A Guide to Literature on Chaotic Routing

Chaotic Routing Project
Department of Computer Science and Engineering
University of Washington, Seattle, WA 98195

The Chaotic Routing Project at the University of Washington encompasses a wide range of research pertaining to multicomputer communication networks. Most research is restricted to routing and flow control topics, but other issues, such as network topology and fault tolerance are covered as well. In general, the network model this research is targeted to is a high-speed, point-to-point network connecting the processors in a single large multicomputer. This does not include wide-area or even local-area networks of workstations.

To date, the primary thrust of work on the Chaotic Routing Project has been in the areas of design, simulation, and implementation of the Chaos Router. The design and theory behind the chaotic routing concept was first introduced in [Konstantinidou & Snyder 90] and a proposed architecture for a chaos router was presented in [Konstantinidou & Snyder 91]. Konstantindou's dissertation provides more in-depth details on the proposed design and architecture [Konstantinidou 91]. A more concise version of Konstantinidou's dissertation work can be found in [Konstantinidou & Snyder 94]. The chaotic router design was expanded to include not only hypercube, but also other k-ary d-cubes such as mesh and torus networks in [Bolding & Snyder 91b, Bolding & Snyder 92]. This work also introduced a streamlined virtual cut-through mechanism in the router, and demonstrated superior performance to oblivious routing techniques (more details can also be found in [Bolding 93a]). Expanded simulation studies of mesh, torus, and hypercube networks were presented in [Fulgham & Snyder 93] which explored the performance of several types of non-uniform traffic such as transpose, bit-revesal, complement, hot spot, and others in chaotic and oblivious routing. All of this work is summarized in [Bolding et al. 94]. The VLSI implementation of the Chaos Router Chip was detailed in [Bolding et al. 93] and [Bolding 93a]. Implementation of a high-performance, clock-skew tolerant version of a channel is discussed in [Wille 92]. The fault-tolerant aspects of chaotic routing were presented in [Bolding & Snyder 91a] and [Bolding & Yost 94a]. A reordering packet-router network interface is being designed by Neil McKenzie and is described in [McKenzie et al. 94].

Other work includes several papers on various aspects of networks and network routing. The invention of the Zenith minimal adaptive router is described in [Konstantinidou 90, Konstantinidou 91]. An analysis of another minimal adaptive router is desribed in [Nguyen & Snyder 94], while worst-case permutations for minimal adaptive routing are explored in [Chinn 94]. An analysis of the performance of hypercube and torus networks, when subjected to pinout and bisection bandwidth constraints, is presented in [Bolding & Konstantinidou 92]. An analysis of the harmful effects of unbalanced loading caused by the common virtual-channel deadlock prevention scheme is presented in [Bolding 92]. Various technologies for channel design are compared in [Bolding 93b].

The chaos router simulator, a detailed simulator model which can simulate many different types on routing on any k-ary d-cube network is available for use by other research institutions via anonymous ftp from shrimp.cs.washington.edu in ~ftp/pub/chaos, as are most of the above mentioned documents.

Abstracts of the Chaotic Routing Project documents follow:

- **[Bolding & Yost 94b] (ebn.ps):** We present the Express Broadcast Network (EBN), a network used for quick and reliable broadcast of control messages in multicomputer networks. The EBN can be implemented with a single extra wire per network link and with minimal extra hardware at each routing node. However, it provides very fast broadcast mechanisms that take advantage of all redundancy in the network to deliver messages regardless of faulty network components. We present extensions of the basic network to include multiple-wire, multiple-bit, and bidirectional wire support, as well as describing basic methods of using the EBN for various applications.

- **[Chinn et al. 94] (UW-CSE-94-07-04.PS):** An adaptive routing algorithm is one in which the path a packet takes from its source to its destination may depend on other packets it encounters. Such algorithms potentially avoid network bottlenecks by routing packets around "hot spots." Minimal adaptive routing algorithms have the additional advantage that the path each packet takes is a shortest one.

  For a large class of minimal adaptive routing algorithms, we present an $\Omega(n^2/k^2)$ bound on the worst case time to route a static permutation of packets on an $n \times n$ mesh or torus with nodes that can hold up to $k \geq 1$ packets each. This is the first nontrivial lower bound on adaptive routing algorithms. The argument extends to more general routing problems, such as the *h-h* routing problem. It also extends to a large class of dimension order routing algorithms, yielding an $\Omega(n^2/k)$ time bound.

