New Developments and Applications in Experimental Design
IMS Lecture Notes – Monograph Series (1998) Volume 34

OPTIMAL MONITORING OF COMPUTER NETWORKS

VALERII FEDOROV[1] AND DEBORAH FLANAGAN[1]

Oak Ridge National Laboratory

We apply the ideas from optimal design theory to the very specific area of monitoring large computer networks. The behavior of these networks is so complex and uncertain that it is quite natural to use the statistical methods of experimental design which were originated in such areas as biology, behavioral sciences and agriculture, where the random character of phenomena is a crucial component and systems are too complicated to be described by some sophisticated deterministic models. We want to emphasize that only the first steps have been completed, and relatively simple underlying concepts about network functions have been used. Our immediate goal is to initiate studies focused on developing efficient experimental design techniques which can be used by practitioners working with large networks operating and evolving in a random environment.

1. Introduction. In most cases a computer network can be represented as a graph with a given number of nodes (vertices, sites) and edges (links, communication channels). Possible objectives of experiment(s) may include evaluating such quantities as the following: delays on a given subset (subset of interest) of edges, processing times at a given subset of nodes, traveling times from one subset of nodes to another, etc. Existing software and hardware allow measuring [see, for instance, Paxson (1997) for details and a comprehensive list of references] a large variety of network performance indicators, so that in general our "measurement" is a vector. Types of measurement strategies may be very different. For instance, a meter can be installed at any chosen node to measure input and output flows; a measurement software or hardware device can be placed at a host node, and a preselected set of nodes or edges can be monitored; a practitioner can cooperate with others (i.e., there are a few host nodes) to monitor a network. Thus, if we have an opportunity to plan (design) experiments, we may look for the best subset of host nodes where devices must be allocated, find the most informative subset of nodes and edges to be monitored by the given host, or in the most general setting select the most effective team of host nodes (sites) and match them with

---

[1] Research supported by the Oak Ridge National Laboratory Directed Research and Development Fund.

Received September 1997; revised March 1998.

AMS 1991 subject classifications. 62K05.

Key words and phrases. Correlated observations, design of experiments.

the sets of nodes and edges to be monitored. In this paper we confine ourself to the single host problem. Other cases may be analyzed in the same setting but notation and calculus become too complicated and too long to be discussed here.

Our approach is essentially based on the theory of optimal experimental design for correlated observations and particularly on the ideas developed and discussed in Fedorov (1996) and Fedorov and Flanagan (1998).

In Section 2 we introduce the model which consists of two random components: the first one models the network variability, and the second one presents observational errors. This model results in criteria of optimality which depend on the variance-covariance matrix comprising both types of randomness. We investigate the properties in Section 3 and develop the numerical method of construction of the corresponding designs in Section 4. In Section 5 we consider an example based on data collected from the Energy Sciences Network (ESnet) backbone [cf. Paxson (1997)].

Note that we consider the design of nonintrusive experiments, which are very close to what is called "sampling" in more traditional areas of statistics. The active experimentation techniques when the network parameters may be varied (controlled) would provide increased insight on how the particular network is functioning. We leave this more exciting approach for the more mature stages of experimentation on networks.

2. Model and optimality criteria.

2.1 *Main assumptions.* Let us assume that we have one host and S nodes $X = (x_1, \ldots, x_S)$ to be monitored. As it was mentioned, at every node we can observe a few response variables, such as flow rates, delays in various types of processing, queue lengths, etc. To keep notations simple only the univariate case is discussed. We admit the possibility of repeated observations. For instance, a selected node may be interrogated several times during a relatively short period. If the long term trends are neglected (or properly eliminated) then the following model may be applied

$$y_j(x_i) = u(x_i) + \varepsilon_j(x_i), \tag{2.1}$$

where the response function $u(x_i)$ describes the state of the i-th node at the j-th observation, $\varepsilon_j(x_i)$ is the corresponding observational error, $j = 1, \ldots, r_i$. It is assumed that no change occurs in the response function while all r_i observations are collected. All components in (2.1) are assumed to be random variables. The first one, $u(x_i)$, describes the random behavior of the monitored network, while the second one is related to observational errors or short time disturbances. The same characters are used both for random variables and their realizations. The latter ones are standardly marked by additional indices: i.e., $u(x_i)$ stands for the random variable, and $u_j(x_i)$ is its realization.

