Static Performance Tuning

By Richard Elling - Enterprise Engineer

Sun BluePrints™ OnLine - May 2000

http://www.sun.com/blueprints

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303 USA
650 960-1300 fax 650 969-9131

Part No.: 806-5525-10
Revision 01, May 2000

Static Performance Tuning

Introduction

Dynamic performance tuning is well documented in books, the trade press, and online. However, there is a class of problems which could affect system performance which is not dynamic by nature and cannot be detected by conventional dynamic tuning tools. The problems discussed in this Sun BluePrints™ OnLine article are of a static nature. This Sun BluePrints article provides a limited set of solutions, and describes ongoing research into tools being developed at Sun Microsystems, Inc. to improve system performance.

Static and Dynamic Performance Problems

Dynamic performance tuning techniques look at live system performance data and make recommendations based on the way that the systems react under stress. Complex systems become increasingly difficult to understand when they are overloaded. Benchmarks are often used to provide baseline performance data and to examine how variations in system components affect performance.

Static performance tuning is complementary to dynamic performance tuning. Static performance tuning looks to solve problems which are known to affect performance, but are not easily detected by dynamic system analysis. The goal is to detect the static configuration problems which impede performance. Such problems are often reported to system administrators as performance problems by end users, however, these do not have any basis in dynamic system resource attributes such as CPU usage, network bandwidth utilization, disk utilization, etc. Since the problems are static, the analysis also can be static — the server does not have to be under load to perform a static performance analysis.

Another way to describe the difference between dynamic and static performance tuning is by using the analogy of tuning an automobile engine. Dynamic performance tuning can be likened to the adjusting of the carburetor, whereas static performance tuning can be likened to checking if the tires are inflated correctly.

Network Performance Tuning

With increased reliance on networking, it's not surprising that network performance issues are becoming important to overall system performance. There are many aspects of networking that are static by nature, which can lead to poor performance on the client side.

Static performance analysis examines problems which cause delays in communication. However, the problems are not due to bandwidth utilization, but rather as the result of changes over time and configuration issues. Fortunately, they are relatively easy to analyze.

Name Service

Name service problems manifest themselves in many different ways. Symptoms can include:

- Slow login.
- Slow connections for services such as mail, http, ftp, telnet, and databases.
- Slow X-window system performance.
- Web browser hanging.
- New window hangs at startup.
- Common Desktop Environment (CDE) login hanging.

Name Service Problems on Name Servers

An efficient name service requires that the name servers are correctly configured, which in some cases requires a detailed analysis. Fortunately, the analysis is generally narrow in scope, as typical enterprises have few name servers.

Flat name services such as NIS have few issues which relate directly to static performance. However, this changes if the server is its own client. If so, then performance issues tend to be related to the server. It is possible for NIS servers to bind to other servers on their network, which can lead to interdependencies which can be difficult to debug. The `ypwhich` command is used to determine client binding. NIS has the ability to forward hostname lookups to DNS. Therefore problems which affect DNS servers and NIS clients can affect the performance of the NIS server. These problems are described below.

Hierarchal name services such as DNS and NIS+ can have complex configuration problems. Also, DNS servers usually act as DNS clients. NIS+ servers are clients to the higher domain. Therefore problems in higher domains can influence the lower domains.

Name Service Problems on Clients

The first place to look for name service problems on clients is the

`/etc/nsswitch.conf` file.

This file sets the policies for name services. The term database corresponds to a type of service needed. For example, the hosts database is used for the

`gethostbyname(3NSL)` and

`gethostbyaddr(3NSL)` library calls.

Multiple name services may be used, but only one successful query response will be used.

Each process using the `nsswitch.conf` file, the entire file is read only once. If the file is later changed, the process will continue to use the old configuration.

Any mistake in spelling names of sources or databases in the `nsswitch.conf` file may cause delays which affect client applications. These delays may not be apparent in dynamic performance tuning as clients will block waiting on the result of their query. This blocking consumes real “wall clock” time, but not CPU time.

NIS

The compiled-in default entries for all databases use NIS as the enterprise level name service and are identical to those in the default configuration shown in List 1.

List 1

```
#
# /etc/nsswitch.nis:
#
# An example file that could be copied over to /etc/nsswitch.conf; it
# uses NIS (YP) in conjunction with files.
#
# “hosts:” and “services:” in this file are used only if the
# /etc/netconfig file has a “-” for nametoaddr_libs of “inet” trans-
ports.

# the following two lines obviate the “+” entry in /etc/passwd and /
etc/group.
passwd:     files nis
group:      files nis

# consult /etc “files” only if nis is down.
hosts:      nis [NOTFOUND=return] files
ipnodes:    files
# Uncomment the following line and comment out the above to resolve
# both IPv4 and IPv6 addresses from the ipnodes databases. Note that
# IPv4 addresses are searched in all of the ipnodes databases before
# searching the hosts database. Before turning this option on, consult
# the Network Administration Guide for more details on using IPv6.
#ipnodes:    nis [NOTFOUND=return] files

networks:   nis [NOTFOUND=return] files
protocols:  nis [NOTFOUND=return] files
rpc:        nis [NOTFOUND=return] files
ethers:     nis [NOTFOUND=return] files
netmasks:   nis [NOTFOUND=return] files
bootparams: nis [NOTFOUND=return] files
publickey:  nis [NOTFOUND=return] files

netgroup:   nis

automount:  files nis
aliases:    files nis

# for efficient getservbyname() avoid nis
services:   files nis
sendmailvars:   files
printers:   user files nis

auth_attr:  files nis
prof_attr:  files nis
```

The policy “`nis [NOTFOUND=return] files`” implies that “if NIS is UNAVAIL, continue on to files, and if NIS returns NOTFOUND, return to the caller; in other words, treat NIS as the authoritative source of information and try files only if NIS is down.” This, and other policies listed in the default configuration above, are identical to the hard-wired policies in SunOS™ releases prior to 5.0.