  To complement these lower bounds, we present two upper bounds. One is an $O(n^2/k)$ time dimension order routing algorithm that matches the lower bound. The other is the first instance of a minimal adaptive routing algorithm that achieves $O(n)$ time with constant sized queues per node. We point out why the latter algorithm is outside the model of our lower bounds.

- **[Chinn 94] (worstperms.ps):** Chaotic routing is a randomized, nonminimal adaptive routing algorithm for multicomputers. An adaptive routing algorithm is one in which the path a packet takes from its source to its destination may depend on other packets it encounters. Such algorithms potentially avoid network bottlenecks by routing packets around "hot spots." Minimal adaptive routing algorithms have the additional advantage that the path each packet takes is a shortest one.

  Chinn, Leighton, and Tompa [Chinn et al. 94] provide a lower bound for permutation routing problems on the $n \times n$ mesh for a large class of deterministic minimal adaptive algorithms. Specifically, they prove that for any such routing algorithm, there exists a permutation that requires $\Omega(n^2/k^2)$ steps to route all the packets in the permutation, where k is the number of packets a node can contain.

  We present experimental results showing the performance of the Chaos router on permutations for which a deterministic minimal adaptive version of the Chaos router performs poorly. The results show that on these worst case permutations, the time the Chaos router takes to deliver all packets in the permutation closely fits a polynomial in n whose degree is 3/2. From these experiments, we conjecture that no practical router for the $n \times n$ mesh can route arbitrary permutations in time proportional to n, even though the mesh topology has the bandwidth to do so.

- **[Nguyen & Snyder 94] (minimal.ps):** Two classes of adaptive routers, minimal and non-minimal, are emerging as possible replacements for the oblivious routers used in current multicomputer networks. In this paper, we compare the simulated performance of three routers, an oblivious, a minimal, and a non-minimal adaptive router, in a two-dimensional packet switching torus network. The non-minimal adaptive router is shown to give the best performance and the oblivious router the worst. Significantly, however, for many traffic patterns, the minimal adaptive router's performance degrades sharply as the network saturates. Based on an analysis made using several visualization tools, we argue that this performance drop results from non-uniformities introduced for deadlock prevention. Furthermore, this analysis has led us to believe that *network balance* is an important performance characteristic that has been largely overlooked by designers of adaptive routing algorithms.

- **[Bolding & Yost 94a] (ftchip.ps):** As interconnection networks grow larger and larger, the need for reliable message delivery in the presence of faults grows as well. Unfortunately, most network routing schemes currently in use do not provide graceful tolerance of even the most common faults. Because routing messages around failed components requires non-minimal routing, it makes sense to examine routers which, by design, allow packets to take non-minimal routes. Such routers provide a basic level of fault-tolerance by allowing messages to be routed around faults, without requiring a priori knowledge of their locations. However, the mechanisms can be slow and clumsy at times. We augment Chaotic routing, a non-minimal adaptive routing scheme, with a limited amount of hardware to support fault detection, identification, and reconfiguration so that the network can automatically reconfigure itself when faults occur. We present a high-level design of these mechanisms, driven by the goal of achieving reasonable reliability without exorbitant cost.

- **[McKenzie et al. 94] (cranium.ps):** Cranium is a processor-network interface for an interconnection network based on adaptive packet routing. Adaptive networks relax the restriction that packet order is preserved; packets may be delivered to their destinations in an arbitrary sequence. Cranium uses two mechanisms: an automatic-receive interface for packet serialization and high performance, and a rendezvous interface for flexibility. To minimize software overhead, Cranium is directly accessible by user-level programs. Protection for user-level message passing is implemented by mapping user-level handles into physical node identifiers and buffer addresses.

- **[Bolding et al. 94] (UW-CSE-94-02-04.PS.Z):** Chaotic routers are randomizing, non-minimal adaptive packet routers designed for use in the communication networks of parallel computers. Chaotic routing is reviewed along with other contemporary network routing approaches, including the state-of-the-art oblivious routers. Each routing approach is evaluated for its effectiveness as a multicomputer message router. The results indicate that the Chaos router is the most effective of known routing methods.