Let the vector

$$U = (u(x_1), \ldots, u(x_S))^T$$

describe the network performance, and let

$$E_u(U) = U_0, \;\; Var_u(U) = E\left[(U - U_0)(U - U_0)^T\right] = K,$$

where the $S \times 1$ vector U_0 and the $S \times S$ covariance matrix K are given. The subscript u (or ε) means that expectation or variance is taken with respect to u (or ε); subscripts $\varepsilon|u$ (or $u|\varepsilon$) are used for conditional expectations. The obvious transform $U \to U - U_0$ zeroes the expectation of U, and, therefore, in what follows we assume that $E_u(U) = 0$. The observational errors $\varepsilon(x_i)$ are assumed to have zero means and to be uncorrelated:

$$E_{\varepsilon|u}\left(\varepsilon_j(x_i)\right) \equiv 0, \quad E_{\varepsilon|u}\left(\varepsilon_j(x_i)\varepsilon_{j'}(x_{i'})\right) \equiv \sigma^2 \delta_{ii'}\delta_{jj'},$$

where $\delta_{\ell\ell'}$ stands for Kronecker's symbol. Introduction of σ^2 depending on x does not lead to any significant changes and is not considered here. The argument x_i may be considered as a descriptor of site i and might be skipped to simplify notations. However, we will continue to use it to make it easier to bridge our results with standard convex design theory.

Note that in this study we do not use any properties of the set X. For instance, we do not introduce a distance between two points x_i and $x_{i'}$ [cf. Fedorov (1996) and Sacks and Schiller (1988)]. The concept of a distance is much less natural in communication network measurements than in meteorology or seismology, where the physical distance $\| x_i - x_{i'} \|$ between observing stations may define the behavior of elements $K(x_i, x_{i'})$ of the matrix K as functions of $\| x_i - x_{i'} \|$. Introduction of concepts similar to "distance", such as number of switches or complexity of routes between x_i and $x_{i'}$, may lead to more efficient and realistic modeling of communication networks, but it is beyond the scope of this paper.

We assume that for all j the identical (experimental) designs are used:

$$\xi_n = \{p_i, x_i\}_1^n, \ p_i = r_i/N, \ N = \sum_{i=1}^{n} r_i, \ x_i \in X, \ n \le S.$$

Frequently p_i is called the weight of the node (point) x_i. Nodes x_i are called supporting points of the design ξ_n. Let $K(\xi_n)$, which corresponds to the nodes $x_1, \ldots, x_n$; be a submatrix of K, let $K(x, \xi_n)$ be a column vector of covariances between $u(x)$ and $u(x_1), \ldots, u(x_n)$, and let the matrix $W(\xi_n)$ be diagonal with the elements $W_{ii} = N\sigma^{-2} p_i \delta_{ii}$. We also use the matrices $K(Z, \xi_n) = (K(x_1, \xi_n), \ldots, K(x_q, \xi_n))$, where $x_1, \ldots, x_q \in Z \subset X$, and $K(Z)$ is a submatrix of K corresponding to the same nodes.

2.2. *Estimation and optimality criteria.* Let $Y(\xi_n)$ be the vector of averaged observations made according to ξ_n:

$$Y(\xi_n) = \begin{pmatrix} \frac{1}{r_1}\sum_{j=1}^{r_1} y_j(x_1) \\ \vdots \ \vdots \ \vdots \\ \frac{1}{r_n}\sum_{j=1}^{r_n} y_j(x_n) \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \ \vdots \ \vdots \\ Y_n \end{pmatrix}.$$

