NEWISYS System Management

A White Paper

Richard O. Simpson, PhD

NEWISYS, Inc.
10814 Jollyville Road
Austin, Texas 78759

March 2003

Abstract

NEWISYS® System Management capabilities allow total hands-off operation after initial installation, with the goals of keeping mission-critical systems running and reducing total cost of ownership. Dedicated, independent hardware in the form of an embedded Service Processor, memory, sensors, and a LAN connection permits NEWISYS 2100 servers to be operated and monitored remotely. System power consumption levels can be set remotely, statically for now and dynamically based on work load in the future. The Service Processor runs an embedded version of Linux®, and all the System Management functions are developed as standard Linux applications. Future enhancements will add more capabilities, such as partitioning large systems and monitoring many systems at once from a single console.

Why System Management?

With the current trend to consolidate more workload onto fewer servers, it is important that those servers have strong system management capabilities. Just a few servers can become “mission critical” for a business, and downtime cost can be out of proportion to the physical hardware cost. Advance warning of a looming problem can allow orderly switch-over to a backup system before an actual failure. If something does fail, rapid diagnosis and repair are important. System management addresses these areas with environmental monitoring, alert messages, remote control of all server operations, and remote diagnostics.

Even for more mundane, non-mission-critical servers, the presence of system management functions can reduce the total cost of ownership. For a rack-mounted server such as the NEWISYS® 2100, one that’s intended to be installed alongside hundreds of others just like it, cost of ownership rises dramatically if the server requires frequent physical attention in the machine room by the operations staff. If the system is well constructed of reliable components, it should operate without hardware failure for several years. There should be no need to touch the box except for initial installation and eventual decommissioning.

System management overview

The NEWISYS 2100 server includes a small embedded PowerPC™ processor, flash memory, RAM, an Ethernet interface, and a significant amount of software that are dedicated to the System Management function.

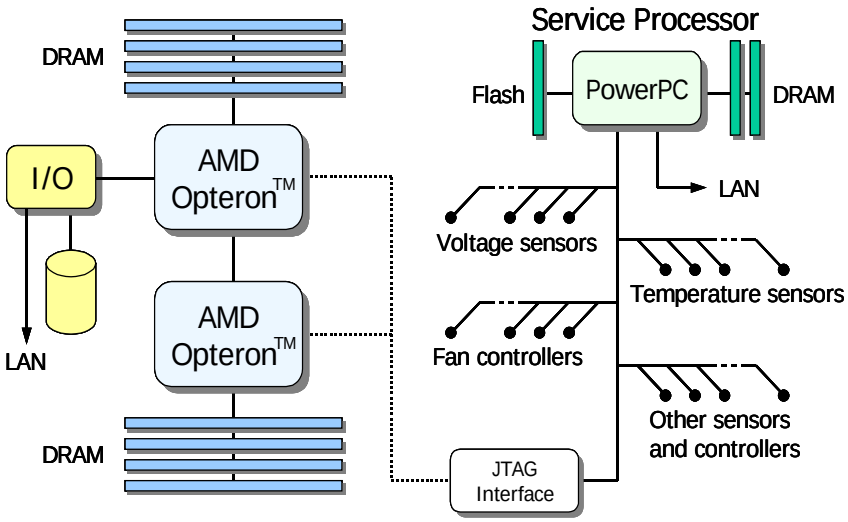
A goal of the independent, dedicated System Management function in the NEWISYS 2100 server is to make it possible to do literally everything required to operate the system over its lifetime from a system administrator’s workstation outside the machine room. It must not be necessary to go into the machine room, locate one server out of hundreds, and hold down a power button for four seconds because the OS refuses to shut down.

To achieve this, System Management must make it possible to do the following by remote control:

- Turn the server’s power on and off. This means that the hardware used for System Management must be independent of the server’s main power supply, so that it is functional even when the AMD Opteron™ processors are turned off.
- Manage the amount of power consumed by the server. At the moment this is a static setting, entered via the System Management console. In the future, with support from the server operating systems, an overall power level can be set for an entire group of servers, with the power level of each individual server being adjusted dynamically based on work load.
- Boot, shutdown, and re-boot the server’s operating system(s). Some of this will happen automatically if System Management is able to turn the power on and

off, but provision must be made for cases where the server's OS "hangs" and is non-responsive.

