A Multi-Ring Framework for Survivable Group Communications[1]

Jun Wang , William Yurcik

{wangj, byurcik}@ncsa.uiuc.edu
National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign

Point of Contact: Jun Wang
B71 Computing Applications Building, MC-476
605 E. Springfield Avenue
Champaign, IL 61820
(217)244-0644, FAX (217)244-1987

[1]This work is funded in part by a grant from the Office of Naval Research (ONR) under the umbrella of the National Center for Advanced Secure Systems Research (NCASSR).

A Multi-Ring Framework for Survivable Group Communications

Jun Wang, William Yurcik
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
{wangj, byurcik}@ncsa.uiuc.edu

***Abstract*— Virtual ring is a preferable design for reliable and survivable group communication. Different implementations of virtual ring have different advantages and constraints. After studying two specific implementations of virtual ring, Virtual Ring on Embedded Tree and Virtual Ring of Traveling Salesman Tour, we propose a novel ring-based survivable architecture for group communication, called the Multi-Ring Virtual Ring (MVR). MVR is capable of tolerating one link or one node failure and is easy to implement and maintain. Performance is analyzed with respect to end-to-end hop-count delays and extra bandwidth that is needed for backup. Results show that the MVR design is a desirable candidate to provide reliability and survivability for a group communication system.**

I. Introduction

Survivable and secure group communications are critical in military environments, especially for Command and Control (C2) systems. Existing solutions to group communications may use different techniques. One most straightforward solution is to use multiple unicast connections connecting all group members together. For example, a mesh may be formed among all of the members. Another option is to use multicast to cover all of the members. The multicast solution has a salient advantage over the unicast solution in terms of efficiency, because a large amount of duplicated packets are involved if the unicast solution is used.

This work is funded in part by a grant from the Office of Naval Research (ONR) under the umbrella of the National Center for Advanced Secure Systems Research (NCASSR).

Please address all correspondences to Jun Wang and William Yurcik at CAB, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Champaign, IL 61820, phone: (217) 244-0644, fax: (217) 244-1987.

However, if multicast is used to provide a group communication in a C2 system, several special requirements must be met, for example, the multicast solution must be survivable, secure, and reliable. Reliable multicast has been intensively studied. ACK's are commonly used to confirm successful transmissions. To ensure survivability, backup connections have to be reserved for recovery from failures and attacks. For instance, if a multicast tree is used for the group communication, another disjoint tree must be found for backup. We will show later in Section III-A.3 that even to find the appropriate disjoint backup tree to make the tree solution survivable is not a trivial task.

Considering the above special requirements, the existing IP multicast can not be a desired solution for the design of survivable and reliable group communication system. This is because of the following reasons. (1) IP multicast is complex to implement and deploy. Some special protocols, such as the IGMP (Internet Group Management Protocol), have to be deployed in the network before the IP multicast can be used, which is not a common-place service. (2) It is even more complex if survivability, reliability and security are involved. For example, the existing IP multicast uses tree architecture. In order to provide survivability, an appropriate disjoint backup tree must be found and maintained for recovery from failures or attacks, which is not a trivial task. Additionally, for security, the key management, especially with group dynamics, adds more difficulty to tree-based IP multicast.

Our solution to the design of survivable and reliable group communication system is based on application layer multicast overlay. When a multicast overlay is built, tree-based or ring-based topologies are commonly used. Tree-based topologies may have better delay performance. However, it has many limitations if security, survivability and reliability are considered. The reasons are similar to the reasons why IP multicast is not desirable: (1) Traditional ACK feedback based reliability schemes are problematic due to associated flow/congestion control problems. (2) Tree-based dynamic key management schemes are arguably too complex for practical implementation. (3) A disjoint backup tree has to be found and maintained, which is very hard.

On the other hand, although ring-based topologies may have longer delays, it has some significant advantages if security, survivability and reliability are concerned. The ring-based architecture has inherent reliability and fault tolerance. First, no ACK is needed at all – a sender can always tell if its transmission is successful or not because if the transmission is successful, the original packets will be looped back to the sender itself. Second, we do not have to find a disjoint backup ring because the counter-looped ring automatically provides a backup solution, assuming that all of the links are bidirectional. Third, pair-wise key management between adjacent members of a ring overlay may prove to be more practical than current tree-based schemes [1].