- **[Konstantinidou & Snyder 94] (chaos.ps):** In this paper the Chaos router, a randomizing, nonminimal adaptive packet router is introduced. Adaptive routers allow messages to dynamically select paths, depending on network traffic, and bypass congested nodes. This flexibility contrasts with oblivious packet routers where the path of a packet is statically determined at the source node.

  A key advancement of the Chaos router over previous nonminimal routers is the use of randomization to eliminate the need for livelock protection. This simplifies adaptive routing to be of approximately the same complexity along the critical decision path as an oblivious router. The primary cost is that the Chaos router is probabilistically livelock free rather than being deterministically livelock free, but evidence is presented implying that these are equivalent in practice. The principal advantage is excellent performance for nonuniform traffic patterns.

The Chaos router is described, it is shown to be deadlock free and probabilistically livelock free, and performance results are presented for a variety of work loads.

- **[Bolding 93b] (UW-CSE-93-12-03.PS):** Massively parallel multicomputers require a high-performance interconnection network. The physical channels which connect nodes in the network are the components that most commonly impose restrictions on the amount of data which can be communicated between nodes of the network. Several different schemes for maximizing the bandwidth of physical interconnection channels are examined.

  Conventional technology centers on 5V electrical signalling. Within this paradigm, there are several different methods of maximizing throughput of a channel, given its limited bandwidth. Simplex and duplex channels are examined, as well as arbitration protocols and transmission line pipelining. Non-conventional technologies which achieve higher bandwidth transmission through low-voltage signalling and fiber optic communication are also examined. For the present, it is expected that most networks will continue to use electrical signalling with low-voltage signalling because fiber optic hardware is too expensive. As technology changes, though, it may become more common for networks to be built out of optical or other new technologies.

- **[Bolding et al. 93] (chip.ps):** Chaotic routers are randomizing, non-minimal adaptive packet routers designed for use in the communication networks of parallel computers. Although adaptive routing, and, specifically, chaotic routing, has been shown to be superior to oblivious routing in most cases, the practical application of adaptive routing to multicomputer networks has been difficult to achieve due to the complex nature of adaptive routers. A prototype two-dimensional (mesh and torus) chaotic router chip has been designed and is being fabricated in a 1.2μ CMOS process. The chip exhibits high bandwidth, limited only by the speed of the off-chip drivers, and low input-to-input latency. To achieve this, much attention was given to reducing the critical path complexity of the router. The resulting chip is shown to be as good or better than state-of-the-art oblivious routers in almost all cases.

- **[Bolding 93a] (boldingPhd.ps.Z):** A crucial component of a massively parallel multicomputer is the interconnection network which links all of the nodes of the computer together. This network provides the primary method of communication between the hundreds or thousands of processing nodes and is, thus, critical to the successful operation of the multicomputer.

  Current state-of-the-art interconnection networks use simple, oblivious routing techniques which achieve very good performance when loading is light, but do not perform well in the presence of non-uniform congestion or faults. Chaotic routing, a non-minimal adaptive routing technique, provides a mechanism which takes into account the presence of congestion and faults when choosing a path for a message and can, thus, achieve better performance.

  Chaotic routing is shown to be applicable to all finite-sized networks of bounded degree with bi-directionally connected links. A design for a chaotic router is presented which includes both the features of virtual cut-through routing and the advantages of internal non-blocking buffering. This design allows messages to follow minimum-latency paths when network loading is light, and provides sufficient buffering to ensure high throughput when the network is heavily loaded.

The resulting design is compared with other minimal and non-minimal adaptive designs, as well as with oblivious routing. Chaotic routing is shown to be superior to most other adaptive designs due to its lower design complexity, and simulations are used to compare it with oblivious and deflection routing. The simulations show that, for mesh-connected networks, chaotic routing performs only slightly better than oblivious routing. However, for torus- and hypercube-connected networks, chaotic routing is superior to oblivious and deflection routing, achieving much higher throughput and lower latency.

