Network Design Using Hierarchical Performance Models and Multi-Criteria Optimization

Mingyan Liu
John S. Baras
Center for Satellite and Hybrid Communication Networks
University of Maryland
College Park, MD 20742, USA *

ABSTRACT

Network design and dimensioning often involve trade-offs between different objectives that sometimes may have opposing effects on design. At the same time, some objectives are mandatory while others are optional. In this paper we use the multi-objective optimization package CONSOL-OPTCAD for the design of trunk reservation parameters. We also study link capacity design by using CFSQP along with our model and its sensitivities generated by Automatic Differentiation (AD). Using AD for sensitivity analysis of our reduced load approximation is a novel approach. Using AD for computing derivatives of the fixed point along with the calculation of the fixed point itself is both accurate and efficient. Putting together these techniques we present a systematic way of network design, regardless of the details of the underlying performance evaluation model.

INTRODUCTION

Network design relates to determining the size, connectivity, configuration and capacity/resource allocation of a network. Network dimensioning on the other hand is a problem of figuring out the number of users a network can support with a certain level of QoS. These problems are closely related to network performance evaluation and studies in bother areas are often carried out side by side, as can be shown by the references in the previous sections. While performance evaluation can be used for analyzing existing networks, our goal here is to use performance evaluation for network optimization, design and dimensioning.

Typically, a design and dimensioning model is formulated into a constraint optimization problem. The objective function can be an overall network revenue in terms of network deployment cost, reward from traffic carried through and penalty on traffic lost. The constraints are often QoS requirements for different traffic classes, whose calculation is determined by the performance evaluation model that may count for multi-rate traffic, adaptive routing or call admission control policy. Generally the performance evaluation is done iteratively to guide the search algorithm and check the feasibility of QoS constraints.

Works on network dimensioning for dynamic routing networks for the single-service case include [1, 2, 3, 4, 5, 6]. Multi-service networks have been addressed in [7, 8, 9, 10].

Due to the iterative nature of most performance models themselves, many of such works try to derive a closed form presentation of the shadow price or implied cost, often involving gradient calculation, to make the optimization model easier to program and compute numerically [8, 9, 11, 12]. Such derivations may vary depending on the performance model used by the optimization. In many cases obtaining the exact form of gradients is not possible and further approximation is required. In general manually deriving gradients makes the design process highly dependent on the underlying performance model.

In [13, 14] Medhi proposed a network dimensioning model that is decomposed into two steps: a bandwidth estimation problem and a multi-commodity flow model or a routing and capacity design problem.

Our emphasis is on integrated services data networks employing QoS routing. While research results are abundant for circuit-switched scenarios, there has not been much study that applies to data networks. There has been extensive research in technologies of design and optimization of traditional circuit switched networks, e.g., [1, 11, 2, 15], and plenty of research results

---

*Prepared through collaborative participation in the Advanced Telecommunications and Information Distribution Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0002. The U.S. Government is authorized to reproduce and distribute reprints for Government purposed notwithstanding any copyright notation thereon.

for fully connected, symmetric networks with fixed, sequential or state-dependent routing [12], especially for networks with no more than two hops on their routes [16], or when network traffic is of single rate [17]. There has been relatively lesser study that considers large random networks with both multiple traffic rates and state-dependent routing [12, 18, 19, 10]. Furthermore, all of such methods are for flat networks and flat routing schemes.

We observe that the evolving integrated service networks typically support traffic with varying bandwidth requirements. They are typically much sparser and have a more hierarchical topology. Correspondingly, routing schemes are becoming increasingly hierarchical in order to scale up with the size of the network. Routes can comprise of a much larger number of hops and there are typically a large number of possible routes between source and destination nodes. Therefore it is important to develop network performance models that take into account the random topology, different bandwidth requirements and hierarchical routing schemes.

There is also a lack of comprehensive tools and systematic methods to use these performance models for network design, planning, control and management.

LINKING PERFORMANCE EVALUATION WITH MULTICRITERIA OPTIMIZATION

Our approach for network design, optimization and dimensioning is to use existing numerical tools. Our reason for this is two-fold: One, an analytical approximation algorithm can be easily linked to mathematical programming tools to get network performance optimization and trade-off analysis. Secondly, we aim at developing a systematic way of network design and optimization, which is relatively independent of the underlying detailed performance model. We believe that by using properly selected mathematical programming tools we can achieve this.

