Characterizing Optimal Topological Structures for a Class of Large Distributed Data Networks

John K. Antonio
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

ABSTRACT

The primary goal of this paper is to present a new and fundamental graph-theoretic result for characterizing optimal topological structures. Here, an optimal topology is defined as one which maximizes the number of origin destination pairs that can communicate concurrently, while satisfying practical constraints related to performance, capital investment, and reliability. For a given network topology, the graph-theoretic result gives a bound for the maximum number of origin destination pairs which can have concurrent communication. This theoretical result is in the form of a simple inequality which relates (among others) the following three parameters: number of concurrently communicating origin destination pairs, number of network links, and network diameter. The novelty of this result with respect to past research in the area is its strong graph-theoretic foundation. The paper aims at developing the mathematical machinery needed to cut to the heart of the topology design problem, as opposed to past approaches which rely heavily on heuristics or "rules of thumb."

I. INTRODUCTION

In virtually all applications for data networks, information is exchanged between a collection of computers which are geographically distributed. Unfortunately, having direct data links connecting every pair of computers in a large data network is usually inconceivable—due primarily to the associated cost. Instead, large data networks are typically sparsely interconnected. Each node (i.e., computing site) is directly connected to only a few neighboring nodes, which are in turn connected to their own sets of neighbors. By viewing a data network as a directed graph $G = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ denotes the set of nodes and $\mathcal{L}$ the set of data links, a necessary condition for enabling communication between any pair of nodes is for G to be a connected graph. That is, for each pair of nodes, say i and j, there must exist at least one path connecting i to j. Therefore, nodes which are not directly connected can still communicate by sending data along one or more interconnecting paths. This brings to light two important questions. First, how should the nodes of the network be interconnected—referred to as the topology design problem. Second, for a given network topology, which path or paths should be used in transmitting data between each pair of communicating nodes—referred to as the routing problem.

In addressing the topology design problem, researchers typically consider three main issues: capital investment, performance, and reliability.[1] Capital investment is dependent on factors such as the number, bandwidth, and physical length of data links used. Performance is typically measured by using an objective function which quantifies, in some sense, the amount of time delay associated with sending data through the network (i.e., the level of network congestion). Reliability is measured in terms of the "connectedness" of the network in the presence of node and/or link failures.

The primary objective of the routing problem is to select routes for the set of origin destination (OD) pairs which request communication (called the active OD pairs), see for example, references [1–6]. This route selection problem is usually formulated in terms of optimizing some measure of network performance.

Although it is convenient to logically separate the problem of topology design from that of routing, their solutions are actually closely interdependent. In particular, it is possible for an "optimal" network topology to exhibit very poor performance measures if some type of intelligent routing is not employed. For the purposes of this paper, the characterization of optimal topologies implicitly assumes that some form of routing be used in practice.

The remainder of the paper is organized as follows. In Section II, the bare essentials of network performance and optimal routing are covered. Section III covers the topology design problem (including a summary of past approaches) and sets the stage for the new approach to characterizing optimal topologies. Section IV includes the main graph-theoretic result, and shows how it can be used to characterize optimality of network topologies. Simulations are included which indicate that the theoretical (upper bound) result provides an excellent means for estimating the number of concurrently communicating OD pairs a particular network topology can support. Section V is devoted to showing how the new *characterization* of optimal topologies can be used as a basis for developing new *design methodologies*.

II. NETWORK PERFORMANCE AND OPTIMAL ROUTING

Queuing theory is the primary methodological framework for analyzing network performance [4]. Perhaps the simplest queuing model is the so-called $M/M/1$ queuing system which consists of a single queuing station and a single server. For link (i,j), it is assumed that packets arrive according to a Poisson process with rate F_{ij}, and the probability distribution of the service time is exponential with mean $1/C_{ij}$. By applying Little's Theorem, the average delay for a packet to traverse link (i,j) is given by

$$D_{ij} = \frac{1}{C_{ij} - F_{ij}}. \tag{1}$$

Jackson's Theorem states that in a network of single server queues in which customers arrive from outside the network at each queue according to independent Poisson processes, the average number of outstanding packets in the system (in the steady-state) can be derived as if each queue in the network is an $M/M/1$ queue. So, for the purpose of measuring network performance, modeling the entire network with simple $M/M/1$ queues is somewhat justified.