It can be verified by direct minimization that the predictor

$$\hat{U}(Z) = K^T(Z, \xi_n)\left(K(\xi_n) + W^{-1}(\xi_n)\right)^{-1} Y(\xi_n) \tag{2.2}$$

minimizes the matrix of expected squared residuals

$$D\left(\xi_n, \tilde{U}(Z)\right) = E_{u,\varepsilon}\left[\left(\tilde{U}(Z) - U(Z)\right)\left(\tilde{U}(Z) - U(Z)\right)^T\right]$$

among all linear estimators $\tilde{U}(Z) = LY(\xi_n)$ such that

$$E_{u,\varepsilon}\left[\tilde{U}(Z) - U(Z)\right] = 0.$$

In other words

$$D(\xi_n) = D\left(\xi_n, \hat{U}(Z)\right) \leq D\left(\xi_n, \tilde{U}(Z)\right), \tag{2.3}$$

where inequality must be understood in the sense of ordering of nonnegative definite matrices (cf. Horn and Johnson, (1985), Chpt. 7.7). From (2.1) and (2.2) it follows that

$$D(\xi_n) = K(Z) - K^T(Z, \xi_n)\left(K(\xi_n) + W^{-1}(\xi_n)\right)^{-1} K(Z, \xi_n). \tag{2.4}$$

The objective of this study is to provide methods which allow the minimization of some given functions of the matrix $D(\xi_n)$, for instance, $\mathrm{tr}D(\xi_n)$, $\ln|D(\xi_n)|$, $\max_i D_{ii}(\xi_n)$, etc. See Fedorov and Hackl (1997) or Pukelsheim (1993) for details about optimality criteria. Thus, we have to consider the following optimization problem

$$\xi_n^* = \arg\min_{\xi_n} \Psi\left[D(\xi_n)\right], \tag{2.5}$$

where Ψ is a selected objective function (criterion of optimality). In (2.5) the number of nodes (or supporting points) n is fixed, and the total number of available observations $N = \sum_{i=1}^{n} r_i$ is assumed to be given. In general, n may be optimized as well.

3. Designs with continuous weights. Two features of optimization problem (2.5) may cause serious computational hurdles: weights p_i are discrete, and the optimal number of supporting points must be found, in general. The problem is simplified both theoretically and numerically if we allow weights to be continuous, so that $0 \leq p_i \leq 1$, $\sum_{i=1}^{n} p_i = 1$, and make $n = S$. If an optimal n is less than S then some of the weights are zeroes. In other words, instead of ξ_n^*, we are going for some approximate solution which usually works well for larger N. Extensive discussions about connections of optimal designs with continuous weights and exact optimal designs may be found in Fedorov and Hackl (1997) and Pukelsheim (1993).

For $n = S$ and Z coinciding with X from (2.4) and the identity $(A+B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1}$, it follows that for any design ξ

$$D(\xi) = \left(K^{-1} + W(\xi)\right)^{-1}, \tag{3.1}$$

where the subscript n is skipped for obvious reasons. The regularity of the matrix K is assumed in the latter formula. If Z is a subset of X then the covariance matrix (2.4) is an obviously defined submatrix of (3.1). The subscript S will be skipped if it does not lead to ambiguity.

Using (3.1) we can reformulate the design problem as

$$\xi^* = \arg\min_{\xi} \Psi\left[D(\xi)\right], \tag{3.2}$$

where ξ can be any probability distribution with support X.

Let us assume that the function $\Psi[D(\xi)]$ is a convex function of ξ and has a directional derivative $\psi(\xi^*, \xi)$ at ξ^* in the direction ξ.

THEOREM 1. *A necessary and sufficient condition for a design ξ^* to be optimal is fulfillment of the inequality*

$$\psi(\xi^*, \xi) \geq 0 \tag{3.3}$$

for any other design ξ.

This result is well known in optimization theory of convex functions and is widely used in experimental design theory (cf. Cook and Fedorov (1995), Fedorov and Hackl (1997) Chpt. 2). For the D-criterion, $\Psi(D) = \ln|D|$, we have

$$\psi(\xi^*, \xi) = \mathrm{tr} D(\xi^*)\left(W(\xi^*) - W(\xi)\right). \tag{3.4}$$

Noting that $\mathrm{tr} D(\xi)W(\xi) = \sigma^{-2}N\sum_{i=1}^{S} D_{ii}(\xi)p_i$ and combining (3.3) and (3.4), we have

THEOREM 2. *A necessary and sufficient condition for a design ξ^* to be D-optimal is that*

$$\max_i D_{ii}(\xi^*) \leq \frac{\sigma^2}{N}\mathrm{tr} W(\xi^*)D(\xi^*),$$

and equality holds at all points where $p_i^* > 0$.

