout

- Dropout is a regularization method for preventing units from co-adapting too much by randomly dropping units from the neural network during training.



$$z_k^{(l+1)} = w_k^{(l+1)} o^l + b_k^{(l+1)},$$

$$o_k^{(l+1)} = f(z_k^{(l+1)}),$$

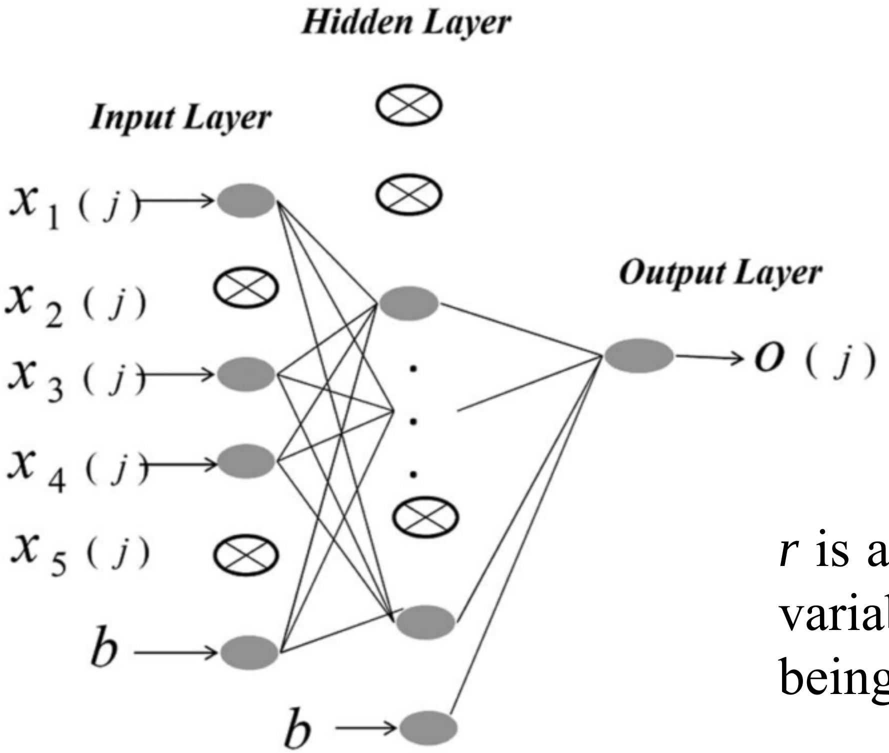
z^l is the vector of the input to layer l and o^l is the output from layer l, w^l is the weight, and b^l is the biases

UCDAVIS NETWORKS RESEARCH LAB

Neural Networks with Dropout

f is the **activation function**

$$f(x) = \tanh(x) = \frac{2}{1+\exp(-2x)} - 1.$$



$$r_v^{(l)} \sim \text{Bernoulli}(p),$$

$$\tilde{\boldsymbol{o}}^{(l)} = \boldsymbol{r}^{(l)} * \boldsymbol{o}^{(l)},$$

$$z_k^{(l+1)} = \boldsymbol{w}_k^{(l+1)} \tilde{\boldsymbol{o}}^{(l)} + b_k^{(l+1)},$$

$$o_k^{(l+1)} = f(z_k^{(l+1)}),$$

r is a vector of independent Bernoulli random variables, each of which has probability p of being 1

Neural Networks with Dropout

Mean-square loss function $$E = \sum_{j=1}^{n} (y(j) - o(j))^2,$$

For learning, **Adam algorithm** is used :

- ✓ Adam algorithm is an efficient stochastic optimization method that only requires first-order gradients with little memory requirements
- ✓ It is well suited for problems that are large in terms of data and/or parameters
- ✓ It has been shown to outperform other stochastic optimization methods

D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. for Learning Representations, San Diego, CA, 2015.

