April 1989 Volume 1 Number 2

UNIX TECHNOLOGY ADVISOR

OPERATING SYSTEMS

Technology Infusion and Integration in AIX 3

By Dr. Charles Sauer

Executive Summary: AIX™3, the core operating system technology announced in the Open Software Foundation's maiden press conference, incorporates a wide array of new features and technology. Here's a preview of what's in store.

Introduction

AIX 1 was derived from AT&T's System V, with additions from other standard sources, especially Berkeley System Distribution (BSD) 4.2. IBM added unique enhancements in virtual memory support, memory mapped files, dynamic configuration, real time system support, shared libraries and other areas [Loucks and Sauer]. AIX 2 provided major new functions from both public sources and IBM development, including Berkeley internet software, SNA, international character set support, distributed system support, X-windows and extended security features.

A major motivation for AIX 3 development is the significant incorporation of new operating system technology. In many areas this is based on concepts pioneered in other operating systems and prototyping efforts, e.g., in MULTICS or the CP.R project from IBM Research. However, the implementations of these concepts have still required new design efforts to provide evolutionary compatibility with existing components, interfaces and semantics.

The Goals

There are many areas where advances have been sought, including:

- Enhanced support for and compatibility with relevant standards, including ANSI XJ311 (C language), IEEE 1003.1 (POSIX), and NIST FIPS 151.
- "99.44%" BSD 4.3 compatibility.
- Integrations and refinement of Virtual Resource Manager (VRM) functions with the other major components of the AIX kernel, thus producing a single unified kernel.
- Provision of real time capabilities, previously confined to the VRM components, to the higher level components of the system.
- Provision of dynamic binding capabilities, previously confined to the VRM components, to the higher level components of the system, and support of dynamic shared libraries based on this technology.
- Reimplementation of the virtual memory system and the file system to integrate the two areas. In addition, the reimplementation provides major new file system functionality
  - Use of data base memory [Chang and Mergen] for file system meta-data (directories, inodes, and indirect blocks) to provide robustness and reduced need for invocation of separate consistency checking utilities such as *fsck*.
  - Management of disk space as "logical volumes". File systems are composed of a collection of logical volumes, possibly spanning physical volumes. File systems may be enlarged while in use by adding logical volumes. Mirroring of file systems at a logical volume granularity is implemented.

ALSO IN THIS ISSUE . . .

- Continuation of the work to provide security characteristics as defined in the National Computer Security Center's *Trusted Computer Systems Evaluation Criteria*.

- Advancement of distributed system capabilities, including support for internet protocols, implementation of location transparent access to memory mapped files, provision of enhanced, configurable authentication/authorization mechanisms, incorporation of Apollo Computer's object oriented remote procedure call facilities (NCS), and implementation of transparent remote execution facilities.

- Incorporation of advancements in higher level user interface facilities based on the X Window System™, including new toolkits and related program development facilities and applications.

- Provision of integrated system management facilities for both the newly developed components and existing system components.

In addition to these functional characteristics, a major objective of the new designs and the redesign of existing components is portability, so that the new system is supportable on a variety of hardware. There is inherently a tension between function, performance, and portability, so "portability" has been sacrificed in some areas, e.g., some of the virtual memory *implementation*, with the expectation that the virtual memory *interfaces* would be portable.

Given the breadth of the above ambitions, it is impractical to try to describe all of the characteristics and technologies incorporated in AIX 3. The remainder of this article will clarify a few of the above topics. Others, e.g., the integration of the file system with virtual memory and the use of data base memory, have been discussed in other papers. Other topics will be covered in future articles.

Enhanced Support of Industry Standards

Just as there is a tension between advanced function and portability, there is a tension between advanced function and standardization. In the current open systems environment, conformance to standards is mandatory. Advancement of standards, e.g., resolving conflicts between *de facto* standards, may also be considered technological advancement, both because of the enabling of other technologies and because of the inherent difficulties in resolving the conflicts.

ANSI XJ311, IEEE 1003.1, NIST FIPS 151

With the acceptance of IEEE 1003.1, there is an agreed to basic set of programming interfaces for system functions. Similarly, when ANSI XJ311 is finalized,

The UNIX Technology Advisor is published monthly by MYOB, Inc. It is a publication dedicated to the needs of technical managers working in the UNIX/open systems environment. POSTMASTER: please send address changes to The UNIX Technology Advisor, POB 955, Hollis, NH 03049. Subscriptions to the Advisor are available in the United States for US$295 per year. Correspondence on editorial or circulation matters should be sent to POB 955, Hollis, NH 03049. Telephone: (603)465-7825. Letters to The UNIX Technology Advisor become the property of the newsletter, are assumed intended for publication, and may be used by the editors as newsletter matter. Letters should include writer's full name, affiliation and telephone number. Letters may be edited to meet space or clarity requirements. Statements of fact and opinions expressed by columnists and writers are their own, and are not necessarily those of The UNIX Technology Advisor. UNIX is a registered trademark of AT&T; The UNIX Technology Advisor is not affiliated with AT&T. The X Window System is a trademark of the Massachusetts Institute of Technology.

code can be ported to the preemptive environment without requiring the level of locking analysis required in the general case.