There are potentially many numerical tools that we could use. For the examples we present in this paper we used CONSOL-OPTCAD and CFSQP. Our choice of these two for this research is based on the following. CONSOL-OPTCAD and CFSQP are both developed at the University of Maryland by groups led by Professor André L. Tits and are freely available. We had considered using OPL (Optimization Programming Language) and its industrial implementation the OPL Studio from ILOG, Inc.. Unfortunately OPL is primarily targeted at linear programming (e.g., CPLEX), integer programming and combinatorial optimization problems, which made it a less favorable choice for our purposes since our model is nonlinear. Same considerations apply to languages like AMPL and GAMS. On the other hand, CONSOL-OPTCAD is a tool for optimization-based design of a large class of systems, which include both linear and nonlinear problems. More importantly, it evaluates the performance of instances of the system under consideration and allows parameters to take on any real value in a given range so that they can be optimally adjusted, and is thus very good for trade-off analysis. CFSQP (C code for Feasible Sequential Quadratic Programming) is a set of C functions for solving large scale constrained nonlinear minimax optimization problems, generating iterates satisfying all inequality constraints. It takes user provided objective and constraint function code. It provides default finite differencing for gradient evaluation of these functions but can also take user provided gradient evaluation code for objective and constraint functions. This last feature makes it a very nice package to be used along with ADIC which generates gradient evaluation code as we described in the last chapter (ADIC is also freely available from Argone National Lab).

In the next section we use the reduced load approximation method with CONSOL-OPTCAD [20] for trunk reservation parameter design. In the second application example we take the reduced load model along with its derivative code generated by ADIC and use CFSQP [21] for link capacity design.

DESIGN OF TRUNK RESERVATION PARAMETERS

As a way of call admission control, trunk reservation regulates individual classes of traffic as well as their inter-relationship. The combination of trunk reservation parameters $\{r_1, r_2, ..., r_S\}$ for each class of traffic s could potentially affect average blocking probability and the total carried traffic by the network. In this example we choose the objective to be the weighted average of blocking probabilities, and the constraints to be the bounds on the blocking probability of individual classes of traffic.

Following our standard notation, this design problem is formulated as follows:

Design parameters: $r_1, r_2, ..., r_S$.

$$\min \quad \frac{\sum_{(r,s)} \lambda_{rs} \cdot B_{rs}}{\sum_{(r,s)} \lambda_{rs}}$$

$$s.t. \quad B_{rs} < \text{bound}_{rs}, \quad \text{all} \quad (r, s)$$

where λ_{rs} is the offered traffic of class s on route r.

In CONSOL-OPTCAD, we can provide two values, namely the good value and bad value, for each $bound_{\mathsf{k,l}}(s)$, indicating our level of satisfaction. Trade-off analysis can be carried out between minimizing the objective function value and satisfying the constraints, and thus decide the combination of $\{r_1, r_2, ..., r_{\mathsf{S}}\}$.

By applying this formula to the following example of a fully connected network shown in Figure 1 below, in the case with trunk reservation admission control, and restricting the trunk reservation parameters to be less or equal to 5, we get the trade-off between the blocking probability of each class vs. the weighted average blocking probability, which is shown in Table 1 (Numbers in italic are individual minimum).

Table 1: Trunk Reservation Parameter Design

Weighted Average	B_1	B_2	B_3	(r_1, r_2, r_3)
0.265702	*0.036159*	0.580163	0.803230	(1,4,5)
0.273409	0.062937	0.526852	0.840663	(1,3,5)
0.290184	0.067968	0.562512	0.792359	(2,4,5)
0.292510	0.112452	0.467300	0.861320	(1,2,5)
0.307262	0.116759	0.485502	0.819291	(1,2,4)
0.309973	0.119625	0.492353	0.824259	(2,3,5)
0.420363	0.571205	*0.284030*	0.565232	(4,1,2)
0.406413	0.319234	0.671934	*0.322844*	(3,5,1)

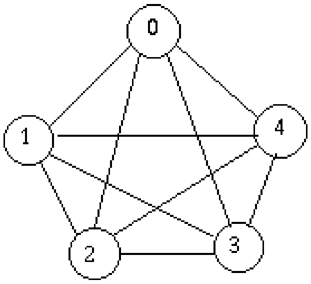

Figure 1: Topology of Example Network.