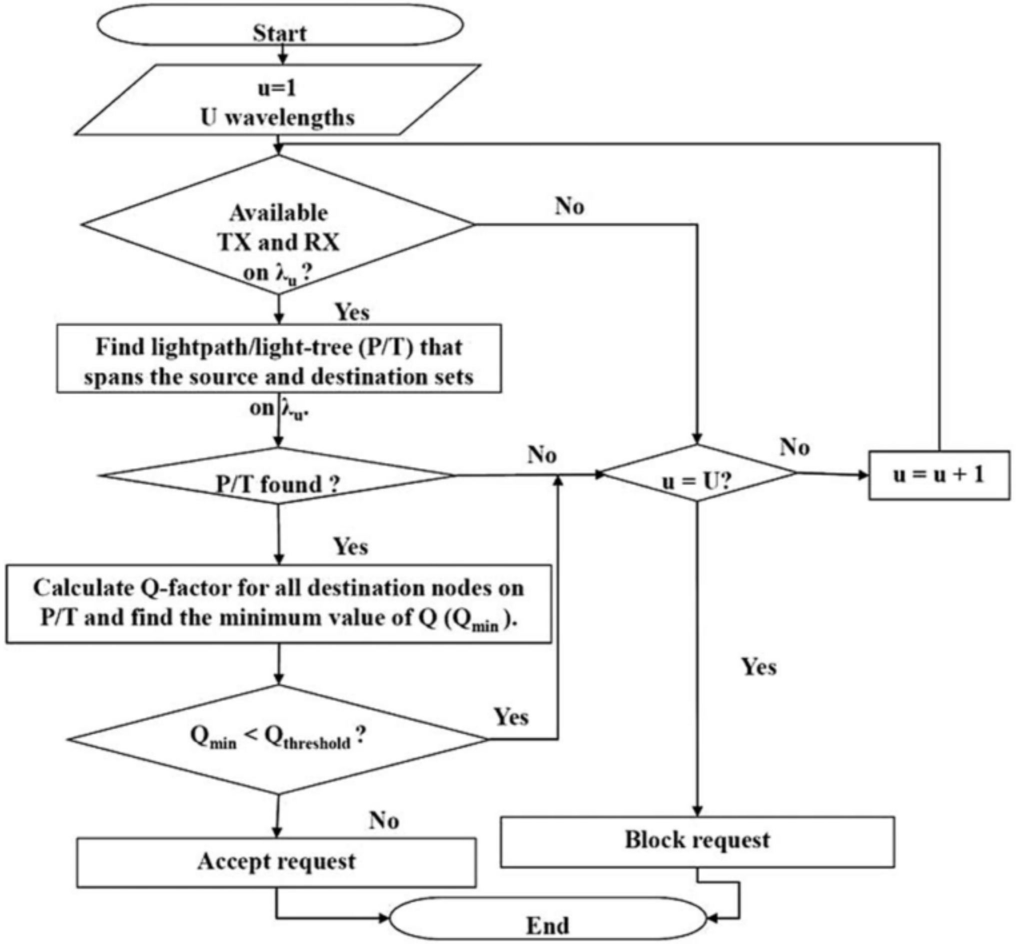
Data Set Generation

NETWORK STATISTICS

Number of nodes	14
Number of bidirectional links	50
Average distance (km)	67
Maximum distance (km)	100
Minimum distance (km)	20
Average node degree	7.15
Minimum node degree	4
Maximum node degree	10
Diameter (km)	160
Diameter (hops)	3

- Requests were generated with the multicast group sizes varying between 1 and 7
- Requests subjects to a Poisson process
- Data were generated assuming that 4, 8, 16, and 32 C-band wavelengths were available
- Steiner tree (ST) heuristic & Dijkstra's algorithm was used & the first-fit algorithm

Dynamic IA-UMC-RWA Algorithm



Dynamic IA-UMC-RWA Algorithm

- $x(j)$ and $y(j)$ were extracted from each lightpath/light-tree j attempted to be established into the network
- For each light-tree, a number of patterns was generated by decomposing each light-tree to its constituent lightpaths
- $y(j) = 0$ for a Q-factor below the predetermined Q-threshold, and $y(j) = 1$ otherwise

