CRI Question and Answer Panel

Julie Larson, Cray Research, Inc., Eagan, Minnesota
Panel Members
Tom Boyle ....Customer Service - Technical Support
Steve Johnson ....Engineering
Dave Judd.................Software Division - Compilers
Kevin Matthews........Software Division - Storage Systems
Dave Wallace ....Software Division - Operating Systems

Introduction

Respondents to the CUG Site Survey produced by the Operations SIC are given the opportunity to comment on any hardware, software, or operations issues, problems, or concerns at their site. Several of these questions/comments are addressed formally during the CRI Q&A Panel discussion at the CUG meeting. In addition, the appropriate individuals at Cray Research have provided written responses in this paper.

If your comment or question has not been included, it is because CRI was unable to determine the exact nature of the concern. In some cases site identification is kept confidential, so we were unable to get clarification from you. Please contact your local service representative who can forward your concern to Technical Support for a response.

1 The reliability of hardware on the 512 MPP has been less than expected. The software is not very good either.

Cray Response: The exact reasons for why hardware problems arise are often hard to identify. When hardware problems do arise, CRI Engineering and Product Support will do their very best to address and rectify them in a timely manner consistent with customer wishes. We believe CRI's past responses prove our commitment to rectifying customer concerns.

Suffice it to say that MPP Hardware Engineering is unaware of latent design related failure mechanisms in the CRAY T3D system. If there were such problems they would be addressed and rectified as quickly as possible. We believe the problems that are occurring are random failures with no common thread. It is unfortunate that these occur, but when they do, CRI Engineering and Customer Service do everything possible to put things right.

An analysis was recently performed on available reliability data for large CRAY T3D systems (greater than 300 PEs). The data spans the last several months and in all cases, supports our theory that the T3D-512 hardware problems are random with no common thread and no "magic bullet" cure.

The concern submitted by this customer deals specifically with 512 processor machines. Reliability calculations show that as systems grow in complexity the probability for random failures increases. This, in some way, explains why 512 processor machines are subject to random failures more than smaller machines. We understand that this type of explanation is somewhat unsatisfying and wish that we had the "magic bullet" answer. However, there is no "magic bullet". CRI will continue to address random failures and fix them as quickly as possible. It is our intention to always rectify customer concerns.

In reference to the comment on software, in particular our compilers and products, there are some areas that need improvement. One is compile time and size, with the worst case being array syntax statements in large CRAFT routines. Some steps have been taken to ameliorate these problems, as in the introduction of the "CDIR$ SERIAL_ONLY" and "CDIR$ PARALLEL_ONLY" directives that allow programmers to tell the compiler that a routine will only execute in one particular region. This lets the compiler generate code for only one region instead of both, which is the default for CRAFT routines. There are still compile time and size problems that need to be addressed.

The debugger had a number of problems when it was first released. Although many of the problems have been addressed in recent or upcoming releases, initial bad experiences are remembered, and this can be difficult to overcome, even after these problems are addressed. We must strive to make our first releases more reliable.

There are areas were the programming environment has been quite reliable. The single PE versions of code perform quite reliably. The shared memory library routines have allowed a number of applications to produce good scalable results. In general, the programming environment along with the T3D hardware have demonstrated that developing scalable algorithms is achievable with modest effort. There are even production codes running that scale quite nicely on a large number of processors.

We are looking at where reliability problems have occurred too frequently, and where reliability has been better. We will learn from these experiences, and improve the reliability of our products by improving our development process. Highly reliable software products are an absolute must before our customers can be truly productive.

2 Memory problems take several incidents before they can be isolated and corrected.

Cray Response: The comments in items 2 and 3 are related. The response below is directed toward both comments.

3 Many hardware problems! 29.16 hours of down-time due to double bit errors, single bit errors.

Cray Response: Items 2 and 3 are related to the diagnosis of memory problems and are from CRAY C90 sites. CRI has addressed this issue in past CUG CRI Q&A Panel discussions. This will serve as an update to this concern.

The architecture of the C90 memory, which provides for the very high bandwidth, also makes diagnosing an intermittent memory failure quite difficult. One year ago, CRI released a new diagnostic tool called SMON, which is an acronym for System Monitor. This diagnostic tool has greatly improved our ability to diagnosis not only memory failures, but CPU failures as well. When a hardware failure does occur, SMON is triggered and a dump of all key registers and areas of memory is taken. The data contained in this dump vastly improves the analysis of the failure thus greatly aiding CRI service personnel in determining which hardware component is failing. It has been clearly demonstrated that SMON aids in the analysis of a failure and reduces interrupts when effectively used. SMON however is not an automatic fault isolation tool. It does require that the user of SMON be knowledgeable in the interpretation of the data provided. CRI provides our service personnel with the necessary training and documentation for the effective use of the SMON tool.