Capacity for each link is set to be 100. There are three classes of connections, which are indexed 1, 2, and 3. They have bandwidth requirements of 1, 2, and 3, respectively. When call admission control is used under heavy traffic, the trunk reservation parameter for each class is 2, 4, and 6, respectively. For each node pair, the direct route and all two-hop routes are allowed. The direct route is listed first in the routing list, and the two-hop routes are listed in random order. The medium traffic rates are listed in Appendix A. Heavy traffic rates are set to be double the medium rates.

Because of the topology symmetry of this example, the same class of traffic encounters approximately the same blocking probability regardless of their source-destination node pair and input rate, although their input rates are counted in calculating the weighted average blocking probability. So the numbers displayed here are only distinguished by their classes but not their associated source-destination node pairs.

As we can see from Table 1, the weighted average blocking probability and the blocking probability of class-1 type of traffic achieve their optimum at the same time with trunk reservation parameter choice of 1,4 and 5. The reason is obvious: since class-1 has much smaller trunk reservation requirement than class-2 and 3, together with its lowest bandwidth requirement, it has the highest priority and chances of being admitted into the network. On the other hand, class-2 and 3 are being jeopardized by their high trunk reservation requirement and also high bandwidth requirement. We may also conclude that class-1 type of traffic occupies the greatest amount of the total traffic throughput since the weighted average blocking probability is the lowest while blocking probability of class-1 is the lowest. Class-2 and 3 types of traffic achieve their minimum separately when their trunk reservation parameters are 1 and others' are higher.

DESIGN OF LINK CAPACITIES

Similar to the first application, link capacities are another set of parameters which is critical in network design. It is natural to expect that we should assign higher capacities to more frequently congested links, and save capacities on rarely used ones. In this example we define the objective as the total carried traffic to be maximized, and the constraints as the total available resource/bandwidth and the capacity bounds on each link. Note that this objective is equivalent to that of the first example, but gives a different objective value in terms of throughput rather than blocking probability. We use the hierarchical network presented in Figures 2 and 3 with the offered traffic given in Table A.2.

There are three clusters in this example, with the dash-circles surrounding each one. Each group has a label/address, e.g., 1.1 indicates Layer 1, Peer Group 1. Each node has an address as well, e.g., 1.1.3 is Node 3 of Peer Group 1.1. All border nodes are shown in black and non-border nodes are shown in white. A cluster can have a single or multiple border nodes. A border node

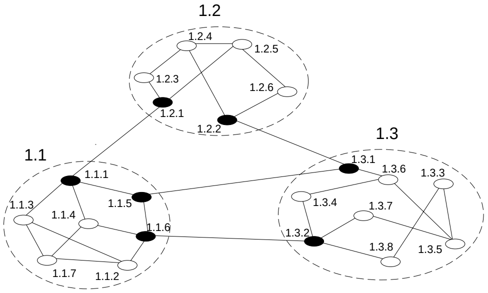

Figure 2: Network with three clusters – Layer One

can be connected to different clusters, e.g., Node 1.3.1. A non-border node does not necessarily have a direct link connected to border nodes, although this is often true with IP networks. Note that all links on this layer are actual, physical links.

The way aggregation and abstraction are done is as follows:

- All border nodes are kept in the higher layer – in this case Layer 2;
- Border nodes belonging to the same cluster are fully connected via "logical links".

This results in the Layer 2 abstraction shown in Figure 3.

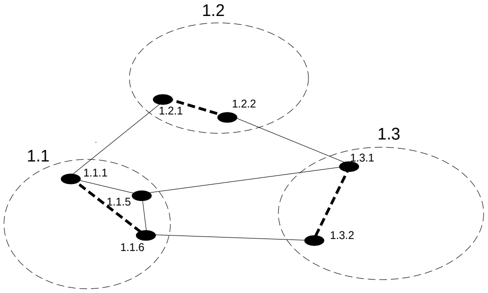

Figure 3: Network with three clusters – Layer Two

A logical link may correspond to an actual link on the lower layer, e.g., Link 1.1.1 $\longleftrightarrow$ 1.1.5. The real logical links are shown in dashed lines, indicating a feasible path rather than a direct link.