- Stop the server's boot process in BIOS, to inspect and modify BIOS settings.
- Upgrade the server's BIOS (that is, "flash the BIOS") using files downloaded via the network.
- Install and upgrade the server's OS, and install drivers and application software.
- Upgrade the System Management software itself via the network.



- Monitor the server's physical condition while it is operating: watch for elevated temperatures, incorrect voltages, and failing components.

The NEWISYS® design puts the responsibility for these functions in software running on a small embedded processor dedicated to the System Management task: the Service Processor. Powered by "standby" power from the power supply, the Service Processor is "up" all the time once the server has been plugged in to the AC supply and the physical on/off switch on the power supply has been turned on.

The System Management function occupies a small portion of the server's main circuit board, and consists of the following:

- The Service Processor, a Motorola® MPC855T PowerPC embedded processor.
- Flash memory and DRAM for the Service Processor.
- A Ethernet LAN connection.
- Various sensors for voltages, temperatures, fan speeds, switch closures, and the like, connected to the Service Processor by a simple serial bus.
- A "JTAG" interface to the main Opteron processors that allows the Service Processor to read internal registers and other state from those processors when necessary.

  The Service Processor uses the JTAG interface to read information from the Opteron processors after a machine check, in the manner of a flight data recorder. The state of the Opteron processors can be captured at the point the error occurred, including trace buffers that indicate what was happening just before the crash.

- A number of lights (LEDs). Some are on the front and rear panels but many are on the main circuit board and can be lighted by the Service Processor to indicate the locations of failing components on the board.
- A small LCD display and a set of push-buttons on the front panel that make up the server's physical user interface.

Remote access to the System Management software is via the Service Processor's LAN connection, in two ways:

- A web browser, running on any system that has access to the Service Processor's LAN, provides a graphical user interface. This is the "browser-based console" that is described in detail below.
- A text command-line interface. All the functions that can be done via the GUI (browser-based console) can be also be done by this text interface. It is intended that this interface be used for remote scripting.

Functions provided

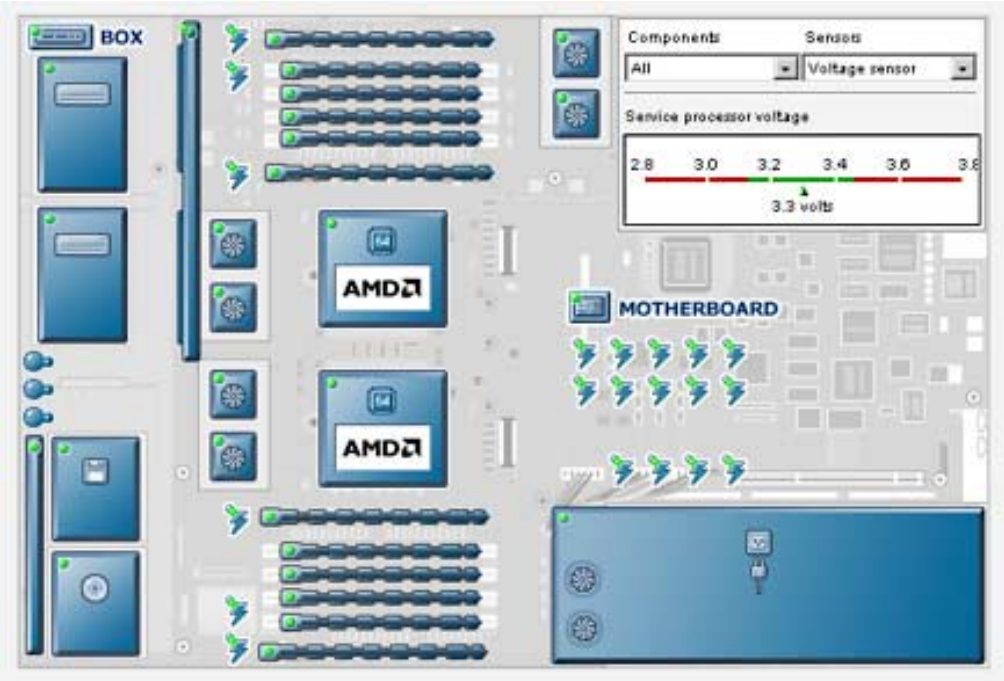
Environmentals

The Service Processor is in charge of monitoring and responding to the server's environment. Its primary responsibility is to prevent physical damage to the hardware (e.g., overheating). Secondary responsibilities include noise abatement by reducing fan speeds when possible and minimizing power consumption.