4 CRI Maintenance is way too expensive!

Cray Response: Customers are, in general, very pleased with the service they receive on their Cray systems; however, some of them believe that the price they pay for that service on certain system types is too high. In response to these customer concerns, Customer Service has developed a new service program that will be introduced during the fourth quarter of 1995. This program was previewed at an Operations SIC session during the Denver CUG and is described in a paper titled “The New Cray Service Program: an Advance Look” starting on page 222 of the *CUG 1995 Spring Proceedings*. This new service program is aimed at improving our competitive position for both services and service prices. It will offer more flexibility in choosing the level of service needed by our customers and provide competitively priced service options with optional service package enhancements. Customer Service is committed to delivering quality service at a competitive price. We believe that the new service options, coupled with our new product offerings, will meet that commitment.

5 Good support after an event. The problem is far too many events!

Cray Response: The comments in items 5 and 6 are very similar. The response below is directed toward both comments.

6 The on-site maintenance service and National Technical Support is excellent. Very good response to problems after the event locally, nationally and from the USA. But—The frequency of problems on this installation causing loss of user work is far in excess of any reasonable level. We do-not expect this level of hardware and software problems on a production system from Cray!

Cray Response: Comments 5 and 6 say nice things about the support provided by Cray Research Customer Service - thank you for that.

The issue raised is a product reliability one and possibly is associated with fault diagnosis and the ability to achieve a “first time” repair. Because no specific CRI product is mentioned, it is impossible to be explicit or discuss any specific actions we are taking.

In general, Cray continues to strive not only to design and ship the best performing products and make them available at the earliest possible date, but also to build in reliability and maintainability. When reliability or maintainability is seen to fall short, we aggressively attack the root causes.

Engineering Diagnostics - Continually modify and improve our on-line and off-line diagnostic tools to better diagnose a problem.

Engineering Reliability - Diagnose FRU failures to identify the failing part and challenge suppliers or our own Manufacturing group if there is a problem.

Technical Support - Monitor problems on a worldwide basis and identify as early as possible failures that may be common across a product range.

Continuation Engineering - Work any suspected product issues and if a problem is identified, issue a Field Change Order to improve the product’s performance or reliability in the field.

Software Division - Focus on fixing problems that have a significant impact on system stability. Implement changes in the development process to identify and resolve problems before release. Increase focus on compatibility, unit and system testing of software fixes. (See the response to item 8 below for more detail.)

Cray continues to strive to improve the Reliability, Availability, and Serviceability of its products.

7 Dissatisfaction with CRI tape support.

Cray Response: Based on the comment submitted, it is difficult to understand the exact nature of this customer's dissatisfaction. In an effort to cover all the bases, our response will include comments on three different tape support issues. First, we provide information regarding the rate of problems found with the UNICOS tape subsystem and general responsiveness to reported problems. Second, CRI service activities for tape software is summarized. Finally, because a number of Y-MP EL and CRAY J90 sites have given poor ratings to CRI tape support, specific tape issues for this product line is discussed.

Tape Sprs

The number of tape related Software Problem Reports (SPRs) coming in from the field has greatly decreased in recent years. We attribute this reduction in incoming problems to several changes which have come about in the last few years. First, the CRI tape development and support groups make an asserted effort to respond to SPRs well within the time set by our internal SPR response guidelines (reference the paper titled "Cray Research Software Support, Delivering Fixes" in the *CUG 1994 Fall Proceedings.*) Also, CRI personnel in the field and in Eagan are making a conscience effort to increase the level of awareness regarding the customer's perspective on a problem and the impact it has at their site. Improvements have also been made in the area of testing; Our tape testing process is more intense.

Tape Service

Historically, providing service for Cray's on-line tape sub-system has been a challenge and in many ways different from servicing our other products. In addition to our own tape subsystem and driver software, hardware and software from other vendor's must always be taken into account during problem isolation and resolution. It's difficult to discuss all of the general issues and solutions in a paper of this nature, but in summary, we can list the following activities and actions that are ongoing efforts to continuously improve tape service.

- Regular Cray Research Service Bulletin (CRSB) articles.

  Each month, a summary is provided of 3rd party microcode and software levels used on the CRAY systems in the CRI Corporate Computing Network (CCN) computer center.

  Most CRSB issues also contain additional technical articles discussing changes in CRI tape software or other information pertinent to interfacing with specific 3rd party hardware and software.

- Network of Central and Field technical experts

  Field analysts who have worked at sites or in regions where tape use is heavy are a very valuable resource in preventing and diagnosing tape problems for CRI's general customer base. CRI Customer Service takes advantage of this resource by working in many ways to link these individuals together, along with Software Developers and Central Technical Support. Through email discussions, internal meetings, courses, seminars, and informal contact, we have formed these individuals into an effective tape service team where technical information and problem solving strategies are shared on a regular basis.