Dynamic Binding/Shared Libraries

Until recently, most binding and shared library implementation in UNIX operating systems have been relatively static, e.g., as compared to the dynamic characteristics of MULTICS. AIX 3 is designed to use a new object module format, based on extensions to the COFF format, to enable dynamic binding of unresolved symbols at load time and, if demanded, during program execution.

An exported symbol strategy is used to avoid the potential overhead of resolving arbitrary entry points. System calls themselves are symbols exported by the kernel and dynamically bound at exec time. The mechanisms used are very similar to those previously used and documented for dynamic binding of device drivers and other code in the VRM.

However, with the new design, these mechanisms are available not only for kernel level code, but also for user level code. These mechanisms also enable a fairly dynamic shared library design, allowing for relatively simple replacement of library components.

Logical Volumes

Rather than the rigid partitioning previously used to manage space allocation on disk, AIX 3 is designed to use file systems consisting of an extensible number of "logical volumes". Logical volumes are intended to be on the order of several megabytes of contiguous disk space. A file system may span physical disk boundaries by consisting of logical volumes on different physical volumes. The design also supports mirroring at the logical volume level.

Distributed System Support

Distributed Services 1

Distributed Services (DS) provides distributed system support for AIX, including remote file access with location transparency and distributed inter-process communication. Some of the characteristics of DS 1 include:

- Support for a variety of administrative models, including "single system image", *ad hoc* associations of workstations, and mixtures of administrative models.
- Preservation of "traditional" semantics: In order to do this, machines sharing resources must exchange state information so that their actions are appropriately coordinated.
- Aggressive caching strategies based on fine granularity state information: Note that the performance achievable by aggressive stateful caching may be more important than preservation of traditional semantics in considering tradeoffs between stateless and stateful mechanisms [Sauer/Johnson *et al.*].
- Provision for authorization and authentication appropriate to the administrative model, including appropriate administrative utilities.
- Network transport based on SNA LU 6.2.
- Kernel structured using Sun "vnode" definition. The AIX 2 kernel implements "gnodes", a superset of the Sun "vnode" data structure [Kleiman] to support multiple file system types in the kernel. This allows a clean division between the local AIX filesystem code and the remote filesystem code. Further, it is designed to support multiple local file system types, e.g., for compatibility with exisitng media such as DS and NFS. The AIX 3 gnode design is essentially the same as that in AIX 2.2.1.
- Coexistence with Sun Microsystems' NFS. In addition to gnodes, this includes other issues. The mount model is heavily influenced by the virtual file system mount model of NFS, but includes added features such as file granularity mounts and inherited mounts [Sauer].

Distributed Services 2 Objectives

The objectives of DS 2 include:

- Portability, including portability to operating systems not based on the AIX operating system or the UNIX operating system.
- Extension and generalization of the concepts in DS 1, including the following additions in particular:
  - Support for remote mapped files. (This is designed to be optional in the sense of allowing systems to not support mapped files at all and still be more or less full participants in the network.)
  - Support for configurable authentication and authorization models. These include id translation models such as in DS 1, *Kerberos* based models and others.
  - Support for configurable network transport protocols. It is intended that relatively standardized protocols, e.g., TCP/IP and LU 6.2 be used. For these protocols to be used and still achieve the desired performance, the typical implementation needs to have significant

performance tuning. Such tuning is of general benefit, where undertaken, and systems without such tuning can function in the DS 2 environment, but at lesser performance levels.

- Provision of new kinds of functions not previously addressed in DS 1, including remote procedure call and remote execution facilities.

Summary

AIX 3 development has ambitious goals. It strives for strict upward compatibility with formal and *de facto* standards related to the UNIX operating system. It also strives to incorporate major elements of operating system technology, both old and new, not normally present in UNIX operating system implementations. Finally, it is intended to be a base for further evolution in operating system development, including offerings from the Open Software Foundation.

Dr. Charles Sauer, Senior Technical Staff Member, IBM

References

A. Chang and M.F. Mergen, "801 Storage: Architecture and Programming," *ACM Transactions on Computing Systems 6*, 1 (February 1988)

S.R. Kleiman, "Vnodes: An Architecture for Multiple File System Types in Sun UNIX," *USENIX Conference Proceedings*, Atlanta, June 1986

L.K. Loucks and C.H. Sauer, "Advanced Interactive Executive (AIX) Operating System Overview," *IBM Systems Journal 26*, 4 (1987)

L.K. Loucks and C.H. Sauer, "Saving the Day," *UNIX Review 7*, 1 (January 1989)

C.H. Sauer, "Distributed File System Issues in UNIX Operating Systems", presented at Uniforum-DC, Washington (August 1988)

C.H. Sauer, K.A. Bohrer, T.G. Lang, C. Minshall, G.L. Owens, K. Solem and B.J. Walker, "Convergence of AIX and 4.3 BSD," *UniForum Conference Proceedings*, San Francisco, February 1989

C.H. Sauer, D.W. Johnson, L.K. Loucks, A.A. Shaheen-Gouda, T.A. Smith, "Statelessness and Statefulness in Distributed Services", *UniForum Conference Proceedings*, Dallas, February 1988