To demonstrate the feasibility of chaotic routing, a prototype chaos router chip is presented. This chip, fabricated in 1.2μm CMOS, implements a two-dimensional chaos router. The chip, in simulations, operates at 66MHz, with the speed being limited only by the speed at which the 5V CMOS pads can be switched. Thus, the chip should operate at the same clock speed as oblivious routers using the same technology, giving an equivalent bandwidth capability.

- **[Fulgham & Snyder 93] (UW-CSE-93-06-01.PS.Z):** The Chaos router is a nonminimal adaptive packet router that is both deadlock free and probabilistically livelock free [Konstantinidou 91]. Unlike its predecessors [Ngai 89], the Chaos router uses randomization to provide an efficient way to avoid livelock [Konstantinidou & Snyder 90]. The Chaos router has been compared in software simulation studies to the state-of-the-art oblivious routers (i.e. dimension order routers) on the hypercube topology [Konstantinidou & Snyder 91], and the mesh and torus topologies [Bolding & Snyder 92]. The results, based on simulations of two different workloads, one uniform and one non-uniform load, show that in general the Chaos router is superior to the oblivious router. However, two workloads alone are not representative of the typical types of traffic encountered in practice. In this paper five additional non-uniform workloads are simulated. Each was chosen to represent one of the various types of communication encountered in typical programs. The workloads, compared on both the 256 node torus and hypercube topologies, show considerable variety and reveal insight on the effectiveness of Chaotic adaptive routing on non-uniform loads. Data presented includes the saturation point of the networks under the various loads, the throughput as a function of presented load, latency, and total delay of packets through the network. Collectively, this data gives a much clearer and complete characterization of the Chaos and oblivious routers' behavior on non-uniform traffic.

- **[Wille 92] (chaosChannel.ps):** The snychronous channel controller for the chaos router requires highly synchronized clocks and short interconnection delays in order to operate efficiently. A phase-adjusting controller is introduced which can tolerate significant clock skew by self-adjusting its operation. The channel also can operate at speeds independent of the interconnect speed. This allows the channels to be connected across arbitrary distances and without the necessity of complex clock synchronization schemes.

  A design is presented and compared with the design of the snychronous chaos router channel controller. The phase-adjusting controller exhibits higher bandwidth at the expense of greater average latency through the controller.

- **[Bolding 92] (UW-CSE-92-07-07.PS):** A common scheme for preventing deadlock in networks is the virtual channel method of Dally and Seitz [Dally & Seitz 87]. Due to the nature of this scheme, an otherwise completely uniform network will have non-uniformities intoduced into it. The variations have a wide range of effects, ranging from limitations on overall network performance to differences in observed network characteristics from node to node and from message to message.

  p

- **[Bolding & Konstantinidou 92] (hyp-tor.ps) :** The binary hypercube and the family of k-ary, n-cubes have long been used as communication networks in massively parallel machines. In an effort to understand the effect of dimensionality on the network performance, k-ary, n-cubes have been compared under constant bisection bandwidth and constant pin count constraints. We show that these constraints are not independent and when either one is ignored the comparison strongly favors some particular network. We propose a new *cost function*, namely the product of the network's bisection bandwidth and pin count, and compare the networks under equal cost functions. We then give specific examples of a comparison between the binary hypercube and the torus networks under this new model, both in terms of their minimum latency and maximum throughput bounds and in terms of average latency and sustained throughput under varying applied traffic.

- **[Bolding & Snyder 91a] (faultOverview.ps) :** The chaos router is an adaptive nonminimal message router for multicomputers that is simple enough to compete with the fast, oblivious routers now in use in commercial machines. It improves on previous adaptive routers by using randomization, which eliminates the need for complex livelock protection and speeds the router. This randomization, however, greatly complicates the fault detection because there is no worstcase bound on the time required to deliver a message. Distinguishing between lost and very slow messages is difficult.

  A new method of fault detection is presented that applies not only to the chaos router but also to other adaptive routers as well. In addition, solutions to several practical fault diagnosis and recovery problems in the chaos router are presented. The presentation supports the claim that fault tolerance can be incorporated into a practical router without harming performance for the normal, fault-free cases.