- UNICOS Field Tests

  For UNICOS 8.0 and UNICOS 9.0 field testing, CRI actively pursued sites who were heavy tape users. While we have a large variety of tape hardware available in CCN, it was recognized that a true production load was needed to completely verify that new tape software was of the highest quality. These field tests went extremely well and as noted above, fewer and fewer tape software problems are being reported.

- Tape Software Training

  A tape workshop has been introduced as a regular course offering and much of the information from this workshop has been incorporated into the *UNICOS Tape Subsystem Administrator's Guide*, publication SG-2307.

CRAY Y-MP EL and CRAY J90 Tapes

The resiliency of the tape subsystem on Y-MP ELs and J90s can be predicted or calculated by taking into account 4 major areas: UNICOS tape subsystem software reliability, IOS software reliability, SCSI controller and device firmware issues and the rate at which we add support for new tape products. Since the UNICOS tape subsystem is common to both the Entry Level Cray systems and mainline machines the discussions above apply.

The IOS software associated with SCSI tape support has not been particularly resilient or flexible when adding support for a new device. To address this, a large restructuring effort was initiated late last year. This new IOS driver was recently introduced on J90s and will soon be available on EL systems. We believe that this effort will pay dividends long term in producing a more stable and supportable tape product offering.

While Cray provides most of the software necessary in this area, the VME SCSI controller vendor and the various tape device vendors provide the firmware for their products. Firmware related problems are fairly common because a change that a tape drive vendor, for example, might make to fix a problem for us or another customer may not only interact with the VME controller but also with other tape drives from other vendors on the same SCSI string. The firmware issues ultimately come down to our ability to work closely with our tape vendors and provide an extensive regression test with multiple configurations before each release of both our software and a new version of firmware.

Finally, we've found that our ability to support our SCSI tape offering at any given point in time (responsiveness to incoming problem reports) is directly proportional to the number of new tape drives we are working to evaluate and add to the supported product list. While this has always been true, we have been asked to add more new tape products in the last 9 months than in the previous 3 years combined. This kind of situation clearly indicates that the developers and support personnel with the proper experience and expertise are not always available. We are currently addressing this issue by training additional support and testing personnel as well as attempting to add to the development-level expertise that is available to work on SCSI tape issues. Additional testing support has also been added during this very busy period.

In conclusion, CRI is very interested in finding out more about the specific causes for any dissatisfaction with our current tape support. We would like to ask that the person who submitted this comment please contact Julie Larson via email to julie@cray.com.

8 Cray needs to supply a X-windows-based batch (NQS) monitor/controller so that operations can point and click to control jobs, queues, limits, etc.

Cray Response: The first release of centralized administration for UNICOS will contain central administration for NQE. This will contain a GUI that will allow management of all the NQE components via a GUI.

9 Software problems, mainly in the MLS area need a very long time to really get fixed!

Cray Response: As of September 14th, 1995, there are 13 problem (non-design) SPRs assigned to the MLS product. The following is a summary of those SPRs:

Severity	Opened by Customer	Opened by Cray	Comment
Critical	0	0	
Urgent	1	0	Opened 16-Mar-94 as minor; changed to urgent on: 25-Jul-95
Major	3	2	Oldest customer SPR is 06-Feb-95
Minor	2	5	Oldest customer SPR is 07-Sep-94
Total	**6**	**7**	

The 6 customer SPRs referenced above were opened by four different sites as follows:

Some problems are more difficult to fix and take longer to address than others. However, we continue to address all reported MLS problems on a continuous basis. Although minor SPRs don’t receive the same priority as SPRs with a higher severity, we continue to address those also.

Site	Number	Comment
A	4	These SPRs were opened as follows: 2 Majors: 13-Jul-95 and 26-Jul-95 2 Minors: 07-Sep-94 and 11-Oct-94
B	1	This site has the Urgent SPR above
C	1	This site has a Major that was opened on: 06-Feb-95
D	1	This site has a Major that was opened on: 18-Apr-95
Total	**7**	**One of the 6 SPRs was submitted by two separate customers.**

Another factor that affects the time to resolve a problem is ongoing development activity in support of new customer requirements.

In summary, given that fixing MLS problems must be balanced against continuing customer requirements and given the above information, we believe that MLS problems are being addressed in a timely manner.