PATTERNS GENERATED FOR EACH WAVELENGTH CASE

	$U=4$	$U=8$	$U=16$	$U=32$
# patterns in $\mathcal{D}$	50,322	82,010	91,221	91,338
# patterns with $y(j)=1$	36,152	60,966	69,594	69,300
# patterns with $y(j)=0$	14,170	21,044	21,627	22,038

Model Accuracy

Accuracy Results for $U = 4$

# of patterns in $\mathcal{D}^r$	6000	12,000	24,000
# of patterns in $\mathcal{D}^t$	2000	2000	2000
Dropout fraction (p)	0.8	0.8	0.8
# of epochs	700	700	700
Training time in min	0.7	1.5	3
Class 1 Acc. (%)	90.8	87.6	89
Class 2 Acc. (%)	87.3	100	99.9
Total Acc. (%)	89	93.8	94.45

Accuracy Results for $U = 8$

# of patterns in $\mathcal{D}^r$	6000	12,000	24,000
# of patterns in $\mathcal{D}^t$	2000	2000	2000
Dropout fraction (p)	0.8	0.8	0.8
# of epochs	700	700	700
Training time in min	0.6	1.5	3
Class 1 Acc. (%)	88.5	85	90
Class 2 Acc. (%)	96	100	99
Total Acc. (%)	92.25	92.5	94.5

Accuracy Results for $U = 16$

# of patterns in $\mathcal{D}^r$	8000	18,000	36,000
# of patterns in $\mathcal{D}^t$	2000	2000	2000
Dropout fraction (p)	0.75	0.75	0.75
# of epochs	800	800	800
Training time in min	1	2.6	5.3
Class 1 Acc. (%)	74.4	87.5	93.3
Class 2 Acc. (%)	100	99.9	96.4
Total Acc. (%)	87.2	93.7	94.8