Ring-based overlays are promising candidates to provide group communications in C2 systems. However, if the number of members is large then hierarchical or interconnected rings must be considered for scalability. In this paper we focus on different approaches to building ring-based overlays on top of given physical networks, highlighting interconnected rings as the scalable and applicable solution. Specifically we propose a novel ring-based overlay solution for situations where group members are scattered in different network domains and a simple ring [2] is hard to find.

The rest of the paper is organized as follows. In Section II, the system and network model will be presented. In Section III, we will investigate two designs of virtual rings based on two well known ring building schemes. In Section IV, we will present a new survivable virtual ring framework based on multiple local rings. Finally, we will conclude the paper in Section V.

[1]The key management scheme is our future work and is out of the scope of this paper.

[2]Following the well known definition, a *simple ring* or *simple cycle* is a ring/cycle that consists of only distinct links and nodes [1].

II. SYSTEM MODEL

Formally, a network is defined as a strongly connected directed graph $G(V, E)$, where V is the set of nodes (routers in the network) and E is the set of edges (links in the network), with cardinalities $|V|$ and $|E|$, respectively. Links are bidirectional with the same capacity in each direction. A link from node x to y is represented as (x, y). The capacity of link (x, y) is denoted by $c(x, y)$. It is clear that $c(x, y) > 0$ and $c(x, y) = c(y, x)$.

Based on the system model, we further assume that there is a subset of the nodes that form the multicast group, called the *member nodes*. The set of member nodes is denoted by V_m, and the number of member nodes is then $|V_m|$.

A path p from v_1 to v_n is denoted as $p(v_1, v_n)$ (p_{v_1,v_n} for short) and $p(v_1, v_n) = \langle v_1, v_2, \cdots, v_{n-1}, v_n \rangle$. $p(v_1, v_n)$ is *simple* if all nodes from v_1 to v_n are distinct. If v_1 and v_n are the same node, $p(v_1, v_n)$ forms a *ring*. If a ring degrades to a simple path by deleting any one link on it, then the ring is called a *simple* ring.

III. VIRTUAL RING BASED FRAMEWORK FOR GROUP COMMUNICATION

As we have discussed in Section I, there exist several different combinatorial designs to provide multicast overlay in group communications. The designs include the VC Mesh (VCMESH) in ATM, the Multicast Server (MCS), the Shared Tree (ST) and the Virtual Ring (VR) [2], [3]. Different designs have different advantages and constraints. For example, the VCMESH has the lowest end-to-end latency and the highest survivability, but it suffers from low efficiency (high waste of bandwidth) and the "ACK implosion effect" [2] at the same time. The ST enjoys low end-to-end latency, but it suffers from the "ACK implosion effect". And it is hard to provide survivability to ST, too, because finding the appropriate disjoint backup tree is difficult. The VR has longer end-to-end delay, but its ring structure provides a natural way for reliable communications. In fact, in the VR model, no ACK is needed at all because the sender is always able to tell whether packet transmissions are successful or not. The virtual ring architecture can also provide survivability if the ring satisfies certain conditions, which will be investigated next.

If we consider to provide survivability to a group communication system, for example, to tolerate at least one node failure or one link failure, then we need to carefully design the virtual ring architecture. For example, if we use a virtual ring embedded on a tree, then we will have difficulties of finding the appropriate disjoint backup tree.

In this section, we will discuss two implementations of virtual rings for a given network based on two well known schemes, the "Ring on Embedded Tree" (RET) scheme and the "Ring of Traveling Salesman Tour" (RTST) [4]. Their advantages and constraints will be studied. Survivability issues will also be investigated for both RET and RTST schemes. In the next section, we will propose a new ring-based architecture to achieve survivability from single failure or attack.