A D-optimal design minimizes the maximal variance of prediction

$$\xi^* = \arg\min_{\xi}\max_i D_{ii}(\xi).$$

In this theorem and in what follows $\max_i$ means maximization over all points from X, i.e., $1 \leq i \leq S$. Thus, observations in a D-optimal or minimax design must be placed at points (sites) where prediction might be the worst.

For linear criteria $\Psi(D) = \mathrm{tr}AD$, where $A \geq 0$ is the so-called utility matrix, we have

$$\psi(\xi^*, \xi) = \mathrm{tr} D(\xi^*)AD(\xi^*)\left(W(\xi^*) - W(\xi)\right),$$

and the following result holds.

THEOREM 3. *A necessary and sufficient condition for a design ξ^* to be linear optimal is that*

$$\max_i \{D(\xi^*)AD(\xi^*)\}_{ii} \leq \frac{\sigma^2}{N}\mathrm{tr} W(\xi^*)D(\xi^*)AD(\xi^*),$$

and the equality holds at all points where $p_i^* > 0$.

If $A = I$, i.e., the average variance of prediction must be minimized, then the theorem tells us that an optimal design ξ^* allocates observations at sites in which the predicted value $\hat{U}(x^*)$ of $U(x^*)$ might have the greatest average squared covariance with all other $\hat{U}(x), x \in X$.

4. First order algorithms. The above theorems help to develop and analyze various first order algorithms for construction of optimal designs. For computer networks the matrices processed during computations have large sizes. It is, therefore, especially important to use recursions which are computationally simple and stable. The most convenient in this sense are algorithms similar to the first order exchange type algorithms (see, for instance, Mitchell (1974), Fedorov and Hackl(1997), Chpt.3). Their main idea is very simple: at each stage add that new point which improves the current design most and delete from the same (or just corrected) design that point which contributes least. Here we formulate the simplest version of that kind of algorithm for D-criteria.

Let the initial design ξ_0 be such that all weights $p_{0i} = b_i\alpha_0$, where b_i is an integer and $\sum_{i=1}^{S} p_{0i} = 1$. For instance, we may choose $b_i \equiv 1$ and $\alpha_0 = 1/N$.

1. Given ξ_t and $D(\xi_t)$, find
$$a = \arg\max_i D_{ii}(\xi_t).$$
Add α_t to the weight of point x_a to construct the design ξ_t^+ and the matrix $D(\xi_t^+)$. Note that the sum of the weights in the latter design is greater than 1.

2. Find
$$d = \arg\min_{i\in I_t^+} D_{ii}(\xi_t^+),$$
where I_t^+ is the set of all supporting points of ξ_t^+, i.e., points with nonzero weights at step t. Delete α_t from the weight of point x_d to modify ξ_t^+ and to construct ξ_{t+1}, in which the sum of the weights is restored to 1 as it was in ξ_t.

3. If $|D(\xi_{t+1})|/|D(\xi_t)| < 1-\gamma$, where γ is a small positive number less than 1, then put $\alpha_{t+1} = \alpha_t$ and go to step 1. Otherwise make $\alpha_{t+1} = \alpha_t/2$, and then go to step 1.

Computations may be stopped when α_t is sufficiently small.

The choice of p_{0i} and α_0 is a matter of convenience. For instance, the above choice guarantees that no more than α_t^{-1} observations are needed to avoid any "fractional" observation in design ξ_t, which is a frequent case in the continuous design theory setting. Actually, in the "classical" version of the exchange algorithm $\alpha_t \equiv N^{-1}$, where N is a preselected number of observations. The algorithm with $\alpha_t \equiv N^{-1}$ was applied to the construction of optimal spatial designs by Sacks and Schiller (1988) in a slightly different setting (repeated observations were not allowed). Unfortunately, in this case the limit design (if it exists) is generally not an optimal one. We, therefore, introduced the possibility of infinitely reducing the step length α.