- **[Bolding & Snyder 91b, Bolding & Snyder 92] (UW-CSE-91-04-04.PS):** The chaos router is an adaptive nonminimal message router for multicomputers that is simple enough to compete with the fast, oblivious routers now in use in commercial machines. It improves on previous adaptive routers by using randomization, which eliminates the need for complex livelock protection and speeds the router.

  The two-dimensional chaos router is shown to be theoretically sound and physically realizable. Extensive simulation studies compare chaos routing with oblivious and deflection routing in mesh and torus networks. Chaos routing is shown to be competitive for mesh networks and superior for torus networks. This high performance is, perhaps, unexpected for the mesh since there is no finite bound on the delivery time of any message.

- **[Konstantinidou 91] (hardcopy only):** *Multicomputers* such as the Caltech Cosmic Cube and its commercial descendants, employing distributed memory and message passing, manage to avoid the scalability problems of shared-memory multiprocessors, thus becoming a promising approach to massive parallelism.

The efficient routing of messages is one of the fundamental issues in multicomputers. State-of-the-art systems implement oblivious routing, where the path of each message is unique and completely determined by its source and destination addresses. Oblivious routers, although simple to implement, exhibit pour performance under traffic known to create hot-spots and they are fault-intolerant. Adaptive routers, an alternative to obliviousness, select message paths based on the local load characteristics. As a result they are better at diffusing congestion and they are more fault-tolerant.

In this thesis, two novel adaptive routing schemes are introduced, a *minimal* and *nonminimal* one. In the minimal router, the *Zenith* router, messages are always routed closer to their destination and message delivery can be deterministically guaranteed. In the nonminimal router, the *Chaos* router, messages can be occasionally sent further from their destination. The hardware requirements and control complexity of the Chaos router are superior to that of any minimal router. This reduction in hardware stems from the fact that the Chaos router does not need to use priority routing, like other nonminimal routers in order to guarantee message delivery. Rather, randomization in the message derouting process is used to guarantee message delivery with high probability. The performance of the suggested routers is evaluated through simulations and their hardware requirements are considered.

- **[Konstantinidou & Snyder 91]:** *This paper is subsumed by [Konstantinidou & Snyder 94].* The Chaos router is an adaptive, randomized message router for multicomputers. Adaptive routers are superior to oblivious routers, the state-of-the-art, because they can by-pass congestion and faults. Unlike other adaptive routers, however, the Chaos router has reduced the complexity along the critical path of the routing decision by using randomization to eliminate livelock protection. The foundational theory for Chaotic routing, proving that this approach is sound, has been previously developed [Konstantinidou & Snyder 90]. In this paper we present the complete design of the router together with (simulated) performance figures. The result show the the Chaos router is competitive with the simple and fast oblivious routers for random loads and greatly superior for loads with hot spots.

- **[Konstantinidou & Snyder 90]:** *This paper is subsumed by [Konstantinidou & Snyder 94].* We present the chaos router, an asynchronous adaptive router, which under certain circumstances can send messages farther from their destinations. The chaos router greatly simplifies the routing logic by removing the livelock protection of previous schemes. Through an effective use of randomness, whose sources include that due to the adaptively processed load, the natural timing difference of selftimed circuitry and explicitly injected randomization, the chaos router avoids long message routes with high probability. In this paper the router is described, it is argued that the chaos router is deadlock free and probabilistically livelock and starvation free, and simulation results are presented showing that the chaos router performs well.

- **[Konstantinidou 90] (hardcopy only):** Abstract not available.

References

[Bolding & Konstantinidou 92] K. Bolding and S. Konstantinidou. On the comparison of hypercube and torus networks. In *Proc. Int. Conf. on Parallel Processing*, pages 62–66, August 1992.

[Bolding & Snyder 91a] K. Bolding and L. Snyder. Overview of fault handling for the Chaos Router. In *Proceedings of the 1991 IEEE International Workshop on Defect and Fault Tolerance in VLSI Systems*, pages 124–127, November 1991.

[Bolding & Snyder 91b] K. Bolding and L. Snyder. Performance study of two-dimensional chaotic routers. Technical Report CSE-91-04-04, University of Washington, Seattle, WA, April 1991.

[Bolding & Snyder 92] K. Bolding and L. Snyder. Mesh and torus chaotic routing. In *Advanced Research in VLSI and Parallel Systems: Proceedings of the 1992 Brown/MIT Conference*, pages 333–347, March 1992.