Now, based on the result of Jackson's Theorem and equation (1), our performance cost function is defined as the sum of all link delays weighted according to their relative importance. So, we get the following function

$$D(F) = \sum_{(i,j)\in\mathcal{L}} \frac{F_{ij}}{C_{ij} - F_{ij}}, \tag{2}$$

where links having more traffic flow are given higher relative weightings. It can be verified that each term in the sum represents the average size of the queue associated with link (i,j). Therefore, $D(F)$ represents the average number of outstanding packets in the network.

For the purposes of this paper, a set of routes which minimize $D(F)$—for a given set of OD traffic demands—will constitute the notion of an optimal routing. It turns out that $D(F)$ is a convex function of the variables F_{ij}, and therefore standard numerical optimization algorithms can be used to determine the global optimal solution. Since the primary focus of this paper is the topology design problem, the reader is referred to [1–6] for detailed formulations and associated solution techniques for the optimal routing problem.

[1] In the present work, another important issue is addressed—maximizing the number of distinct node pairs which can communicate concurrently—more on this in subsequent sections.

CH2868-8/90/0000-2396$1.00 © 1990 IEEE

III. The Topology Design Problem

Unlike the optimal routing problem, most formulations of the topology design problem do not yield mathematically tractable problems. In fact, most formulations have many locally optimal solutions, and require solving hard combinatorial problems to arrive at the global optimal solution. For this reason, past research in the area of designing data networks has resulted primarily in heuristic algorithms which iteratively perturb an initial topology, according to a set of rules, in search of at least a local optimal solution [4,7,8]. Unfortunately, this type of approach does not give much (if any) analytical insight into the underlying structure of a true optimal solution. The underlying topological structure of the solutions obtained by using these hueristic algorithms can change drastically depending on the choice of the initial network.

A. Past Formulations

In formulating the topology design problem, past researchers have typically considered three main issues:

1. Network Performance: Usually some measure of overall network congestion. The function $D(F)$ derived in the previous section is a popular choice.
2. Reliability: A measure of the connectedness of the network in the presence of node and/or link failures. The concept of a k-connected graph is commonly used. A graph is said to be k-connected if every subgraph obtained by deleting $k-1$ nodes is connected.
3. Capital Investment: Dependent on the number, capacity, and physical length of the data links used.

Based on the above issues, most of the past formulations of the topology design problem may be categorized into one of the three formulations below.

Given: (a) The geographic location of the nodes, and (b) The input traffic demands of the OD pairs.

(F1) Maximize network performance, subject to reliability and capital investment as constraints.

(F2) Maximize reliability, subject to network performance and capital investment as constraints.

(F3) Minimize capital investment, subject to network performance and reliability as constraints.

As mentioned previously, formulations such as (F1), (F2), and (F3) typically result in problems that are NP hard. Therefore, heuristic techniques have been the standard approach in the past.

B. The New Proposed Approach

One of the major flaws with past formulations (besides the fact that they are difficult to solve) is that they rely heavily on having precise knowledge of the expected input traffic demands for all OD pairs. Due to the "bursty" and dynamic nature of actual traffic demands, this single constant parameter is not adequate for describing the actual phenomenon. That is, in practical situations the value of this parameter for each OD pair is extremely time dependent. So, previous formulations (such as F1 through F3) face a difficult dilemma: (1) The eventual solutions depend strongly on the constant average values used to estimate the input traffic demands. (2) In reality, the input traffic demands are time varying. Result: A static network is designed based on an over-simplified model of the actual dynamic input demands.

In this paper, we propose designing networks based on the idea of maximizing the expected number of OD pairs which can have concurrent communication. In so doing, the aforementioned dilemma is avoided. The issues of network performance, reliability, and capital investment can be incorporated as constraints in the new formulation.