A. Ring Based on Embedded Tree (RET)

"Ring on Embedded Tree" is the most straightforward approach to providing a virtual ring in a given network. In this approach, we should first find an embedded tree in the given network, which covers all the member nodes in V_m. Note that such an embedded tree can be found within polynomial time. Based on the embedded tree, the virtual ring is formed easily by conducting an Euler Tour (like a Depth-First-Search) on the tree. The example is shown in Figure 1. Sometimes if cost is taken into consideration, then to find the optimal tree becomes NP-hard. Actually, it is easy to see that it is the famous Steiner tree [5] problem.

1) Advantages of using RET: The RET itself is simple to find [4] if cost is not considered. Even if cost is considered (the original problem then becomes the Steiner tree problem that is NP-hard), many existing heuristic algorithms [6], [7], [8], [9], [10] can yield near-optimal results with respect to different cost functions.

2) Disadvantages of using RET: (1) Since RET is based on tree structure, it suffers from even longer end-to-end delay than a simple ring. In a simple ring, a packet travels $|V_m|$ hops before looping back to the source. However, in an RET, the hop count will be $2(|V_m| - 1)$. (2) If we want to guarantee single failure survivability, this design suffers from

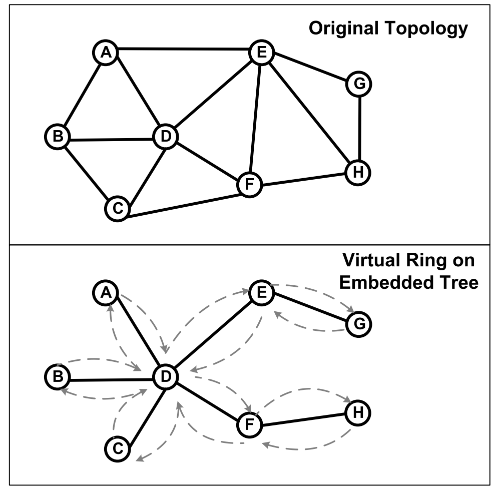

Fig. 1. Example of virtual ring based on embedded tree (RET)

the same difficulties as the tree design, because another disjoint backup tree has to be found, which is hard. We will discuss this in the next subsection. Otherwise, the virtual ring could fail to survive from a single node or link failure. For example, in Figure 1, if node D or link (D, E) fails and there is no disjoint backup tree, the entire multicast group will be partitioned.

3) Survivability of RET: Survivability of RET is not as trivial as it seems to be. Let us first define the connectivity of a network [1], [11].

Definition 1: The connectivity of a network, $\kappa(G)$, is the minimum number of nodes whose removal makes G disconnected.

For example, in a 2-connected network, there must be more than 2 disjoint paths between any pair of nodes. That means, if we design appropriately, a multicast overlay that is survivable from one link or one node failure can be found in a 2-connected network. We will show next, however, this is not the case of RET. That is, RET is not a good design in terms of providing survivability.

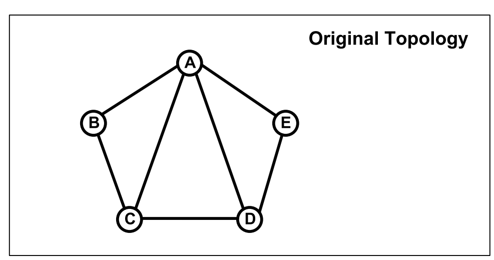

Fig. 2. 2-connectivity cannot guarantee existence of disjoint backup tree for RET

Figure 2 shows an example of a network. It is easy to verify that it is indeed 2-connected. However, it is impossible to find a disjoint backup tree if RET is used. To see this, we first notice that there are totally 7 links in the network. Then, if RET is used, we have to at least find 2 disjoint trees embedded in the network, which requires at least 8 links. Therefore, it is very straightforward to see that we cannot provide survivability if RET is used.