MONITORING. The Service Processor can monitor nearly two dozen voltages, a dozen temperatures, the operations of eight fans, various switches and operator controls, total system power consumption, and a "heartbeat" from the server operating system.

- Data displays via the browser-based console

  The status of all the monitored items is displayed in summary fashion when a system administrator logs in to the browser-based console. Actual readings of each of the sensors can be displayed in tabular and graphical form.



- Alerts via the browser-based console

  Alerts are messages telling of significant events, such as a temperature measurement rising above a preset warning level. The Service Processor sends alerts to all logged-in browser-based consoles. These alerts are persistent, meaning that they remain outstanding on the console's events screen (and are displayed to those who log in subsequently) until the condition causing the alert is resolved and the alert is then acknowledged by operator action (clicking on the line in the events screen).

- Alerts via SNMP

  The same alerts can be sent to the standard SNMP port at a system administrator-specified IP address. An external management system such as HP OpenView or IBM® Director can process the alert notification in its normal way.

- Power consumption

  The Service Processor monitors a signal from the power supply that provides a direct reading of the total power being consumed by the system. The current level of system management software displays this reading along with the other sensor readings. Future software can use the measured power level and a maximum power level set by the system administrator to modify the server's power consumption. Changing the power level dynamically like this requires support in the server operating system.

CONTROL. The Service Processor can control the speed of six of the eight fans (the two that are integral to the power supply are under the control of its microprocessor) to ensure sufficient cooling for the sensitive electrical components such as the AMD Opteron CPUs and memory. If a monitored temperature approaches an alert level, the Service Processor will increase the fan speeds in an attempt to keep the temperature down. If the fans are already running at maximum speed and a critical high temperature level is reached, the Service Processor will try to shutdown the server operating system in an orderly manner. When the OS has shut down, or if the OS fails to respond, the Service Processor will turn off the server's main power. All these situations result in alerts being sent to the browser-based console(s) and via SNMP to external management systems.

The Service Processor also has limited control over certain operating voltages, and can be directed by field service personnel to move such voltages up or down slightly to assist in problem diagnosis.

The Service Processor has control of the main power supply to the server (can turn it on and off), and can reset and re-boot the server CPUs.

Server CPUs

MONITORING.

- Heartbeat

  A driver program installed in the server operating system communicates periodically with the Service Processor to exchange brief messages meaning “all is well.” If this “heartbeat” signal is not received from the server OS within a specified amount of time, the Service Processor will generate an alert to inform system administrators that the server is not responding. An administrator can then investigate and, if necessary, force the server to reboot.

- Serial console

  The server I/O complex includes a standard serial port, intended for debugging use. While it is possible to connect a serial terminal such as a laptop computer to this port and use the terminal to control the server operating system[1], doing so would require that the terminal be physically close to the server and plugged in to the serial port connector on the server’s back panel.

  Instead, the server’s serial port can be monitored and interacted with remotely using a facility provided by the Service Processor, which makes the traffic on the server’s serial port available via a terminal emulator connected over the LAN. OS debugging can thus be done remotely.

- SNMP

  The SNMP package on the Service Processor makes its status (primarily environmental values such as temperatures) available via SNMP queries, and sends out SNMP alerts. In addition, it acts as a proxy for SNMP on the server, and makes all the server’s SNMP data (the “MIB”) available as well. Thus an external management system can see all the SNMP data for the entire NEWISYS® server via the one SNMP client on the Service Processor.

CONTROL.

- Power On/Off/Cycle

  The server operating system can shut itself down and turn off the power, as it can on most computers. Such an orderly shutdown and power-off can be initiated by the usual means that the server OS provides, and by the Service Processor’s browser-based console. If the shutdown fails (server OS hangs and stops responding), the Service Processor can forcibly turn off the power.