IV. The Main Result

A. Theoretical Development

A graph-theoretic result has been derived which characterizes the inherent tradeoffs and inter-relationships of the topology design parameters. (Due to the space limit, the formal proof of this result is not included.) In order to state the main result we need two preliminary definitions. The first defines a *valid* routing solution as one in which the flow on each network link $(i,j) \in \mathcal{L}$, denoted as F_{ij}, is strictly less than the capacity of the associated link, C_{ij}.

Definition: A valid routing solution is one in which $F_{ij} < C_{ij}$, for all $(i,j) \in \mathcal{L}$. Note: From queuing theory we know that if $F_{ij} \to C_{ij}$, then the size of the queue associated with link (i,j) will grow without bound, see equation (1).

Next, the definition of a valid routing solution is used to define the concept of concurrently communicating OD pair sets (CCOD pair sets).

Definition: Given a network graph $G = (\mathcal{N}, \mathcal{L})$ with link capacities C_{ij}; we say that an OD pair set W (with associated constant traffic demands $r_w, w \in W$) is a CCOD pair set, if there exists a valid routing solution.

Now we are ready to state the main result.

The Main Result: For a given network graph $G = (\mathcal{N}, \mathcal{L})$ with link capacities C_{ij}, the size of all possible CCOD pair sets, denoted as $|W_{cc}|$, is bounded above by

$$|W_{cc}| \leq \frac{C_{max}}{r_{min}} \frac{|\mathcal{L}|}{h_{avg}}, \tag{3}$$

where $|\mathcal{L}|$ is the number of network links, C_{max} is the maximum link capacity, defined by $C_{max} = \max_{(i,j)\in\mathcal{L}}\{C_{ij}\}$, r_{min} is the minimum active OD pair traffic demand, defined by $r_{min} = \min_{w\in W_{cc}}\{r_w\}$, h_{avg} is the average minimum hop distance between OD pairs in W_{cc}, defined by $h_{avg} = \frac{1}{|W_{cc}|}\sum_{w\in W_{cc}} h_w$, where h_w is the minimum hop distance associated with OD pair w.

The main result, i.e., equation (3), coincides with intuition. For example, for a fixed network topology one would expect that if the value of C_{max} is increased, then the potential for accommodating more CCOD pairs should increase as well. Similar statements can be made regarding the parameters r_{min} and $|\mathcal{L}|$, i.e., increasing $|\mathcal{L}|$ and/or decreasing r_{min} should increase the potential for more concurrent communication. Finally, the fact that the bound is inversely proportional to the average minimum hop distance of the OD pairs in W_{cc} also makes intuitive sense. For instance, if the set of OD pairs are such that each OD pair is "close" in terms of minimum hop distances, then fewer links will be utilized when all active OD pairs are concurrently communicating. On the other hand, if the OD pairs tend to be "far away" from each other, then the number of links utilized for communication by just a single OD pair may be on the order of the diameter of the network.[2]

In general, the value of h_{avg} is bounded above and below by

$$1 \leq h_{avg} \leq d, \tag{4}$$

where d denotes the diameter of the network. Substituting these bounds into equation (3), we have

$$|\overline{W_{cc}}| \leq \frac{C_{max}}{r_{min}} \frac{|\mathcal{L}|}{1} \tag{5}$$

and

$$|\underline{W_{cc}}| \leq \frac{C_{max}}{r_{min}} \frac{|\mathcal{L}|}{d}, \tag{6}$$

where $|\overline{W_{cc}}|$ and $|\underline{W_{cc}}|$ denote the maximum and minimum possible upper bounds for $|W_{cc}|$, respectively. Note that the bound for $|\underline{W_{cc}}|$ in equation (6) can be thought of as a conservative (or worst case) bound for the parameter $|W_{cc}|$ of equation (3). For general CCOD pair sets and general network topologies, one can only assume that $h_{avg} = O(d)$; therefore, equation (6)

[2]The diameter of a network is defined as the maximum value of h_w, taken over all possible OD pairs w.

provides a realistic bound (generically) for $|W_{cc}|$. Also, the fact that this bound is inversely proportional to the diameter of the network provides an excellent means of characterizing the optimal topological structure (i.e., optimal in terms maximizing $|W_{cc}|$). Namely, choose the diameter as small as possible, while maintaining capital investment, performance, and reliability at acceptable levels.