10 CRI’s ability to diagnose and repair software problems in a timely manner is unacceptable!

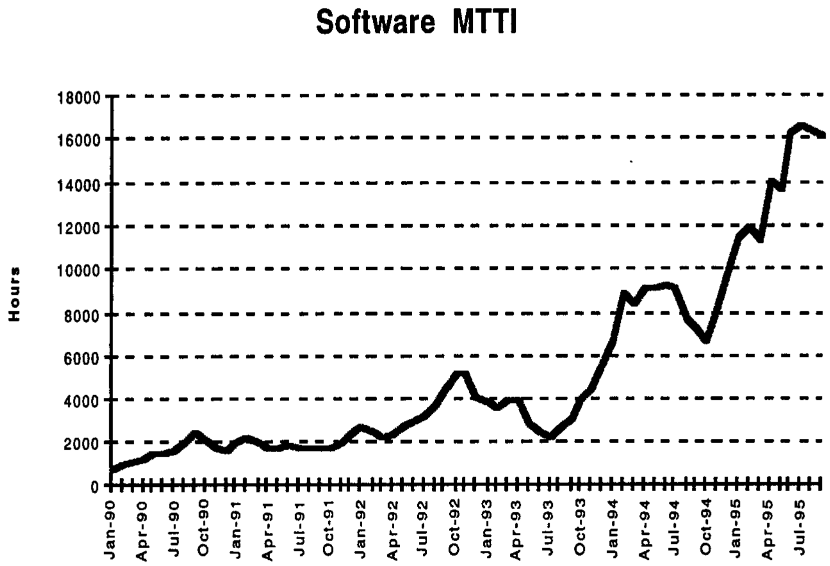
Cray Response: The OS development groups have implemented plans to address the reliability, availability and serviceability of the UNICOS Operating System. This effort was begun prior to the release of UNICOS 6.1.6 and continues today. A number of steps were taken to address the quality of the product including:

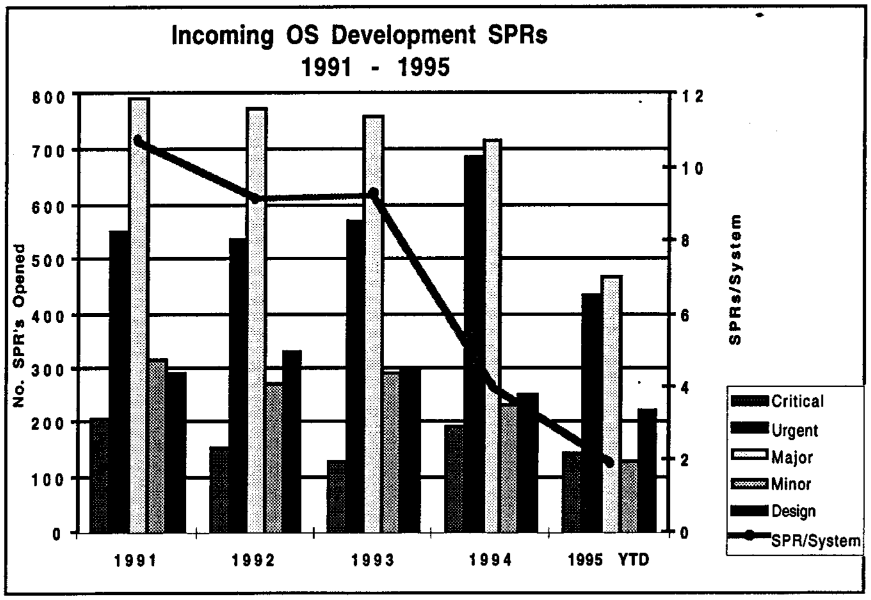
- implementing changes in the development process
- increased focus on design documents and design reviews, code inspections and reviews, unit and global system testing resulting in eliminating problems at the source
- increased control on distribution of individual fixes
- focus on distribution of “tested units” of fixes with emphasis on compatibility, unit and system testing
- focus on problems that have significant impact on system stability

We believe that these steps have been largely successful in improving the quality of the system as measured by Software MTTI and a decrease in the incoming rate of SPRs.

[See the MTTI and Incoming SPR charts on the following page.]

The policies and plans put in place to manage the quality improvement effort has been targeted to affect the greatest number of systems. A consequence of this is that some sites may experience delays in the response to some classification of problems. Our emphasis has been to give highest priority to critical problems, and selected urgent problems. This general policy is carried forward into the support plans for update and revision releases. There are some exception cases that should be reviewed with respect to the general support policies and plans not meeting the needs of these specific customers. We will examine these with the aid of the appropriate account managers and Service Managers to determine a plan of action, if warranted, for these sites.


Software MTTI
Hours
Jan-90
Apr-90
Jul-90
Oct-90
Jan-91
Apr-91
Jul-91
Oct-91
Jan-92
Apr-92
Jul-92
Oct-92
Jan-93
Apr-93
Jul-93
Oct-93
Jan-94
Apr-94
Jul-94
Oct-94
Jan-95
Apr-95
Jul-95


Incoming OS Development SPRs
No. SPR's Opened
SPRs/System
1995 YTD
Critical
Urgent
Major
Minor
Design
SPR/System