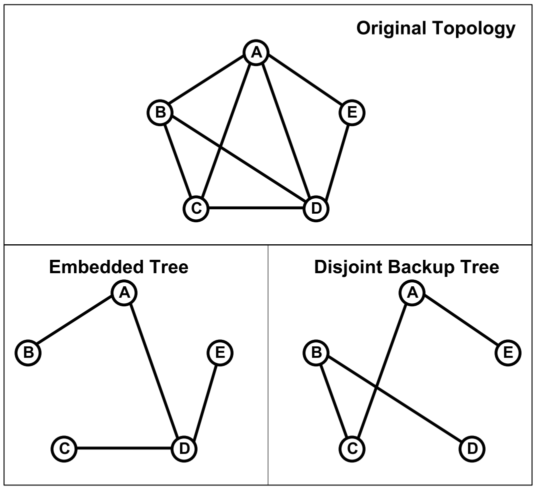

Fig. 3. Disjoint backup tree is not sufficient for survivability of RET

Furthermore, even if two disjoint trees are found in a network, we still cannot guarantee the survivability. An example is shown in Figure 3. Note that although we have found two disjoint trees embedded in the original topology, the network will still be partitioned if Node A is taken down. (If Node A is down, then both embedded trees will be disconnected at the same time.)

B. Ring of Traveling Salesman Tour (RTST)

In this implementation, a least cost (e.g., in terms of hop count) *simple ring* is formed by solving the Traveling Salesman Tour problem. An example is shown in Figure 4 using the same network topology as in the previous example of RET. Note that the Traveling Salesman Tour problem itself is an NP-hard problem. Even if the cost is not concerned, the feasibility problem – to find any feasible simple ring that covers all the member nodes in a given network – remains NP-hard. Essentially, the feasibility problem is similar to the Hamiltonian Cycle problem.

1) Advantages of using RTST: (1) This design is ideal in terms of cost and end-to-end delay. Compared with the delay of $2(|V_m| - 1)$ in the previous RET design, this ring has only $|V_m|$ delay in terms of hop count. (2) It automatically tolerates

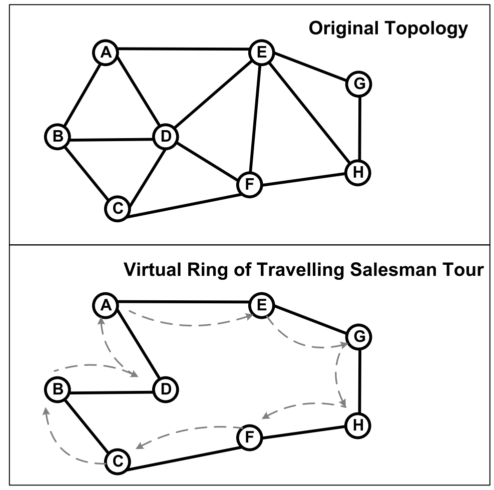

Fig. 4. Example of virtual ring based on traveling salesman tour (RTST)

one node or one link failure and thus provides survivability (just like the dual ring architecture in an FDDI network).

2) Disadvantages of using RTST: It is difficult to find such a ring for a given network. In fact, the Traveling Salesman Tour itself is a well-known NP-hard problem [5]. Such tours may not even exist in some real networks. The problem remains NP-hard even when the least cost requirement is lifted. That is, to find a simple ring that covers all member nodes in a network is already NP-hard, which is similar to the Hamiltonian cycle problem.

3) Survivability of RTST: The survivability from one node or one link failure is inherent in RTST itself, as long as the RTST simple ring is found. This is because there always exist two disjoint paths between any two nodes on the simple ring. However, to find any simple ring covering all member nodes for a given network is NP-hard [12], not to mention the optimal cost simple ring.

Another important issue about survivability is the feasibility of finding the solution. As we have seen, the feasibility of RTST is essentially the Hamiltonian cycle problem that remains NP-hard. The search of a good sufficient condition for the existence of such simple rings still continues to be a research problem [1], [11].

IV. New Multi-Ring Framework: MVR

In this section, we propose a novel ring-based architecture to achieve single-failure survivability, which is called the Multi-Ring Virtual Ring (MVR). The idea is, to find one (optimal) simple ring for the entire multicast group is hard, but we can come up with a near-optimal non-simple ring more easily, especially the group members are scattered in different subnets in the Internet.

A. Algorithm of finding MVR

We can first use some search algorithm to form multiple local simple rings. These rings could be disjoint to each other. Then, we find "bridges" to connect these local rings together into a non-simple ring. Finally, we find "backup bridges" among the local rings, making the MVR scheme survivable in the case of one link or one node failure. More specifically, MVR requires local ring search and so called ear-composition [13], which is much easier than finding one single simple ring for the entire group. The bridges can be found by using Dijkstra's algorithm.