B. *Characterizing Optimal Topologies*

The above graph-theoretic result can be used as a basis for characterizing optimal topological structures. In particular, the fact that the least upper bound for $|W_{cc}|$ is inversely proportional to the diameter of the network will be exploited. Of course one could always accommodate larger CCOD pair sets by increasing the number of links $|\mathcal{L}|$ and/or increasing the maximum link capacity, C_{max}. However, in practice $|\mathcal{L}|$ and C_{max} can not be increased past certain values, because of the associated economic constraints and/or physical limitations.

Preliminary numerical studies were done that indicate "on the average," the least upper bound of the main result is asymptotically tight. The numerical studies were set-up as follows. First, a simple topological structure is chosen: linear, star, or square mesh network.[3] The value of C_{max}/r_{min} is fixed, and the number of nodes in the network, say n, is fixed. The set of active OD pairs W is initialized to the empty set. Then, an active OD pair and associated demand is chosen at random and added to the set of active OD pairs. The optimal routing problem is solved (using the gradient projection based algorithm of [5]). If the solution to the optimal routing problem is valid (i.e., $F_{ij} < C_{ij}$, for all $(i,j) \in \mathcal{L}$), then another OD pair is chosen at random and added to the set of active OD pairs. This procedure is continued until the solution of the optimal routing solution is no longer valid. At this point, the number of active OD pairs are stored and the procedure is done again (with W initialized to the empty set). The average of the number of OD pairs accommodated for each run is then computed; denote this quantity as $|W|_{avg}$.

Let $K = C_{max}/r_{min}$ in the following description. For the n-node linear network depicted in Fig. 1(a), note that the number of links is given by $|\mathcal{L}| = n-1$ and the the diameter is obviously $d = n-1$. Thus, according to equation (6) the least upper bound for $|W_{cc}|$ is given by

$$|\underline{W_{cc}}| \leq K\left(\frac{n-1}{n-1}\right) = K. \tag{7}$$

So, assuming that K is independent of n (which is reasonable), we see that the least upper bound for $|W_{cc}|$ is a constant. Fig. 2(a) shows the simulation results for the linear network, which indicate that the simulated value of $|W|_{avg}$ is independent of n as well.

Similar numerical studies were done for the n-node star network shown in Fig. 1(b). For this network note that the number of links is $|\mathcal{L}| = n-1$ and the diameter is $d = 2$. Therefore, according to the graph-theoretic result, we have

$$|\underline{W_{cc}}| \leq K\left(\frac{n-1}{2}\right). \tag{8}$$

The results of the numerical studies for the star network are summarized by the graph of Fig. 2(b). Note that the value of $|W|_{avg}$ increases linearly with n, as expected.

To further justify the theory, preliminary studies were done for the n-node mesh network, see Fig. 1(c). In an n-node mesh network, we have $|\mathcal{L}| = K_1 n$ and $d = K_2\sqrt{n}$, thus the least upper bound becomes

$$|\underline{W_{cc}}| \leq K_3\sqrt{n}, \tag{9}$$

where K_1, K_2, and K_3 are constants independent of n. The results of the simulations for the mesh network given in Fig. 2(c), agree with the theoretically derived least upper bound. Namely, $|W|_{avg} = O(\sqrt{n})$.

[3]These simple topologies were used because their diameters are well defined as a function of the number of nodes. Also, their high degree of regularity simplified the programming task as the number of nodes was increased.

