Network and business modelling under traffic forecast uncertainty: a case study

Neil Geary¹, Andreas Antonopoulos², John Mitchell¹

¹Communications Engineering Doctorate Centre, University College London, London WC1E 7JE, UK
²Central European University Graduate School of Business, H-1051 Budapest, Nádor utca 21, Hungary

Contact e-mail: n.geary@ee.ucl.ac.uk

Abstract – Uncertainty is a problem for network planning since networks are dimensioned around traffic forecasts looking several years into the future. In this paper we investigate a methodology which can quantify the effects of traffic forecast uncertainty on the network design cost and also its robustness in terms of traffic carrying capability. Both traffic volume and distribution uncertainties are encompassed in the analyses for a case study network topology. Different sized networks are also investigated. The paper concludes by relating this network modelling work to a more general business modelling process under uncertainty.

Index Terms – network planning, traffic forecast, uncertainty, risk, recourse, business modelling.

I. INTRODUCTION

Telecom network operators are now part of a competitive environment and this brings extra pressures such as uncertainty of future demand and pricing. In the monopoly days, traffic forecasts were more reliable and network deployment plans could be based around them. Now it is clear that a more detailed planning process is required to rise above the ineffective art of overprovisioning. This is especially true in a post-boom telecoms investment environment where lowest cost network solutions must be sought.

Planning a network requires information as the inputs into the planning process, and often these are predictions of future parameters. Examples of the information vital for optical network planning include the level of expected demand (traffic volume), the cost of optical networking equipment, and regulatory implications such as interconnect agreements. All of these inputs to the planning process can suffer from uncertainty when predicting for the future.

In [1] an excellent overview relates how the quality of planning information (i.e. how accurate or certain it is) has declined in utility industries since deregulation. Under the previous monopolies, prices were stable, and demand could be accurately forecast. One of the conclusions of [1] is that when planning in an uncertain world, simulation and optimisation should be given more emphasis over the traditional reliance on forecasting methods.

The important uncertainties for optical core network planning are:

1. Price. The price elasticity of bandwidth demand and the impact of competition both affect the price of telecoms services.

2. Market, competitive, regulatory and macroeconomic risks. These risks increase the uncertainties in the business modelling process and can be considered external to the network planning process. However, the discount rate (or cost of capital) will vary according to these factors, so this is uncertainty directly relating to network investment planning and strategy.

3. Technology. Will there continue to be disruptive decreases in the per-bit cost of transmission and switching?

4. Demand. Related to price, the level of expected demand for telecoms services is the vital planning input for network planning. Uncertainty may result in networks being built that are either underprovisioned, meaning that network operators have to turn away customers, or expensively overprovisioned. The adoption rate of innovative new telecom services, for example mobile video calling, also has a high degree of uncertainty attached and adds to the uncertainty of bandwidth demand.

5. Traffic. The level of demand will determine the traffic volume required on the network. However, the distribution of traffic (the pattern between cities) is also subject to uncertainty, and may vary away from the forecast, particularly in a high customer churn environment.

These five planning inputs are all highly inter-related, for example regulation influences price. However, if one overriding uncertainty could be identified, it would be the traffic forecast, since all of the uncertainties identified above are related to it. Networks must be designed to be sufficiently robust to variations in traffic [2].

Traditional solutions to the problem involved overprovisioning to deal with uncertainty. This may lead to excessive capital investment and generally only deals with uncertainty in traffic volume. Typical approaches include adding an extra 10% of capacity to cope with uncertainty.

The problem with this approach is that it is a blind estimation of the level of uncertainty and its effects on the network design. It is also unscientific and may result in networks that are overly expensive. Considering that a survey in [3] reports that network planning is one of the largest risk
factors involved in operating a network, overprovisioning in this fashion no longer seems viable.

A finding from [4] is presented here as being highly relevant:

"A wise and pragmatic policy is to accept incoherence (in the traffic forecast), trying to quantify its effects on the planning results."