[Bolding & Yost 94a] K. Bolding and W. Yost. Design of a router for fault-tolerant networks. In *Proc. Parallel Computer Routing and Communication Workshop*, pages 226–240, May 1994.

[Bolding & Yost 94b] K. Bolding and W. Yost. The express broadcast network: A network for low-latency broadcast of control messages. Submitted for Publication, November 1994.

[Bolding 92] K. Bolding. Non-uniformities introduced by virtual channel deadlock prevention. Technical Report UW-CSE-92-07-07, University of Washington, Seattle, WA, July 1992.

[Bolding 93a] K. Bolding. *Chaotic Routing: Design and Implementation of an Adaptive Multicomputer Network Router.* PhD dissertation, University of Washington, Seattle, WA, July 1993.

[Bolding 93b] K. Bolding. Multicomputer interconnection network channel design. Technical Report UW-CSE-93-12-03, University of Washington, Seattle, WA, December 1993.

[Bolding et al. 93] K. Bolding, S.-C. Cheung, S.-E. Choi, C. Ebeling, S. Hassoun, T. A. Ngo, and R. Wille. The Chaos Router Chip: Design and implementation of an adaptive router. In *Proceedings of the IFIP Conf. on VLSI*, pages 311–320, September 1993.

[Bolding et al. 94] K. Bolding, M. Fulgham, and L. Snyder. The case for chaotic adaptive routing. Technical Report CSE-94-02-04, University of Washington, February 1994.

[Chinn 94] D. Chinn. The performance of adaptive routers on worst case permutations. In *Proc. Parallel Computer Routing and Communication Workshop*, pages 60–71, May 1994.

[Chinn et al. 94] D. D. Chinn, T. Leighton, and M. Tompa. Minimal adaptive routing on the mesh with bounded queue size. Technical Report UW-CSE-94-07-03, University of Washington, July 1994.

[Dally & Seitz 87] W. Dally and C. Seitz. Deadlock-free message routing in multiprocessor interconnection networks. *IEEE Trans. on Computers*, C-36(5):547–553, May 1987.

[Fulgham & Snyder 93] M. L. Fulgham and L. Snyder. Performance of Chaos and oblivious routers under non-uniform traffic. Technical Report CSE-93-06-01, University of Washington, Seattle, WA, June 1993.

[Konstantinidou & Snyder 90] S. Konstantinidou and L. Snyder. The Chaos Router: A practical application of randomization in network routing. In *Proc. Symp. on Parallel Algorithms and Architectures*, pages 21–30, 1990.

[Konstantinidou & Snyder 91] S. Konstantinidou and L. Snyder. Chaos Router: Architecture and performance. In *Proc. Int. Symp. on Computer Architecture*, pages 212–221, May 1991.

[Konstantinidou & Snyder 94] S. Konstantinidou and L. Snyder. The Chaos Router. *IEEE Trans. on Computers*, (To appear) 1994.

[Konstantinidou 90] S. Konstantinidou. Adaptive, minimal routing in hypercubes. In *Proc. 6th MIT Conf. On Advanced Research in VLSI*, pages 139–153, 1990.

[Konstantinidou 91] S. Konstantinidou. *Deterministic and Chaotic Adaptive Routing in Multicomputers*. PhD dissertation, University of Washington, Seattle, WA, May 1991.

[McKenzie et al. 94] N. McKenzie, K. Bolding, C. Ebeling, and L. Snyder. CRANIUM: An interface for message passing on adaptive packet routing networks. In *Proc. Parallel Computer Routing and Communication Workshop*, pages 266–280, May 1994.

[Ngai 89] J. Y. Ngai. *A Framework for Adaptive Routing in Multicomputer Networks*. PhD dissertation, California Institute of Technology, Pasadena, CA, May 1989.

[Nguyen & Snyder 94] T. Nguyen and L. Snyder. Performance of minimal adaptive routers. In *Proc. Parallel Computer Routing and Communication Workshop*, pages 60–71, May 1994.

[Wille 92] R. Wille. A high-speed channel controller for the Chaos Router. Master's thesis, Univeristy of Washington, 1992.