  Since the Service Processor runs on standby power and remains up when the server power is off, it can turn the server power back on and reboot the server OS.

---

1. If the server is running Linux®, the port connects to the standard Linux serial console. If the server is running Microsoft® Windows®, the port connects to the Windows EMS function.

- Boot/Shutdown/Reboot OS

  Independent of whether the power is to be turned off, the Service Processor can shutdown and reboot the server OS.

- BIOS update

  Using a special communications path provided for the purpose, BIOS code running on the Opteron processors communicates with code running on the Service Processor during the server boot process. The Service Processor can direct BIOS to enter its "update" mode, allowing a system administrator to modify BIOS settings or to load a new version of BIOS. This has the effect of pressing the keyboard key that would interrupt BIOS' start-up on an ordinary PC, but without requiring a physical keyboard or that there be a human ready to press such a key.

- BIOS/Service Processor interaction during boot

  BIOS and the Service Processor exchange information about the system configuration as discovered by BIOS and as known by the Service Processor from previous boots. The Service Processor can direct BIOS to omit a known bad component from the discovery list that BIOS passes to the server OS.

  In the future, as servers become larger and partitioning of a server into more than one logical system is implemented, the Service Processor will prevent BIOS from "seeing" components (Opteron processors, memory, busses, and adapters) that, while physically present, logically belong to a different partition.

- Power management

  One of the environmental values that the Service Processor monitors is a signal from the power supply indicating how much power is being consumed. Initially, this value can be displayed via the browser-based console or retrieved via a remote script. The system administrator can interact with BIOS at server boot time to set the server's power level (primarily the Opteron processors' voltage and frequency).

  In the future, as server operating systems gain more power management facilities, the Service Processors on several NEWISYS® systems can coordinate their systems' power consumption as a group. The desired maximum power limit for a rack of systems could be set by a system administrator's command or by a time-of-day event or other event, and the systems within the rack would adjust their power levels to comply.

Standard interfaces, standard Linux components

The Service Processor runs an embedded version of the widely-available open source Linux operating system. Linux, rather than a proprietary embedded system, was chosen for several reasons.

- Linux itself is royalty-free, thus lowering the overall cost of a NEWISYS® system.
- Large amounts of software available for Linux are also free, including the Apache web server and a widely-used SNMP implementation.
- The use of familiar tools and a familiar programming environment simplifies software development. This is true both for our own software development and for customers who might write shell scripts for execution on the Service Processor.
- The cross-platform portability of Linux and programs for Linux allows development and testing on ordinary PCs.
- Remote access to the Service Processor, both by logging in to it and via remote command execution (SSH) are built in.

SNMP

The standard UCD version of SNMP for Linux is used. A MIB is defined for the Service Processor, and the MIB data from both the Service Processor and the Opteron processors is made available through the Service Processor's IP address.

SSH

The standard Linux version of SSH (the "secure shell") is used for authenticated, encrypted remote access to the Service Processor. The traditional Unix and Linux means of remote access, telnet and rsh, are not used because they lack security.

SSH provides two ways in which the Service Processor may be accessed:

- Interactively, as a remote login mechanism. This is the telnet equivalent; from a remote system, it presents to a system administrator a command-line interface to an interactive shell running on the Service Processor.
- Remote command execution. This is the rsh equivalent; a single command entered on a remote system is executed on the Service Processor. A shell script running on such a remote system can cause commands to be executed on an entire list of NEWISYS servers by issuing "ssh" to each.

NFS and Samba

These standard Linux components provide the Service Processor with access to a file system on a remote server. NFS allows the SP to access a Unix or Linux file system; Samba allows it to access a Windows (CIFS) file system. The customer chooses which one, if any, to use. Either provides a place to store event logs, NEWISYS diagnostics, NEWISYS BIOS and Service Processor code updates, and customer-written shell scripts. Except for execution of diagnostics, provision of an external file system via NFS or Samba is not required. The Service Processor is able to run entirely from code stored in its on-board flash.