This implies a risk analysis approach as described in [5], where deterministic methods should be used to design and optimise networks around a forecast. However, it is vital to characterise the robustness to uncertainty of the design and quantify any recourse costs as a consequence of the uncertainty. Recourse costs are those extra investment costs necessary to carry all of the actual traffic. These in practice may vary from incurring SLA penalties, paying 'spot' prices for leasing extra temporary bandwidth, or an investment in extra equipment. Indeed the recourse cost may encompass all such elements over different timescales, and network operators are likely to choose the most cost-efficient option when faced with an inability to carry traffic.

For simplicity in this paper we assume the recourse costs to be the extra line system deployments and upgrades required to carry all actual traffic. This process is applied to circuit-switched core optical network designs.

In sections II and III the proposed methodology is introduced for analysing network designs under the presence of traffic forecast uncertainty. The details of the case study network traffic forecast is outlined in section IV. Section V presents the results, including a comparison between different sized networks with otherwise similar characteristics. Finally, section VI relates the results of this network modelling to a more general business modelling scenario under uncertainty.

II. CHARACTERISING TRAFFIC FORECAST UNCERTAINTY

In order to characterise traffic distribution uncertainty, we use the metric proposed in [6], termed the Distribution Forecast Accuracy (DFA). It is based on the statistical correlation of two traffic matrices: a forecast matrix P and an 'Actual' matrix A. A perfect match with the forecast is indicated by DFA=1, implying no traffic forecast uncertainty.

An advantage of the DFA metric is that it can be employed when the P and A matrices contain different volumes of total network traffic. A drawback of the DFA is that it does not capture certain extreme traffic variations well, as highlighted in [7].

In this paper we aim to assess the robustness of network designs to simultaneous traffic volume and distribution uncertainty, so the DFA metric is used. We extend the work in [6] in order to quantify the recourse costs, in addition to assessing the proportion of traffic that cannot be routed due to uncertainty as described in [6].

III. ANALYSING NETWORK DESIGN ROBUSTNESS TO TRAFFIC VOLUME AND DISTRIBUTION UNCERTAINTY

A. Overview

A restricted Monte-Carlo analysis is used to assess network designs to the presence of traffic uncertainty. Random variations are applied to the predicted traffic matrix P, and this traffic, an instance of A, is then routed near-optimally over the capacitated network topology. If not all of the traffic could be routed, the required optical line systems will be upgraded until all of the traffic can be carried. This extra upgrade cost is termed the recourse cost, and it represents the extra network infrastructure investment required to cope with the traffic uncertainty.

A Monte-Carlo analysis over the full range of uncertainty space would be computationally unfeasible [8]. Therefore the randomly modified A matrix is only accepted if it falls within a specified range of traffic prediction accuracy, e.g. a DFA of between 0.5 and 0.6. This means that a full Monte-Carlo analysis is not needed, since the level of uncertainty can be restricted to a certain band.

Since the DFA correlation metric is independent of the total traffic volume, network designs can be assessed for a combination of traffic volume and traffic distribution uncertainty. A traffic generator module is used to determine traffic at the required volume levels whilst maintaining a similar distribution to the original matrix P. The random variations can then be applied to generate the A matrices in the DFA range required.

B. Assumptions

An opaque optical mesh architecture is assumed. Each traffic demand is assumed to be a bi-directional wavelength. The networks are designed to carry the predicted traffic P using a combination of 40, 80 and 160-wavelength optical line systems. Following the rule of thumb that as capacity quadruples, cost is multiplied by 2.5; the line system costs are taken as 1/1.5/2.5 arbitrary units, for the 40/80/160 channel systems respectively.

C. Near-optimal routing

Each traffic matrix A is routed over the capacitated network topology designed for the P traffic. This is a Multi-Commodity Flow problem, and can be solved by Integer Linear programming techniques [4][9][10]. However this is unfeasible for large networks, so a Simulated Annealing (SA) heuristic is used here to route the traffic on the network. In order to decrease the computation time, the SA initial position is determined by pre-sorting the demands by a metric that reflects the shortest-path hop distance and demand volume [10]. Individual routing operations are performed using the Dijkstra algorithm for unprotected demands, and Suurballe’s algorithm for protected traffic which requires a simultaneous calculation of two disjoint paths [11].
The SA 'temperature', indicating the locus of the immediate search neighbourhood, is proportional to the total volume of demands which cannot be at that point routed.