This is a 21-node, 30-link, 3-clusters, 2-layer network model. We have used both fixed hierarchical routing as well as dynamic hierarchical routing in our experiments. We use a single class of traffic requiring unit bandwidth. Link capacities vary from 80 to 160, which are listed in Table 2. We also provide the offered traffic load between node pairs at a "nominal" level in Appendix A. The intensity of this traffic load is shown in Table A.2, in which *load* is defined as the ratio between the total rate of traffic coming out of a node and the total out-going link bandwidth connecting to this node. At the nominal level, the value of this ratio for each node is around 0.05. In addition to this offered traffic load we also define a "weight" in our experiment as a multiplier to the nominal traffic, so that we get twice, three times of the nominal traffic, etc..

Design parameters: $\{C_j\}$, all j

$$\begin{aligned} &\max && \sum_{(r,s)} \lambda_{rs}(1-B_{rs}) \quad \text{or} \\ &\min && \sum_{(r,s)} \lambda_{rs}B_{rs} \\ &s.t. && \sum_{j} C_j \le 3700; \\ &0 \le C_j \le 1000 && \text{for all } j. \end{aligned}$$

We use CFSQP to solve this problem. The objective function we provide is the reduced load model, and the gradient function for the objective is generated by ADIC with necessary modification. We compare the results as follows. Table 2 is a random allocation of link capacities similar to what we used for numerical experiments. Tables 3 through Table 8 are optimization results on link capacities with the traffic load weight being 3, 5 and 7 respectively, and comparison of total carried traffic between optimal and non-optimal capacity allocations. Original results are in floating point type and we present rounded-off numbers here. Entries in Italics indicate cross-group links.

Table 2: Link Capacities Used for Comparison with Optimal Allocation

Link (i,j)	$C_{i,j}$	Link (i,j)	$C_{i,j}$	Link (i,j)	$C_{i,j}$
(0,1)	140	(1,2)	120	(1,3)	120
(0,2)	120	(2,4)	100	(3,4)	100
(3,6)	100	(0,6)	100	(4,5)	100
(5,6)	120	(6,7)	120	(5,13)	100
(4,20)	120	(7,8)	120	(7,10)	120
(9,12)	120	(11,12)	140	(8,9)	140
(9,10)	100	(10,11)	140	(12,13)	160
(13,14)	100	(14,15)	140	(14,18)	120
(16,19)	100	(16,18)	160	(15,20)	160
(17,20)	160	(17,18)	120	(19,20)	140

CONCLUSIONS AND FUTURE WORK

In this paper we showed how typical network design and dimensioning problems can be formulated as

Table 3: Optimized Link Capacities with Weight = 3.

Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$
(0,1)	0	(1,2)	111	(1,3)	111
(0,2)	111	(2,4)	138	(3,4)	103
(3,6)	165	(0,6)	111	(4,5)	111
(5,6)	155	(6,7)	211	(5,13)	152
(4,20)	161	(7,8)	111	(7,10)	111
(9,12)	111	(11,12)	131	(8,9)	131
(9,10)	91	(10,11)	131	(12,13)	172
(13,14)	186	(14,15)	131	(14,18)	111
(16,19)	71	(16,18)	71	(15,20)	131
(17,20)	131	(17,18)	111	(19,20)	111

Table 4: Throughput Comparison with Weight = 3.

	Optimal	Non-optimal
Total Carried Traffic	534.3	501.49767
Total Capacities Allocated	3683	3700

Table 5: Optimized Link Capacities with Weight = 5.

Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$
(0,1)	0	(1,2)	67	(1,3)	92
(0,2)	101	(2,4)	184	(3,4)	83
(3,6)	179	(0,6)	106	(4,5)	148
(5,6)	200	(6,7)	279	(5,13)	207
(4,20)	220	(7,8)	101	(7,10)	132
(9,12)	107	(11,12)	90	(8,9)	88
(9,10)	47	(10,11)	86	(12,13)	235
(13,14)	218	(14,15)	122	(14,18)	103
(16,19)	29	(16,18)	38	(15,20)	95
(17,20)	127	(17,18)	102	(19,20)	103

Table 6: Throughput Comparison with Weight = 5.