The rule for adding and deleting weights becomes obvious if we note that

$$|D(\xi_t^+)| = \frac{|D(\xi_t)|}{1+\zeta_t D_{aa}(\xi_t)} \quad \text{and} \quad D(\xi_t^+) = D(\xi_t) - \frac{\zeta_t C^+(\xi_t)}{1+\zeta_t D_{aa}(\xi_t)}, \tag{4.1}$$

where $C_{ij}^{+}(\xi_t) = D_{ai}(\xi_t)D_{aj}(\xi_t)$ and $\zeta_t = \sigma^{-2}N\alpha_t$. The above formulae may be derived using the fact that

$$D(\xi_t^{+}) = \left(K^{-1} + W(\xi_t) + \zeta_t \ell_a \ell_a^T\right)^{-1} = \left(D^{-1}(\xi_t) + \zeta_t \ell_a \ell_a^T\right)^{-1},$$

where $\{\ell_a\}_i = \delta_{ia}$, and some exercises in matrix algebra (cf. Fedorov and Hackl (1997), Chpt. 3). In the versions of (4.1) for the deleting procedure, ζ_t must be replaced by $-\zeta_t$ and ξ_t by ξ_t^{+}.

Similar to the classical results of experimental design theory, the following result may be established.

THEOREM 4. *The sequence* $\{|D(\xi_t)|\}$ *converges and*

$$\min_{\xi} |D(\xi)| \leq \lim_{t\to\infty} |D(\xi_t)| \leq (1-\gamma)^{-1} \min_{\xi} |D(\xi)|.$$

The proof is based on monotonicity of the iterative procedure, convexity of $\ln|D(\xi)|$ as a function of ξ, and Theorem 2.

Note that formulae (4.1) and their siblings for the deleting steps are very handy recursions for large size problems.

5. Example. We have used the Department of Energy's ESnet backbone, a portion of the Internet, as a testbed for the numerical procedure proposed in Section 4. Using a network host computer at Oak Ridge National Laboratory, we interrogated 39 other sites (see Appendix A) to construct a reasonable estimate $\hat{K}$ for the matrix K. We have used the Packet Internet Groper (*ping*) software (authored by M. Muss, U. S. Army Ballistic Research Laboratory in 1983) to measure the response time for each interrogation. Because of the lower priority that network routers may give to *ping* requests, the minimum response time (among three *ping* requests per interrogation) is used as the response variable. This also reduces the probability of "missing" observations (i.e., there is more hope that at least one *ping* out of three will result in a response). All 39 sites were interrogated 50 times in random order. We estimated the elements of matrix K for each pair of sites separately without imposing any conditions like positive-definiteness of $\hat{K}$, for instance. We find this very simple approach is sufficient for the current exercise and understand that the estimation of K is a special and very delicate problem which is beyond the scope of this paper. The value of the standard error σ was estimated through averaging differences between results of neighboring in time interrogations over the whole set of interrogations for all sites. We found that $\sigma \simeq 8.0ms$. It is not the best estimator, especially if one takes into account an obvious heterogeneity of ESnet. However, for our illustrative purposes it is not important. Note that the structure of an optimal designs depends on the ratio σ^2/N (see Theorems 2 and 3).

The algorithm from Section 4 is used with K replaced by $\hat{K}$, and results appear in Appendix B. Only the diagonal elements of matrix $\hat{K}$ are listed in Appendix B in order to conserve space. Appendix B also reports the optimal weights for each site and

the variance of prediction (the diagonal elements D_{ii}) for the D-optimal continuous and rounded designs. Also reported is the variance of prediction for the uniform design containing all 39 sites. The optimal design is nearly four times more efficient than the 39-point uniform design. The number of *pings* sent to a particular site must be proportional to the corresponding weight. Obviously, in practice we used "rounded" weights. Of course, this rounding may lead to some increase of the maximal variance of prediction. For instance, when only 10 points with the largest weight are selected and all their weights are set to 0.1, then $\max_i D_{ii} = 74.7$, which is not significantly larger than $\max_i D_{ii}$ for the continuous D-optimal design.