Each iteration of the annealing process de-allocates a number of routed demands according to the temperature, and then re-allocates all demands not yet routed in a random order. The objective function to minimise, or the threshold on which an iteration is accepted, is defined as:

\[ \text{No. of WLs not routed} + \text{Network Utilisation} \]

(where Network Utilisation is a value between 0 and 1 representing the network-wide line system capacity used by the solution)

Therefore the annealing aims to route all the traffic whilst providing the lowest network utilisation figure. The annealing stops when either all demands have been routed, or when 1500 iterations of the annealing have been completed. As an indication, a converged result is usually found after a few hundred iterations.

Routing the traffic near-optimally in this way assumes the use of reconfigurable network elements such as Optical Cross Connects at all nodes. This is because in reality demands will arrive in a discrete sequence at random, and a full dynamic traffic analysis would be required should there exist the constraint that traffic must stay on its initially chosen route. In order to remove the dependency on the ordering of the demand arrival, the capacitated routing algorithm described here tries to route as much traffic as possible over the installed network capacity. In an operations sense, this means that demands might be reconfigured to achieve this, implying the presence of optical switches.

D. Incremental build-out to determine recourse cost

If the Simulated Annealing algorithm converges to a situation where there is still some traffic that is unable to be routed, a further incremental network design step is performed. The remaining demands are routed on their shortest-path routes. Following a First-Fit wavelength allocation, it may be necessary to upgrade certain line systems or install new line systems.

The total cost of the upgrades and new deployments needed to carry the 'unroutable' traffic is the recourse cost for that particular traffic matrix \( A \). The upgrade costs are taken as the difference between the line system costs indicated in section III.B above. For example, if the only upgrade required is one line system needing 80 channels capacity instead of the predicted 40, the recourse cost will be \( 1.58 - 1 = 0.58 \).

If all of the traffic could be routed in the Simulated Annealing phase, then a recourse cost of zero is taken since no further network upgrades are required.

E. Results validation

In a Monte-Carlo analysis, it is important to analyse enough randomly-chosen \( A \) matrices in order to determine the mean recourse cost for each combination of traffic volume and traffic distribution uncertainty. The process is repeated using different \( A \) traffic matrices until a mean value with a specified confidence level can be calculated for all levels of traffic volume and distribution uncertainty.

The advantages of using this process to assess the robustness of network designs to the presence of traffic uncertainty are:

- It can be used to assess any optical network design that has been based on a forecast, whether it was dimensioned manually or using design software. Different candidate network designs, topologies and architectures can be appraised for their robustness.
- The level of information required relating to the traffic forecast is minimal; only a 'best-guess' traffic forecast and a network-wide confidence metric is required. The confidence metric is represented by the DFA range chosen for analysis.

IV. CASE STUDY SPECIFICS

A 30-node network, with an average node degree of 2.8, is dimensioned to carry the 'predicted' traffic matrix \( P \), which comprises 600 wavelengths of traffic in the following mix of idealised traffic patterns:

- 50% uniformly distributed traffic
- 25% node to adjacent node traffic (between two nodes that share a link)
- 25% hubbed traffic (sourced from the topologically central node, representing a capital city)

This mix of three different idealised traffic patterns gives a realistic approximation to real network traffic distributions in optical transport networks. 50% of the traffic is chosen to require optical layer 1+1 disjoint-path protection. These demands were chosen at random.

The network is dimensioned to carry the predicted traffic \( P \) using a combination of 40, 80 and 160-wavelength optical line systems. Demands are routed on their shortest paths. If more than 160 channels are needed on any link, line systems are stacked. Figure 1 shows the line system deployments chosen and individual link loads.

The total network line system cost can then be calculated, for the case study network it is 70.42 arbitrary cost units. The transponder cost is not included in this analysis.

A linear sweep of traffic volume uncertainty was used, at various intervals from -50% to +50% relative to the predicted traffic volume of 600 wavelengths.