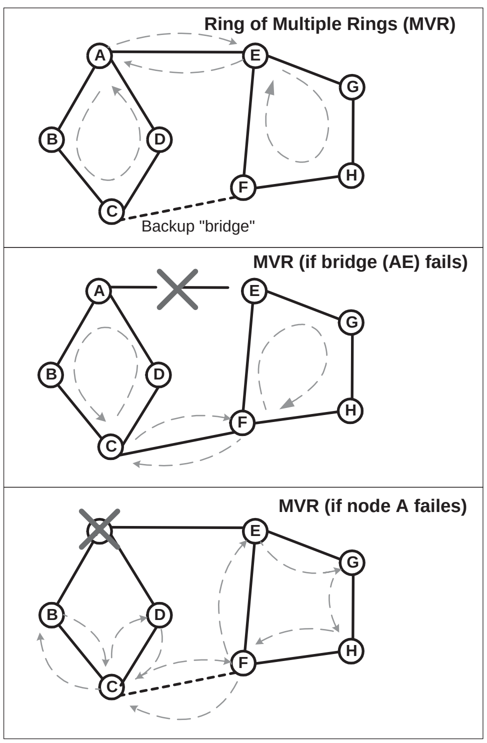

Fig. 5. Design of Multi-Ring Virtual Ring and its survivability

Using the previous example topology, Figure 5 illustrates the basic idea of the MVR design, as well as its survivability from single node or single link failure. In the example, the original MVR is formed as $\langle A, B, C, D, A, E, G, H, F, E, A\rangle$. If the primary bridge (A, E) is broken, because we assume that $c(a, b) = c(b, a)$, for $\forall(a, b) \in E$, the MVR can automatically degrade into the virtual ring of $\langle C, D, A, B, C, F, E, G, H, F, C\rangle$ by using the backup bridge (C, F). If node A fails, then the MVR degrades into the virtual ring of $\langle B, C, D, C, F, E, G, H, F, C, B\rangle$. The example shows clearly that the MVR design is single-failure survivable even without a fully disjoint backup ring/tree.

B. Survivability of MVR and Implementation Issues

1) Survivability: Unlike the RET scheme, the survivability of MVR is inherent. The condition for the existence of an MVR in a given network is simpler than that of the RTST. We can see that as long as the given network is 2-connected, then we can always find embedded MVR's on it.

2) Implementation Issues: Although to find simple rings is basically NP-hard, it is easier to find MVR because searching for local simple rings is easier due to much smaller sizes of local rings. Actually, the simplicity requirement for each local ring in MVR can even be relaxed by using some techniques such as the so called ear-composition [13]. In this way, MVR can be even easier to build while survivability is still guaranteed.

In real implementation of MVR, some issues need to be addressed. For example, if multiple local rings are involved, in order to forward packets correctly at each node in the presence of a failure, we need to let every node know where the failure is, and to activate the correct backup bridge so that the entire overlay will survive from the failure. To address this issue, we first use "heart-beating" messages between each pair of nodes to monitor the possible link/node failure. Then, we can use some bits in each packet header [3] to identify the location of the failure if it happens. For instance, if a failure happens in a certain local ring, then the node right before the failure marks the corresponding bit in every packet header and forward it onto the backup path. Then, every node in the overlay will know the location of the failure by checking the marked bit in the packet headers, and will forward the packets accordingly. In this way, the right backup bridge will be activated and the entire overlay will survive from the failure.

C. Comparisons and Analysis

Assuming we find k disjoint local rings in the MVR and they are connected by $(k - 1)$

[3]Since we are on the application layer, this is not difficult to implement.

TABLE I

ASYMPTOTIC ANALYSIS OF DIFFERENT VIRTUAL RING IMPLEMENTATIONS

Ring Type	End-to-end Hop-count	Extra Bandwidth
RET	$2(\|V_m\| - 1)$	$4(\|V_m\| - 1)b$
RTST	$\|N_a\|$	$2\|V_m\|b$
MVR	$\|N_a\| + 2(k - 1)$	$2\|V_m\|b + 4(k - 1)b$

primary bridges and $(k-1)$ backup bridges, and assuming the bandwidth requirement of the group multicast communication is b, Table I summaries and compares the asymptotic analysis for the three different Virtual Ring designs in terms of end-to-end hop-count delay and extra bandwidth reserved for backup.