	Optimal	Non-optimal
Total Carried Traffic	889.48874	633.50409
Total Capacities Allocated	3689	3700

Table 7: Optimized Link Capacities with Weight = 7.

Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$	Link (i,j)	$C_{\mathrm{i,j}}$
(0,1)	0	(1,2)	75	(1,3)	106
(0,2)	81	(2,4)	153	(3,4)	72
(3,6)	193	(0,6)	94	(4,5)	110
(5,6)	193	*(6,7)*	*302*	*(5,13)*	*191*
(4,20)	*245*	(7,8)	118	(7,10)	149
(9,12)	108	(11,12)	96	(8,9)	97
(9,10)	55	(10,11)	95	*(12,13)*	*244*
(13,14)	230	(14,15)	114	(14,18)	108
(16,19)	35	(16,18)	36	(15,20)	99
(17,20)	116	(17,18)	92	(19,20)	92

Table 8: Throughput Comparison with Weight = 7.

	Optimal	Non-optimal
Total Carried Traffic	1074.2879	728.947842
Total Capacities Allocated	3699	3700

a multi-objective constrained optimization problem, using performance estimation network models, and we described our research results in the development of a general network design and dimensioning methodology by linking our performance models with Automatic Differentiation and multi-objective optimization algorithms and tools. We examined the applicability of automatic differentiation under the scenario of our interest. We presented examples and applications that demonstrate the speed and versatility of our methodology and algorithms.

Future work remains in the following areas. We need further numerical experiments for both the non-hierarchical and hierarchical model for validation purposes. Based on the performance and design model we developed, we are interested in extensive trade-off studies, as well as the property of the problem, e.g., convexity, feasible regions, etc.. Trade-off study helps in answering many realistic questions. For example, for a single connection, what is the trade-off between delay and bandwidth requirement? Is there a region where both metrics can be improved, or reduction in one inevitably results in increase in the other? Does increase in the trunk reservation parameter for a certain class of traffic benefit the overall network performance in terms of throughput, and within what range? Given existing network size and configuration, as well as the number of subscribers, what is the range within which it is profitable to increase network size or increase subscription or both? We plan to address these problems in future studies. It's also important to investigate the trade-offs involving networks' ability to handle peak demand or worst case scenario.

REFERENCES

[1] André Girard. *Routing and Dimensioning in Circuit-Switched Networks.* Addison-Wesley, 1st edition, 1990.

[2] A. Girard and B. Liau. Dimensioning of Adaptively Routed Networks. *IEEE Trans. Networking*, 1(4):460–468, 1993.

[3] A. Girard and R. Pagé. Dimensioning of Telephone Netwroks with Nonhierarchical Routing and Trunk Reservation. *Proc. of Networks'86*, pages 85–93, 1986.

[4] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Netwroks. *IEEE J. on Selected Areas in Comm.*, 9:968–981, 1991.

[5] F. P. Kelly. Routing and capacity allocation in networks with trunk reservation. *Mathematics of Operations Research*, 15:771–793, 1990.

[6] E. Rosenberg. A Nonlinear Programming Heuristic for Computing Optimal Link Capacities in a Multihour Alternate Routing Communications Network. *Operations Research*, 35:354–367, 1987.

[7] D. Medhi. Network Dimensioning for a Multi-Service Integrated Digital Network with Dynamic Routing. *Proc. of ACM/IEEE-CS Symposium on Applied Computing*, pages 33–36, 1991.

[8] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan. ATM Network Design and Optimization: A Multirate Loss Network Framework. *IEEE Trans. Networking*, 4(4):531–543, 1996.

[9] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan. Virtual Private Networks: Joint Resource Allocation and Routing Design. *Proc. INFOCOM*, 1(1):480–490, 1999.

[10] M. Liu and J. S. Baras. Fixed point apporoximation for multirate multihop loss networks with adaptive routing. Technical report, Institute for Systems Research, University of Maryland, College Park, MD, 1999. Technical Report TR99-44.