At the same time, bands of DFA were assessed to model the effect of traffic distribution uncertainty, with the following ranges:

- Low level of distribution uncertainty: \( 0.8 < \text{DFA} < 0.9 \)
- Mid level of distribution uncertainty: \( 0.6 < \text{DFA} < 0.7 \)
- High level of distribution uncertainty: \( 0.4 < \text{DFA} < 0.5 \)
V. RESULTS

Two measures of robustness are assessed and quantified in this paper. Firstly, the amount of actual (A) matrix traffic that cannot be routed, due to capacity being installed in the wrong part of the network. This will be measured as the complement, i.e. the proportion of total traffic in A that was able to be successfully routed. This is termed the 'servability', as introduced in [7]. A servability of 100% indicates that the network could carry all demands in the A matrix.

The second measure is the recourse cost, the extra investment cost required in order to carry the 'unrouteable' traffic. If the servability is 100%, the recourse costs will be zero. The recourse cost is expressed as a percentage of the original (forecast) network design cost of 70.42.

In this example, simulation of different A traffic matrices continues until a 95% confidence level on the mean recourse cost is within 3%; and for the servability, simulation continues until it is within 1%. These confidence levels are shown as error bars in the charts that follow.

Figure 2 shows the mean servability rolloff as the ‘actual’ traffic volume increases. The gradient of the rolloff is approximately 0.075% per wavelength. This is true for all DFA bands shown since the traces become parallel above 700 wavelengths actual traffic.

For the highest DFA band, the network design has the ability to absorb more traffic beyond the 600 originally predicted. This is shown in Figure 2 by the point at 650 wavelengths where the servability remains almost 100%.

This is due to the ‘capacity overhead margin’ present in the network, as it is only 81.9% utilised when carrying the P traffic.

For the lower DFA bands, the knee on the servability curve is at a lower volume. This is because the capacity installed in the network is in the wrong place for the ‘actual’ traffic, which has a significantly different distribution. With the lowest DFA band, 2% of traffic is not routable even when the traffic volume is spot on at 600 wavelengths (the servability is 98%).

Figure 3 shows the mean relative recourse cost required to carry all of the presented ‘actual’ traffic. A recourse of 0% means that no extra network investment is required.

There are high recourse levels for the low DFA band, even when traffic volume is on target or lower. With an ‘actual’ traffic volume of 550, which is lower than predicted, there is still a mean recourse requirement of 5%. This is due to line systems being installed in the wrong part of the network for the presented ‘actual’ traffic, and so extra line systems are required to compensate even though the volume is below that forecasted.

The traces converge at high actual traffic volumes, showing that the distribution accuracy is less important when the traffic volume is severely underestimated. It can be seen that when the traffic volume is underestimated by at
least 25%, the level of traffic distribution uncertainty is irrelevant. But when the 'actual' traffic volume is near the prediction or even overestimated, the level of traffic distribution uncertainty is critical and determines the level of expected recourse costs.

It is interesting to note that the 95% confidence level on the mean, shown in these charts as error bars, is larger as the DFA range decreases (the error bars get longer). This is consistent with the findings reported in [7], where the 5th and 95th percentile traces diverge from the mean as the distribution accuracy decreases.

This seems intuitively correct since one would imagine there to be many more ways for a distribution to be different from the P traffic matrix the further away it is in accuracy from the prediction. This larger space or spread manifests itself in these stochastic analyses as a wider variation around the mean.

A. Effect of network size

In this section, the results from two other networks also having an average node degree of 2.8 are assessed. One network has 15 nodes (network M2.8), the other 45 (network B2.8). The aim is to repeat the analysis keeping all other parameters the same.

In order to compare between the three networks fairly, a predicted traffic volume of 20N wavelengths is used, where N = the number of nodes. The networks are dimensioned in a similar way around their respective traffic forecasts in the same way as described for the case study network, and the resulting network designs all had a network utilisation figure between 80.5 and 82.5%.

The x-axis of the charts becomes a relative traffic volume measure. For example, the data point at 750 wavelengths in Figure 2 becomes a point at a relative traffic volume of 125% since it involves 25% more traffic than the 30-node network was designed for.

Although not presented here for brevity, all three networks show a similar servability trace to that of Figure 2. However, in all DFA bands, the rolloff occurs earliest for the B2.8 network with 45 nodes. Furthermore, the M2.8 curve has a knee at a later volume than L2.8. The rolloffs occur in order of decreasing network size.