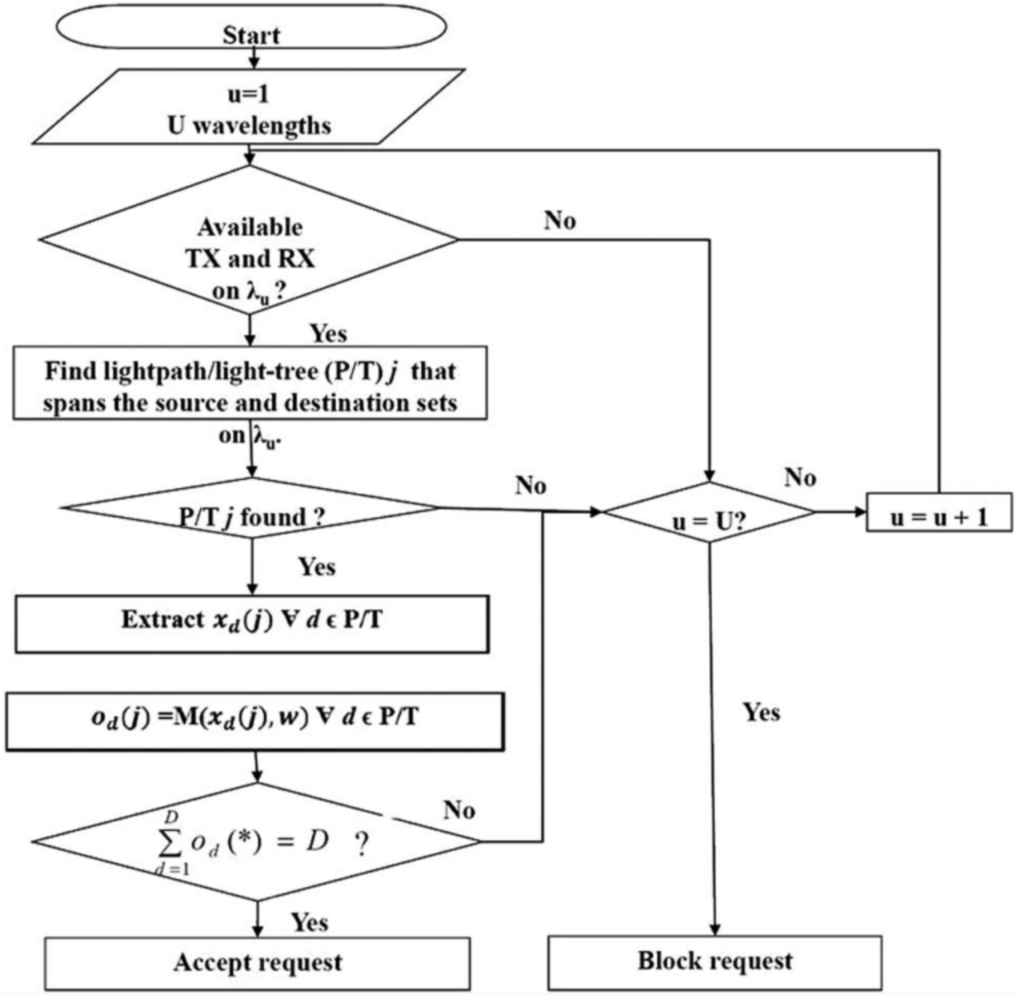
Accuracy Results for $U = 32$

# of patterns in $\mathcal{D}^r$	8000	18,000	36,000
# of patterns in $\mathcal{D}^t$	2000	2000	2000
Dropout fraction (p)	0.75	0.75	0.75
# of epochs	800	500	800
Training time in min	1	2	5.3
Class 1 Acc. (%)	91.7	87.7	93.7
Class 2 Acc. (%)	86.6	99.1	97.4
Total Acc. (%)	89.15	93.4	95.5

Performance Evaluation

- Case 0: The QoT constraint is not considered during the dynamic UMC-RWA algorithm.
- Case 1: The Q-factor model is utilized in the dynamic IA-UMC-RWA algorithm for the QoT decisions.
- Case 2: The data-driven QoT model is utilized in the dynamic IA-UMC-RWA algorithm for the QoT decisions

Performance Evaluation



$$\Sigma_{d=1}^{D} o_d(j) = D$$

$o_d(j)=M(\mathbf{x}_d(\mathbf{j}),\mathbf{W}),\forall d \in$ P/T, indicating that the QoT is sufficient for every destination node in the connection.

Performance Evaluation

BLOCKING PROBABILITY RESULTS FOR $U = 4$

	Case 0	Case 1	Case 2 (93.8)	Case 2 (94.45)
Overall $Pr\{\text{blocking}\}$	0.42	0.44	0.44	0.44
$Pr\{\text{blocking}\}$ due to QoT	–	0.15	0.25	0.21
$Pr\{\text{blocking}\}$ due to wav.	0.42	0.29	0.19	0.23

BLOCKING PROBABILITY RESULTS FOR $U = 8$

	Case 0	Case 1	Case 2 (92.5)	Case 2 (94.5)
Overall $Pr\{\text{blocking}\}$	0.09	0.11	0.15	0.12
$Pr\{\text{blocking}\}$ due to QoT	–	0.04	0.1	0.06
$Pr\{\text{blocking}\}$ due to wav.	0.09	0.07	0.05	0.06

BLOCKING PROBABILITY RESULTS FOR $U = 16$

	Case 0	Case 1	Case 2 (93.7)	Case 2 (94.8)
Overall $Pr\{\text{blocking}\}$	0	0.0002	0.0014	0.0004
$Pr\{\text{blocking}\}$ due to QoT	–	0.0002	0.0014	0.0004
$Pr\{\text{blocking}\}$ due to wav.	0	0	0	0

BLOCKING PROBABILITY RESULTS FOR $U = 32$

	Case 0	Case 1	Case 2 (93.4)	Case 2 (95.5)
Overall $Pr\{\text{blocking}\}$	0	0	0.0004	0.0004
$Pr\{\text{blocking}\}$ due to QoT	–	0	0.0004	0.0004
$Pr\{\text{blocking}\}$ due to wav.	0	0	0	0

Conclusion

- Employ machine learning in unicast and multicast provisioning with QoT constraint, which can also be applied in other areas, such as estimating the failure probability.
- Data set generation can be based on results of other traditional schemes, which is not difficult to achieve.
- The paper also consider the practical feasibility, although it is not enough and detailed.

Thanks!