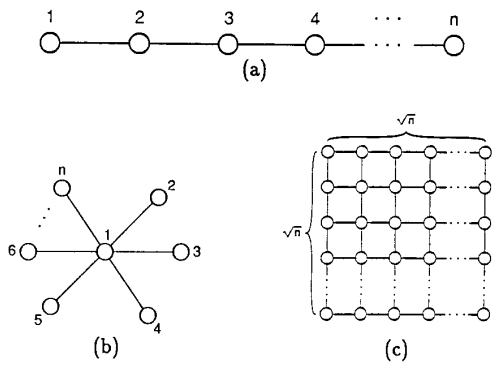
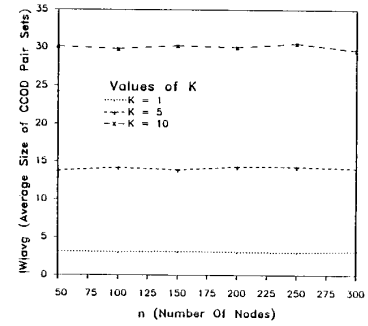
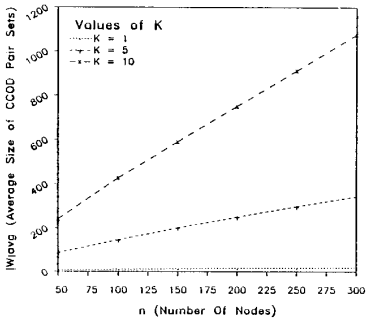


Fig. 1 (a) The n-node linear network. (b) The n-node star network. (c) The n-node mesh network.





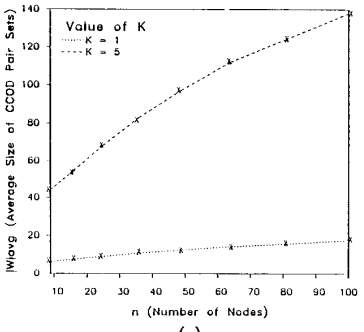


Fig. 2 The results of the simulation studies showing values of $|W|_{avg}$ vs. n for the n-node (a) linear, (b) star, and (c) mesh network.

Note that the linear, star, and mesh networks all have $|\mathcal{L}| = O(n)$. However, even though the star network can accommodate the most OD pairs (on the average), there is a severe tradeoff in terms of reliability. That is, if the the center node of a star network fails, the entire network becomes completely disconnected. While the linear network is also 1-connected, the overall integrity of the linear network is better than that of the star network. Finally, the mesh network strikes a reasonable compromise in that the it is 2-connected, while providing $|\underline{W_{cc}}| = O(\sqrt{n})$.

It is noted further that the linear, star, and mesh topologies were used in this preliminary study primarily because their simple and regular structures simplified the programming task (as the number of nodes was increased). Future simulation studies are planned for more irregular and complicated structures. Of particular interest is the balanced hierarchically clustered topology, which is described in more detail in the next section.

V. Future Work: Design Methodologies

The next logical step is to develop and explore various design methodologies based on the new graph-theoretic result. One promising approach is to use the balanced hierarchically clustered (BHC) topology developed in [1], as a sort of structural target for topology design. In [1] we showed that there exists a sub-class of BHC topologies for which the diameter is $O(\log n)$. Also, it can be shown that that such topologies contain only $O(n)$ links. So, in cases where only $O(n)$ links can be afforded, the BHC topology seems to be a viable topological structure for maximizing the number of OD pairs which can communicate concurrently, i.e., on the order of $\frac{n}{\log n}$. (Note: This $O(\frac{n}{\log n})$ least upper bound result for the BHC topology is larger than, for example, the $O(\sqrt{n})$ result associated with the mesh network.)

The BHC topology characterizes networks which are formed by recursively interconnecting existing smaller networks. Fig. 3 shows an example of a 47-node network which is clustered as a 3-level BHC topology. The clusters have been circled for clarity. A general k-level BHC topology is created by interconnecting an $O(1)$ collection of $(k-1)$-level BHC topologies so as to create a connected graph. By recursively applying this rule, a k-level BHC topology can be constructed for any $k \geq 1$. Note: It turns out that $k = O(\log n)$.