NIS, Active Directory

To log in to a NEWISYS® browser-based console, a userid must be defined on the Service Processor and assigned to one of four groups defined for NEWISYS System Management. Membership in a particular group defines the level of permissions granted to a user: some users can only observe, while others can observe and take actions.

When installing many servers, it would not be cost effective to require the system administrator to define each user individually on each system. The Service Processor software can be directed to use an enterprise's existing authentication server, either NIS (for Unix/Linux complexes) or Active Directory (for Windows complexes). The enterprise's existing group names can be "mapped" to the four NEWISYS group names.

Web server

A web (or HTTP) server provides the primary means of access to the Service Processor, via the browser-based console. We use the standard Linux version of the Apache 2.0 web server. Apache supports encryption via SSL, the "secure sockets library", so that management traffic (screen contents and data from the keyboard) is encrypted. The web server provides the execution environment for CGI programs, the means by which the browser-based console causes most user-directed actions to happen on the Service Processor.

Web browser

This isn't software on the Service Processor, but software on the system administrator's computer. An ordinary web browser such as Microsoft® Internet Explorer, Netscape® Navigator, or Mozilla is all that's needed to access the browser-based console. The only requirement is that the browser must support Java™ and SSL, as virtually all modern browsers do. No additional software need be installed, meaning that a system administrator can access any NEWISYS system's browser-based console from any system, anywhere, as long as a network connection can be established.

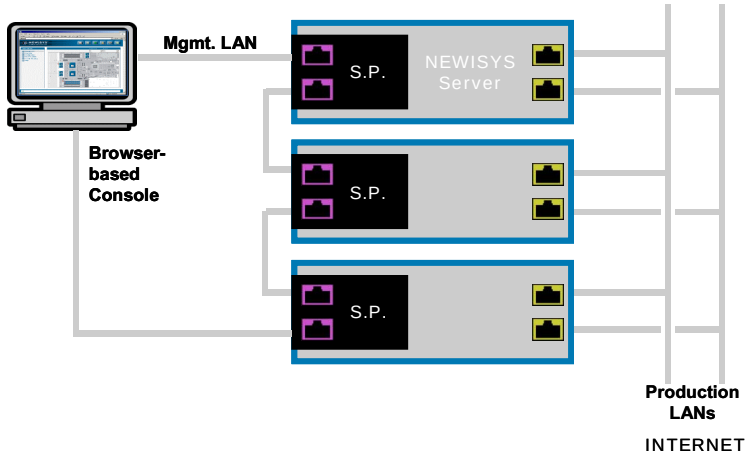
No "state" is saved in the browser — nothing remains on a system that has been used to access a browser-based console once the connection is terminated, so nothing can be observed by a subsequent user about userids, passwords, or system management actions taken.

The browser-based console's graphical user interface is built entirely from standard components:

- HTML is used for the simpler pages.
- JavaScript™ is used to produce and process "forms" on the browser-based console, such as the initial login screen.
- Java applets provide much of the functionality of the browser-based console's user interface, particularly the processing of alert events and the graphical display of the system's status.

Scripting

Each function provided by the browser-based console can also be invoked from the Linux command line. While a system administrator could actually log into to the Service Processor via SSH and issue such a command, it is more likely that the commands would be part of a shell script invoked remotely. A central management system could cause many systems to power on and boot at a given time or because a certain condition has occurred. Likewise, an updated version of BIOS could be loaded into the flash memory on dozens of systems at once via a shell script that queries each system's software inventory for the BIOS revision number, boots any out-of-date systems into BIOS update mode, and re-flashes the BIOS over the LAN.



Security

Separate Ethernet LAN connections are provided for the Service Processor and the server. These can be bridged together or kept completely isolated, at the customer's option.

All access to the Service Processor is encrypted: SSL communications via the browser-based console, SSH for secure remote login and remote command execution. A customer-provided authentication server can be used to validate access to the Service Processor, thus allowing a customer's existing system administrator login IDs to access the Service Processor.

Diagnostics

NEWISYS® supplies a set of hardware diagnostic routines that run on the AMD Opteron processors and on the Service Processor itself. The browser-based console provides a means of invoking the diagnostics and reviewing their results. Running the server diagnostics requires shutting down the server OS, if it is running, so requesting diagnostic execution via the console will perform an orderly shutdown of the server OS followed by a reboot of the server into the diagnostic routines.