In terms of the recourse cost, this trend against network size is repeated. Figure 4 shows the recourse cost for the lowest DFA band, relative to each network's original forecast design cost. Significant relative recourse levels are seen for the B2.8 network at relative traffic volumes of less than 100%. The trace is almost linear for the 45-node network as relative volume increases, whereas the other networks have a more S-shaped nature.

In this part of the study all other variables are equal: the average node degrees are all 2.8, and the utilisation of the network designs under the P traffic matrix are within the tight range of 80.5-82.5%. Therefore it can be concluded that larger networks are less robust to traffic volume and distribution uncertainty.

VI. NETWORK MODELLING TO BUSINESS MODELLING

The results obtained for the case study network (L2.8) can be applied to a business modelling process. The goal is to determine a distribution for the likely network Capital
Expenditure (here simplified to just line system costs).

We assume a 'fuzzy' estimation of the expected level of uncertainty, based on some simplifying assumptions about the distributions on the important uncertainty metrics and the probabilities of these states occurring. Then stratified sampling is applied based on the results from section V.

The estimated traffic volume probability distribution is shown in Figure 5. It has been approximated to a discrete distribution at the volume intervals used in section V. It is a triangular distribution centred around the predicted volume of 600 wavelengths — with a slight skew favouring the underestimate side, since historically this has been likely when forecasting demand for newer technologies [12]. The use of a triangular distribution accentuates the role of the predicted traffic volume peak at 600. Other distributions could also be used; e.g. normal or uniform, depending on how accurate the traffic volume prediction was felt to be.

![Figure 5 - Estimated traffic volume PDF.](image)

For the distribution (traffic pattern) uncertainty, we use in this example an estimation of a 20% chance of a high DFA, and a 40% chance for both the medium and lower bands of DFA (as defined in section IV). This is shown graphically in Figure 6, and might be sensible for a long term planning horizon of 3-5 years, since it is unlikely that the traffic distribution will be accurate looking that far ahead.

Using the data gathered on the L2.8 network from section V, it is possible to determine a probability distribution for the total Capital Expenditure (CAPEX) requirement based on random choices weighted by the two distributions in Figures 5 and 6. A recourse cost of zero implies a total CAPEX cost of 100%, whereas a recourse cost of 50% means a total CAPEX of 150%. Figure 7 shows the final CAPEX distribution as a histogram.

The mean (or expected value) of this distribution is 119.3%. An expected 19.3% increase in required CAPEX should be of concern to network operators, as this will negatively impact profits. However, the median is 113.3%, and the mode is 100% (zero recourse). It is therefore a difficult histogram from which to extract generalities. Yet for the purposes of business modelling this distribution itself can be converted into real dollar figures and then used in a Monte-Carlo business modelling analysis. An example process is shown in Figure 8.

The telecoms business modelling process described in [13] uses a Monte-Carlo analysis with many input variables subject to uncertainty. One can also specify some of these input variables to be correlated with each other, for example expenses and capital expenditure. Figure 8 includes three input variables taken from this reference — distributions for price, volume and expenses. The CAPEX distribution determined through network uncertainty modelling (Figure 7) is simply a further input distribution for this process.

With this extra information on the CAPEX requirement, the business modelling process can be carried out as described in [13] to determine the expected distribution of profit or the project NPV. The spread on the calculated rate of return on investment will quantify the riskiness of the network investment project [14]. This can then be used in the Capital Asset Pricing Model to determine the required investor rate of return (or cost of capital) for such a network investment project [14].

**VII. CONCLUSIONS**

A statistical method to determine the cost of network designs under the presence of traffic volume and distribution uncertainty is presented. Results from the case study networks suggest that, with the analysed traffic pattern, larger networks are less robust to traffic forecast uncertainty, all other things being equal.

The process of network modelling under traffic uncertainty provides more reliable input information for a business modelling process. It can also help to quantify the risk of a network investment project.
ACKNOWLEDGEMENTS

This work is funded by Lucent Technologies UK, and the UK research council EPSRC. Thanks also to the reviewers who helped improve the quality of this paper.

REFERENCES


Figure 8 - The results of detailed network modelling under traffic uncertainty can be linked into a business modelling uncertainty analysis, with the goal of finding the expected distribution of profit. The spread of the profit measure distribution is useful for determining the riskiness of the investment project.