[11] F. P. Kelly. Routing In Circuit-Switched Networks: Optimization, Shadow Prices and Decentralization. *Adv. Appl. Prob.*, 20:112–144, 1988.

[12] F. P. Kelly. Loss Networks. *The Annals of Applied Probability*, 1(3):319–378, 1991.

[13] D. Medhi. Multi-Hour, Multi-Traffic Class Network Design for Virtual Path-based Dynamically Reconfigurable Wide-Area ATM Netwroks. *IEEE/ACM Trans. Networking*, 3(6):809–818, 1995.

[14] D. Medhi and S. Guptan. Network Dimensioning and Performance of Multi-Service, Multi-Rate Loss Networks with Dynamic Routing. *IEEE/ACM Trans. Networking*, 5(6):944–957, 1997.

[15] K. R. Krishnan and T. J. Ott. State-dependant routing for telephone traffic: theory and results. In *Proceedings of 25th Conference on Decision and Control*, volume 10a, December 1986. Athens, Greece.

[16] D. Mitra, R. Gibbens, and B. D. Huang. State-Dependent Routing on Symmetric Loss Networks with Trunk Reservations – I. *IEEE Trans. Communications*, 41(2):400–411, 1993.

[17] S. Chung, A. Kashper, and K. W. Ross. Computing Approximate Blocking Probabilities for Large Loss Networks with State-Dependent Routing. *IEEE Trans. Networking*, 1(1):105–115, 1993.

[18] A. G. Greenberg and R. Srikant. Computational Techniques for Accurate Performance Evaluation of Multirate, Multihop Communication Networks. *IEEE/ACM Trans. Networking*, 5(2):266–277, 1997.

[19] M. Liu, J. S. Baras, and A. Misra. Performance evaluation in multi-rate multi-hop communication network with adaptive routing. *Proc. Annual ATIRP Conferrence*, 1:136–140, February 1998.

[20] M. K. H. Fan, A. L. Tits, J. Zhou, L. S. Wang, and J. Koninchx. CONSOLE User's Manual. Technical Report TR 87-212r2, Systems Research Center, 1990.

[21] C. Lawrence, J. L. Zhou, and A. L. Tits. User's Guide for CFSQP Version 2.0. Technical Report TR 94-16, Institute for Systems Research, 1994.

APPENDIX

---

[0] The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory or the U.S. Government.

Table A.1: Traffic Rates Used in Example One

Node Pair (k, l)	Class s	$\lambda_{k,l}(s)$	Node Pair (k, l)	Class s	$\lambda_{k,l}(s)$
$(0, 1)$	1	20.0	$(1, 3)$	1	30.0
	2	15.0		2	7.0
	3	10.0		3	17.0
$(0, 2)$	1	5.0	$(1, 4)$	1	3.0
	2	38.0		2	20.0
	3	9.0		3	20.0
$(0, 3)$	1	16.0	$(2, 3)$	1	0.0
	2	17.0		2	15.0
	3	16.0		3	20.0
$(0, 4)$	1	6.0	$(2, 4)$	1	15.0
	2	0.0		2	15.0
	3	32.0		3	20.0
$(1, 2)$	1	37.0	$(3, 4)$	1	51.0
	2	20.0		2	26.0
	3	5.0		3	0.0

Table A.2: Nominal offered traffic load for hierarchical network

Node	Rate	Cap.	load
1.1.1	11.15	270	0.041296
1.1.2	7.20	160	0.045000
1.1.3	10.60	180	0.058889
1.1.4	7.90	150	0.052667
1.1.5	8.60	210	0.040952
1.1.6	8.95	220	0.040682
1.1.7	10.45	180	0.058056
1.2.1	8.95	210	0.042619
1.2.2	8.80	220	0.040000
1.2.3	6.65	130	0.051154
1.2.4	9.15	180	0.050833
1.2.5	9.70	180	0.053889
1.2.6	7.95	140	0.056786
1.3.1	9.55	230	0.041522
1.3.2	12.0	290	0.041379
1.3.3	3.70	80	0.046250
1.3.4	7.90	140	0.056429
1.3.5	8.25	160	0.051562
1.3.6	9.00	180	0.050000
1.3.7	8.00	130	0.061538
1.3.8	5.75	100	0.057500