Future Tape Support in the SGI/Cray Environment

Neil Bannister, William E. Kennedy and *Laraine MacKenzie,* Software Development,
Cray Research, Inc., 655F Lone Oak Drive, Eagan, MN 55121, USA

ABSTRACT: Customers find tape support on Cray Research machines a critical software component. For several years, customers have been asking Cray to share those devices across mainframes. With the SGI-Cray merger, an opportunity now exists for GigaRing customers to share tape devices across many mainframes. This talk will describe new goals for future tape support in an SGI-Cray environment.

Introduction

As the amount of data in online storage continues to grow, so does the need to store that data on tape, and the need to share tape data has increased dramatically. To support this demand for tape storage, Cray continues to add support for new devices and has implemented a new architecture for TMF. Cray will protect customers' investments in tape peripherals by supporting those peripherals on the SN1 platform. Cray recognizes that the IRIX platform is highly desirable as storage server and is porting its tape management products to IRIX.

New Features and Status

One of several areas we have focused on for the last year has been the tape SPR backlog. At this point, the SPR backlog is very low and we can now keep it low.

GigaRing tape support was released in three phases: J90, T3E, and T90; these were coordinated with support for the SCSI, ESCON, and BMX GigaRing IO nodes. Now, all previously supported tape devices and loaders are supported over GigaRing for J90se, T90 and T3E.

During 1997, we will release the new Tape Management Facility (TMF) architecture in UNICOS/mk 1.5 and a minor release of UNICOS 9.3 (which will include 10.0). The new architecture is described later in this paper.

Cray Research will be adding support for the IBM ESCON 3590 to UNICOS and UNICOS/mk later this year. This support will include full translation of the 3590 sense bytes. The sense

bytes will be displayed correctly via ERRPT, which will allow for much easier debugging of device problems. The ESCON 3590 will be supported in a minor release of UNICOS 9.0, in UNICOS 9.3 during 3QCY97, and in UNICOS/mk during 4QCY97.

Shared Tape Device Support

For several years, Cray customers have been requesting that we dynamically share tape devices between more than one Cray mainframe. Today, with the use of multi-ported control units and multiple GigaRing ESCON/BMX IO nodes, it is possible (via the use of an ESN/BMN for each mainframe and a multi-port control unit) to share tape devices between GigaRing and Model E IOS mainframes.
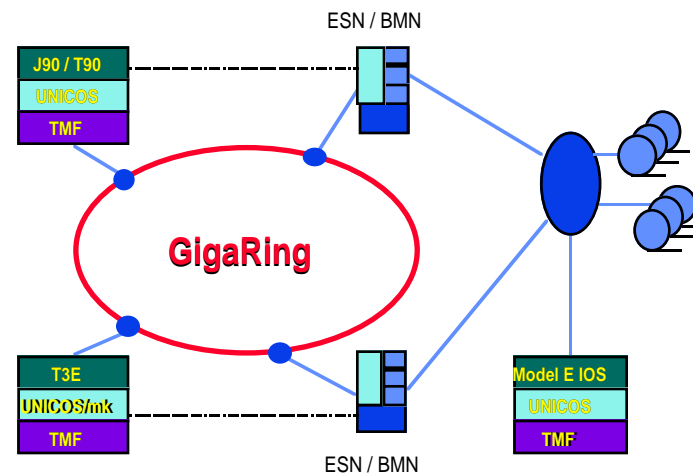
Besides requiring one IO node (of the device type to be shared) per mainframe on the GigaRing, it also requires operator intervention to configure the tape drives up on one mainframe and down on the other. This process is cumbersome and does not support a "lights out" operation. Further, where there can easily be several hosts on the same GigaRing as an IO node (SCSI, BMX, ESCON), it is imperative that we allow dynamic sharing.

In this model, we might propose sharing one ESCON or BMX IO node between multiple GigaRing mainframes on the same ring and a Model E IOS mainframe. However, we believe that there is sufficient hardware available to enable this capability to be supported. This will still require operator intervention, but it would have reduced the number of IO nodes required to one per GigaRing. This sharing capability is not available today and would require IO node and GigaRing mainframe software changes to support it.
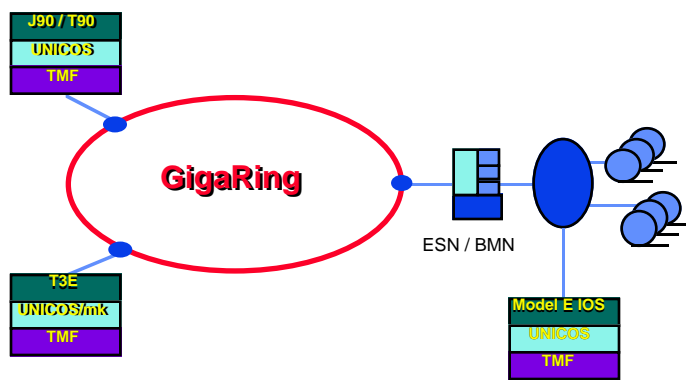