The asymptotic results show clearly that, although the MVR can not achieve the optimal end-to-end hop-count delay and the optimal backup bandwidth as the RTST, it is superior to the RET design for both performance metrics. However, as we have discussed before, it is much easier to implement and maintain the MVR than the RTST, especially when the multicast group is scattered into different subnets in the Internet. Therefore, the MVR is a desirable candidate design to provide survivable and secure group communications in a Command and Control system.

V. Conclusion

In this paper, we focused on the design of survivable group communication system. We argued that the existing IP multicast architecture is not applicable especially when survivability and security issues are involved. We proposed the virtual ring based overlays as our solution. To investigate how to form virtual rings in a given network, we first studied two implementations of virtual rings that are based on two well known designs, the virtual ring on embedded trees and the virtual ring of traveling salesman tour. Both advantages and disadvantages were discussed. Survivability issues were also studied. Then, we proposed a new survivable virtual ring architecture based on multiple local virtual rings, called the Multi-Ring Virtual Ring (MVR). The MVR is generally easier to implement and maintain (compared to the scheme using traveling salesman tour). Moreover, asymptotic analysis show that the MVR is also preferable with respect to the smaller end-to-end delay (in terms of hop count) and the less extra bandwidth reserved for survivability (compared to the scheme using embedded trees). The future work includes: (1) Detailed implementation of the MVR; (2) Simulations and performance evaluations of MVR in real networks.

References

[1] T. Corman, C. Leiserson, and R. Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1990.

[2] W. Yurcik and D. Tipper, "A Survivability Framework for Connection-Oriented Group Communications," in *IEEE Pacific Rim International Symposium on Dependable Computing (PRDC)*, Riverside, CA, 2000, pp. 53 – 58.

[3] W. Yurcik and D. Tipper, "Providing Network Survivability To ATM Group Communications Via Self-Healing Survivable Rings," in *7th Intl. Conference on Telecom. Systems*, 1999, pp. 501 – 518.

[4] M. Baldi and Y. Ofek, "Ring versus Tree Embedding for Real-Time Group Multicast," in *Proceedings of IEEE INFOCOM 1999*, New York, USA, March 1999.

[5] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, CA, 1979.

[6] B. Gavish, "Formulations and Algorithms for the Capacitated Minimal Directed Tree Problem," *Journal of the ACM*, vol. 30, pp. 118–132, 1983.

[7] Leslie Hall, "Experience with a Cutting Plane Algorithm for the Capacitated Spanning Tree Problem," *INFORMS Journal on Computing*, vol. 8, no. 3, Summer 1996.

[8] C.H. Papadimitriou, "The complexity of the capacitated tree problem," *Networks*, vol. 8, pp. 217 – 230, 1978.

[9] J. Wang and K. Nahrstedt, "Hop-by-Hop Routing Algorithms For Premium Traffic," *ACM Computer Communication Review (CCR)*, vol. 32, no. 5, pp. 73–88, November 2002.

[10] Jun Wang, Li Xiao, King-Shan Lui, and Klara Nahrstedt, "Bandwidth Sensitive Routing in DiffServ Networks with Heterogeneous Bandwidth Requirements," in *Proceedings of IEEE ICC 2003*, Anchorage, Alaska, May 2003.

[11] Melissa DeLeon, "A Study of Sufficient Conditions for Hamiltonian Cycles," *Undergraduate Math Journal*, vol. 1, no. 1.

[12] A. Sen, B. Hao, B.H. Shen, and G.H. Lin, "Survivable Routing in WDM Networks – Logical Ring in Arbitrary Physical Topology," in *Proceedings of IEEE ICC 2002*, New York, US, June 2002.

[13] Bernard Fortz, *Design of Survivable Networks with Bounded Rings*, Kluwer Academic Publishers, The Netherlands, 2000.