We selected a relatively small number of available observations ($N = 10$) to emphasize the difference between continuous and discrete designs. In reality, it is a matter of a few minutes to send hundreds of *pings* to different sites, and, therefore, the approximation of reasonable weights is not a serious problem in that type of experiment. Actually, one may talk about the optimal partitioning of the time available for a given experimental session for monitoring various sites instead of the selection of an optimal number of *pings*.

Acknowledgments. We thank Steve Batsell for his valuable comments and Tom Dunigan for writing the program to collect the raw *ping* data. The very constructive remarks and recommendations from the two referees were extremely useful in preparation of the final version of our paper.

REFERENCES

COOK, R. D. AND FEDOROV, V. (1995). Constrained optimization of experimental design. *Statistics* **26** 129-178.

FEDOROV, V. (1996). Design of spatial experiments: model fitting and prediction. In *Handbook of Statistics, Volume 13* (S. Ghosh and C. R. Rao, eds.), 515-553. Elsevier, New York.

FEDOROV, V. AND FLANAGAN, D. (1998). Optimal monitoring network design based on Mercer's expansion of covariance kernel. *Journal of Combinatorics, Information and System Sciences*, in press.

FEDOROV, V. AND HACKL, P. (1997). *Model Oriented Design of Experiments.* Springer-Verlag, New York.

HORN, R. AND JOHNSON, C. (1985). *Matrix Analysis.* Cambridge University Press, London.

MITCHELL, T. (1974). An algorithm for construction of D-optimal design. *Technometrics.* **16** 203-210.

PAXSON, V. (1997). *Measurement and Analysis of End-to-End Internet Dynamics.* University of California, Berkley (doctoral thesis).

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

SACKS, J. AND SCHILLER, S. (1988). Spatial design. In *Statistical Decision and Related Topics IV* (J. Berger and S. Gupta, eds.), 385-395. Academic Press, New York.

COMPUTER SCIENCE AND MATHEMATICS DIVISION
OAK RIDGE NATIONAL LABORATORY
P.O.BOX 2008
OAK RIDGE, TN 37831-6367 U.S.A.

APPENDIX A

Site Identifiers

ID	Site Acronym	Site Name
1	JLAB	Thomas Jefferson National Accelerator Facility (Newport News, VA)
2	ARM	Atmospheric Radiation Measurement Project (Lamont, OK)
3	FNAL	Fermi National Accelerator Laboratory (Batavia, IL)
4	SNL	Sandia National Laboratories Albuquerque (Albuquerque, NM)
5	KEK	KEK, Japan
6	NYU*	New York University Courant Institute (New York, NY)
7	MSRI*	Mathematical Sciences Research Institute, Univ. CA (Berkeley, CA)
8	ANL-MR1	Argonne National Laboratory-Main Router 1
9	AMES	AMES Laboratory, Iowa State University (Ames, IA)
10	FSU*	Florida State University (Tallahassee, FL)
11	CIT	California Institute of Technology (Pasadena, CA)
12	MIT*	Massachusetts Institute of Technology (Cambridge, MA)
13	FNAL-MR1	Fermi National Accelerator Laboratory-Main Router 1
14	GAT	General Atomics (San Diego, CA)
15	UTA	University of Texas at Austin (Austin, TX)
16	SRS*	Savannah River Site (Aiken, SC)
17	SLAC	Stanford Linear Accelerator (Stanford, CA)
18	INEL*	Idaho National Engineering Laboratory (Idaho Falls, ID)
19	LLNL	Lawrence Livermore National Laboratory (Livermore, CA)
20	AUCK*	University of Auckland (Auckland, New Zealand)
21	DOE	Department of Energy (Washington, DC)
22	PPPL	Princeton Plasma Physics Laboratory (Princeton, NJ)
23	UTK	University of Tennessee (Knoxville, TN)
24	LANL-MR1	Los Alamos National Laboratory - Main Router 1 (Los Alamos, NM)
25	NASA*	AMES Research Center, NASA (San Francisco, CA)
26	BNL	Brookhaven National Laboratory (Upton, NY)
27	PPPL-local	additional PPPL site
28	CU	Columbia University Academic Information Systems (New York, NY)
29	ANL	Argonne National Laboratory (Argonne, IL)
30	Pro.PPPL	additional PPPL site
31	PNNL*	Pacific Northwest National Laboratory (Richland, WA)
32	OSTI	Office of Scientific and Technical Information (Oak Ridge, TN)
33	NEVIS*	Columbia University Nevis Laboratory (Irvington, NY)
34	LBNL-MR1	Lawrence Berkeley National Laboratory - Main Router 1
35	LLNL-MR2	Lawrence Livermore National Laboratory - Main Router 2
36	NERSC*	National Energy Research Scientific Computing, LBNL (Berkeley, CA)
37	LBNL	Lawrence Berkeley National Laboratory (Berkeley, CA)
38	SNL/LLNL	Sandia National Laboratories at LLNL (Livermore, CA)
39	YALE*	Yale University (New Haven, CT)