The server diagnostics run as Linux applications on a special version of the Linux kernel built for the AMD Opteron processors. The diagnostics are too large to be packaged in the Service Processor's flash memory; instead, the diagnostics are loaded by the Service Processor from an external file system via NFS or Samba.

The System Management scripting facilities allow a customer to write a script that, at some appropriate off-peak time, will shut down the server OS, reboot into

diagnostics, run a selection of diagnostic routines, record the results in a log for later analysis, and reboot the server OS.

Through the browser-based console, or through commands issued remotely to the Service Processor, the diagnostic functions can all be invoked by remote control. It is not necessary to physically access the system being diagnosed unless a particular test requires interaction (insertion of a CD-ROM, for example).

Staged release environment

The programs that implement the System Management functions reside in flash memory attached to the Service Processor, and on a remote (network) file system made available to the Service Processor by NFS or Samba. Both can be replaced by newer versions as needed. The flash memory can be updated by a command to the Service Processor, once the proper flash image file has been copied from the NEWISYS® web site. Likewise, the programs on the remote file system can be replaced by newer versions by merely copying them over the network.

What this means is that the System Management functions can be enhanced over time and error corrections can be incorporated into new download images. New functions can be made available for existing systems; it isn't necessary to tie System Management software releases strictly to hardware releases. More complex functions can be made available as their programs are completed and tested.

What if the Service Processor isn't functional?

The hardware is designed to be fail-safe with respect to the Service Processor. If the Service Processor doesn't respond when the front panel power-on button is pressed, a hardware sequencer will take over and boot the server normally. What's missing in this case are the Service Processor's management functions: there will be no monitoring, no SNMP alerts, no response to external commands such as remote power on. The hardware will run the fans at full speed all the time for maximum cooling.

A heartbeat is exchanged between the Service Processor and the hardware sequencer during normal operations, so that if the Service Processor fails after power-on the hardware will take over control of the fans and the front panel power button.

Future directions

NEWISYS' initial product is a one- or two-processor 1U (1.75 inches high) server, an entry-level system. As the product line expands, so will the System Management functions:

- In the initial offering, each system stands alone: a system administrator can connect to the web server on a given Service Processor and from there monitor and control the system, but there is no means to control a group of systems jointly. A more powerful browser-based console that communicates with a number of Service Processors at once will permit a system administrator eas-

ily to monitor an entire group of machines and to perform actions that affect some or all of them at once.

- As the servers get bigger (more processors, memory, and I/O), it will be possible to partition a single server into two or more logical servers, with each of the logical servers potentially running different operating systems (Linux and Windows on the same box, for example). Partitioned servers can operate together as a cluster, or completely independently. The box still will have a single Service Processor, though, and the System Management code will need to be enhanced to deal with several "server sides" at once.
- As noted above, a group of Service Processors working in concert can adjust the power consumed by their servers, in order to keep the total power for the group of servers under a specified limit. This function will require operating system enhancements to be fully effective.
- The environmental data captured by the System Management functions can be analyzed over time in order to predict certain failures before they occur, allowing timely replacement before a problem takes down a machine.

Conclusions

The NEWISYS® 2100 server is supplied with an independent System Management processor and supporting hardware. The System Management goals are to facilitate hands-off monitoring and management of mission-critical servers and to reduce the total cost of ownership of NEWISYS systems. It does this by permitting truly hands-off operation after initial installation, including environmental monitoring and remote control of operations that would normally require physical access to the server. As larger servers are introduced, remote System Management will play an even larger role in hands-off operation, reconfiguration, power consumption, and failure prediction.

About NEWISYS®

NEWISYS, Inc., a technology company dramatically changing the server-computing environment, is dedicated to adding enterprise level systems management and modular scalability into the enterprise server market through development of robust server designs. NEWISYS offers a family of products targeted for integration into OEM and system builder server product families. Founded in August of 2000, NEWISYS is headquartered in Austin, Texas. For more information, visit www.newisys.com.