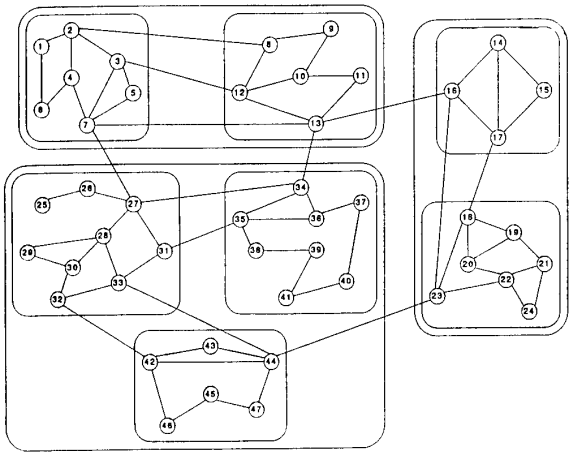


Fig. 3 An example of a 3-level BHC topology.

In using the BHC topology as a structural target for topology design, the following main issues will be addressed:

1. Reliability Analysis: Derive general connectivity bounds for the BHC topology. It seems plausible that (under realistic conditions) the BHC topology is at least 3-connected.

2. The Clustering Problem: Even though the theoretical least upper bound for the size of CCOD pair sets (for the BHC topology) is $O(\frac{n}{\log n})$, it is nevertheless an *upper* bound. Therefore, work needs to be done in determining conditions for which this upper bound is (and is not) met. An obvious situation that might prevent the size of CCOD pair sets from achieving this bound is the case where there is too much inter-cluster traffic. For instance, consider a situation in which the majority of the traffic originates at nodes in a particular high level cluster, and is destined for nodes in a neighboring high level cluster. Since there are only a few data links connecting these neighboring clusters (by definition of a cluster) a bottleneck can result which may severely limit the amount of concurrent communication. A possible remedy to this dilemma is to re-clusterize the network so that those OD pairs which communicate often and/or which send vast amounts of data are in the same (or nearby) low-level cluster. A simple clustering scheme based on sorting the expected OD traffic demands may be viable approach. This implies that some OD pairs which are not in the same *geographical* low-level cluster may need to belong to the same *logical* low-level cluster.

VI. Summary

A graph-theoretic result is presented which shows that in order to accommodate more concurrently communicating OD pairs, one must generically decrease the diameter of the network. Simulation studies done using several topologically distinct networks justify this theoretical result. Finally, the balanced hierarchically clustered (BHC) topology is proposed as a viable structural target for topology design.

Acknowledgment

The author wishes to thank Mr. Nelson Ge for his help in preparing the numerical simulations.

References

[1] J. K. Antonio, G. M. Huang, and W. K. Tsai, "A fast distributed shortest path algorithm for a class of hierarchically clustered data networks," submitted to the *IEEE Trans. on Computers*, 1988. A condensed version in the *Proceedings of the 1989 INFOCOM*, April 1989, Ottawa, Ontario.

[2] W. K. Tsai, G. M. Huang, J. K. Antonio, and W. T. Tsai, "Distributed iterative aggregation algorithms for box-constrained minimization problems and optimal routing in data networks," *IEEE Trans. Automatic Control*, vol. 34, pp. 34–46, Jan. 1989.

[3] J. K. Antonio, G. M. Huang, and W. K. Tsai, "Asymptotic time complexity of the path formulated gradient projection algorithm," to appear in the *Proceedings of the 1990 ACC*, May 23–25, 1990.

[4] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[5] D. P. Bertsekas, B. Gendron and and W. K. Tsai, "Implementation of an optimal multicommodity network flow algorithm based on gradient projection and a path flow formulation," Mass. Inst. Technol., Cambridge, MA, LIDS Rep., p. 1364, 1984.

[6] D. P. Bertsekas, "Optimal routing and flow control methods for communication networks," in *Analysis and Optimization of Systems*, A. Bensoussan and J.L. Lions, Eds. New York: Springer-Verlag, pp. 615–643, 1982.

[7] M. Gerla and L. Kleinrock, "On the topological design of distributed computer networks," *IEEE Trans. Comm*, vol COM-25, pp. 48–60, Jan. 1977.

[8] R. Hinden, J. Haverty, and A. Sheltzer, "The DARPA internet: Interconnecting heterogeneous computer networks with gateways," *IEEE Computer Magazine*, vol. 16, pp. 38–48, Sept. 1983.