APPENDIX B

Various Designs to Monitor ESNet Sites ($N = 10, \sigma = 8.0ms$)

Site ID	Site	Site Variance	D-optimal Weight	Variance of the Prediction, D_{ii}		
				D-optimal Continuous	Rounded 10 points	Uniform All 39 points
1	JLAB	96.8	.0000	23.7	24.2	12.8
2	ARM	72.0	.0000	35.1	42.0	25.1
3	FNAL	90.7	.0000	21.2	23.2	8.2
4	SNL	8.8	.0000	2.8	2.8	1.9
5	KEK	56.2	.0000	31.1	31.5	18.4
6	NYU	1042.7	.0970	58.0	56.4	176.7
7	MSR1	289.2	.0128	57.9	61.2	39.8
8	ANL-MR1	100.9	.0000	32.7	32.6	19.8
9	AMES	83.5	.0000	26.0	27.9	11.7
10	FSU	1497.7	.1010	58.0	58.5	194.8
11	CIT	19.3	.0000	8.3	8.9	4.9
12	MIT	1002.6	.1000	57.8	57.9	187.6
13	FNAL-MR1	6.1	.0000	5.4	5.5	3.8
14	GAT	60.5	.0000	21.5	23.0	13.2
15	UTA	54.0	.0000	17.7	18.4	12.3
16	SRS	978.5	.0985	57.9	57.1	181.3
17	SLAC	93.0	.0000	26.0	31.8	14.9
18	INEL	1537.9	.1010	58.0	58.5	194.7
19	LLNL	57.8	.0000	27.7	28.2	15.5
20	AUCK	1967.8	.1036	57.9	60.0	210.8
21	DOE	25.3	.0000	12.4	12.7	6.9
22	PPPL	41.9	.0000	11.4	12.2	5.2
23	UTK	0.4	.0000	0.3	0.3	0.3
24	LANL-MR1	51.7	.0000	21.1	21.8	10.7
25	NASA	949.6	.0978	58.9	56.8	177.8
26	BNL	183.8	.0000	56.7	56.8	33.9
27	PPPL-local	126.0	.0000	48.2	51.3	23.1
28	CU	75.3	.0000	23.9	23.9	15.7
29	ANL	88.3	.0000	22.4	24.1	10.1
30	Pro.PPPL	35.3	.0000	15.6	15.6	9.4
31	PNL	365.1	.0853	57.9	51.0	130.1
32	OSTI	0.4	.0000	0.3	0.3	0.3
33	NEVIS	402.1	.0809	57.9	52.8	70.3
34	LBNL-MR1	62.4	.0000	16.4	17.4	7.0
35	LLNL-MR2	58.9	.0000	25.5	26.9	12.7
36	NERSC	121.2	.0236	58.0	74.7	47.2
37	LBNL	86.7	.0000	21.3	22.0	10.7
38	SNL/LLNL	131.8	.0000	43.1	46.3	20.9
39	YALE	1137.1	.0987	58.0	57.3	183.6