For SCSI GigaRing nodes in the future, we see value in sharing the MPN IO node between multiple mainframes on the same ring and we are prepared to cost this item if there is suitable demand from our customer base.
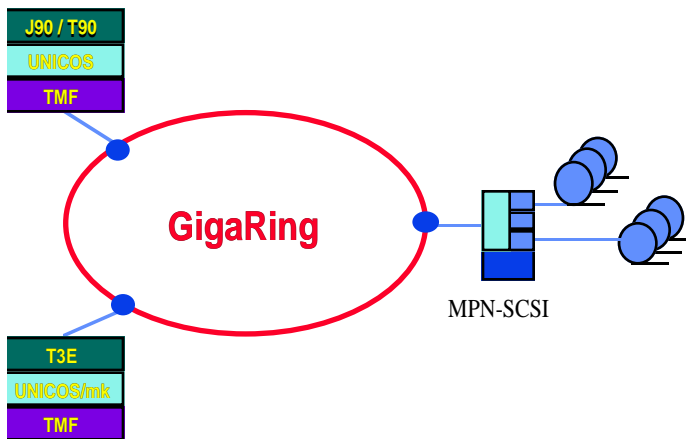
Finally, we want to eliminate the need for operator intervention. Next year we plan to provide a shared tape manager capability to solve this problem. We expect this capability to be provided by the OpenVault product from Mountain View and we intend to interface CRAY TMF to OpenVault to accomplish dynamic tape device sharing. OpenVault will be providing drive allocation, reservation, and deadlock prevention. At this time, it is not clear if OpenVault will run on UNICOS or UNICOS/mk. Because the connection between Cray TMF and OpenVault is a TCP socket, no technical need exists for OpenVault to run on UNICOS.

In summary, we want to move to a tape device sharing model that provides dynamic arbitration of tape devices and eliminates operator intervention. OpenVault will interoperate with these mainframes via a TCP network.
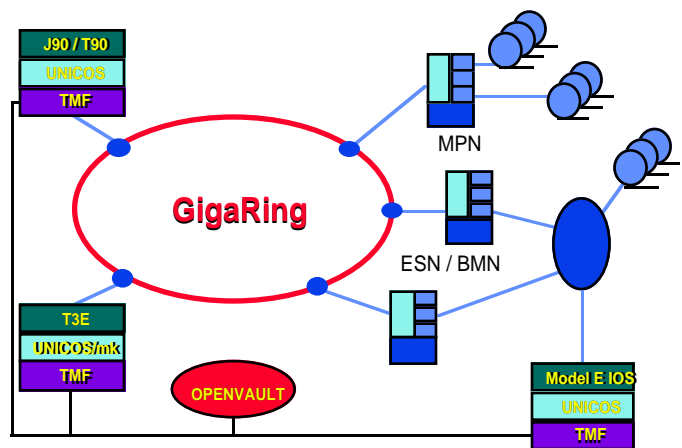
New TMF Architecture

The present TMF architecture fork/execs a separate child process to handle each tape event encountered, such as openfile, read/write volume, tape mark processing, and others. There is also a large context moved between the tape daemon process and each of the child processes. Therefore, TMF uses a great deal of system resources in an operational environment.

The new TMF architecture replaces the fork/exec cycle with a persistent process that handles all tape events. This also means that there is very little context data movement between the parent and child processes. Only state changes are transmitted across this link.

The new TMF architecture provides savings in many ways. There is a reduction in disk space from ~22 Mbytes to ~2 Mbytes. Secondly and more importantly, the user/system time used by TMF was by ~72%. We arrived at the 72% figure by running an internal regression test suite comprised of sixty steps. The user/system time went from 212 seconds (110 user /

102 system) to 60 seconds (40 user / 20 system). In summary, we believe that there will be substantial savings in a busy operational TMF or DMF environment.

Cray Tapes on IRIX

Cray Research is currently porting Cray TMF to IRIX 6.2 Challenge and 6.4 Origin platforms. The goal is to provide the same capabilities on IRIX that Cray customers have on UNICOS. There are some limitations, however, such as features depending upon asynchronous device IO or features that are controlled via limits or the UDB. Our strategy is to support these capabilities as they become available on IRIX.

In porting TMF to IRIX, we have developed a layered driver to sit above the IRIX SCSI driver. This driver provides the linkage between the tape daemon and device driver to handle tape events and make them transparent to the user. Front End Servicing (FES) will be supported. We are also making plans to port Cray REELlibrarian to IRIX. We plan to release TMF and CRL on IRIX platforms during 3Q/4QCY1997.

Device / Loader Direction

Cray Research continues to support its existing SCSI, BMX, and ESCON devices on GigaRing with few exceptions. These supported devices are now available to the J90, T90, and T3E GigaRing mainframes. At this time, the largest blocksize supported for SCSI GigaRing tape devices is 4MB and 256 Kbytes for ESCON and BMX devices. We have also added DLT7000 support for SCSI GigaRing. In third and fourth quarter 1997 we will be adding support for the IBM ESCON 3590 for GigaRing and Model E IOS customers. Cray Research plans to support ESCON and probably BMX devices on the SN1 platform; however, these plans are not final this time.

Summary

We continue to address the support of new devices and keep SPRs low. We are also taking the strong strategy of providing high quality, production-oriented operational tools to the SGI environment. Our goals continue to be "Spin tapes at device speeds" **and** "Spin tapes as well as IBM does for recovery."