The default NIS policies may not be appropriate for all systems. In particular, the default policy does not use local files for host name resolution unless the NIS server becomes unavailable. An error in the entry for a client in the NIS hosts database may cause performance problems on the client, as it may not be able to resolve its own address. It may be desirable to use `dns` and `files` as authoritative sources, in addition to NIS.

NIS+

NIS+ has a hierarchal design and supports multiple replicated servers. It is also very flexible, which creates opportunities for operator induced errors. A number of problems which can be encountered while configuring and operating NIS+ services are listed in the *Solaris Naming Administration Guide*, Appendix A.

Some problems directly affecting performance are:

Too Many Replicas

Too many replicas for a domain degrade system performance during replication. More than five replicas in a domain may cause noticeable performance problems.

Recursive Groups

A recursive group is one that contains the name of some other group. Including the name of other groups within a group may reduce system administrator workload, however, doing so slows down the system. Therefore, recursive groups should be avoided.

NIS_PATH Variable

The `NIS_PATH` environment variable should be simple. For example, the default: `org_dir.$:$`. A complex `NIS_PATH`, particularly one that itself contains a variable, will slow NIS+ queries.

DNS

The DNS configuration file is /etc/resolv.conf. Typically, the domain name and one or more name servers are listed. An example DNS configuration file is shown in List 2.

List 2 - Sample /etc/resolv.conf File

```
domain west.sun.com
nameserver 192.168.33.2
nameserver 192.168.47.199
```

A missing or incorrect domain name can cause multiple DNS queries. If a query is made which is not fully qualified, then the query will append the domain name. For example, a query could be for host.west.sun.com. If that query fails, the following query will strip the first subdomain from the domain name and try again with host.sun.com. This process will continue until either a match is made, or the domain name is exhausted. For fully qualified name queries, the first query will be the name, then if that fails, the additional queries will append the domain name as described above. Incorrect, or domain names with many subdomains may cause DNS queries to take longer than necessary.

The name server list is a set of DNS servers which can respond to queries. It is suggested that at least one name server be listed to provide some fault resiliency. The network latency between the client and each name server should be considered when determining the order of the list.

The failure of a DNS server will result in a delay before the query is submitted to the next name server in the list. An incorrect name server entry or a network routing problem between the client and server will be perceived as a delay by the client. This delay affects both host to address queries and address to host queries. Therefore, services which log network access, such as telnet and ftp will be affected.

For example, a client configured for files and DNS for the hosts database with only a single, incorrect name server entry in /etc/resolv.conf file will suffer a delay of approximate 75 seconds before displaying the login prompt for a telnet session. To the end user this would appear as if the client was down.

Name Service Cache Daemon

The name service cache daemon, nscd, provides a cache for the most common name service requests. It is important to know if nscd is configured when troubleshooting any name service related performance problems. nscd may mask name service problems. It is suggested to disable nscd when troubleshooting client name service problems.

Configuration

System hardware and software configuration have an important effect on performance. The measurements of system bandwidth and utilization are important factors in dynamic performance tuning. Static performance tuning concentrates on the health of the configuration and architectural issues.

A configuration issue might be:

- Is one of the dual redundant paths to a disk subsystem down?

An architectural issue might be:

- How many nodes per hub are being used?

NFS Client Problems

NFS clients can be configured to perform very poorly. An example would be that of nested mounting. List 3 shows an example of nested mounts and how convoluted they can become if incorrectly managed.

In this example, if a user tries to run a SPARC™ executable binary in the project directory then three NFS servers and the network must be operational. Also, as NFS stats the path, each server will respond with information about its section of the path. In such a scenario, performance would be enhanced by using a flatter hierarchy.

It is not uncommon to see nested mounts in environments that have grown over time. Networks that have grown from workstations to workstations with multiple servers are particularly susceptible. It is also not uncommon to see desktop workstations providing file service. In the author’s experience, the worst case found was a network of workstations and servers in which the nested mounts were so complex that a client would mount itself underneath a remote server’s mount. Clearly, such scenarios should be avoided for both performance and availability reasons.

List 3 - Nested NFS Mounts

```
Filesystem                    Mounted on
server1:/export/opt              /opt
server2:/opt/project             /opt/project
server3:/export/local/bin        /opt/project/bin
server1:/export/sparc.bin        /opt/project/bin/sparc
```

Network Topology Problems

A network plays an important role in the performance of name services. The network latency between the client and server will have a negative impact on client lookup performance and should be minimized.

The `traceroute` utility is useful for determining the distance in hops from the client to the server, and gives a relative feel for the dynamic latency. One can only infer dynamic latency from `traceroute`, however, full characterization would be of little practical value because of the dynamic nature of network traffic. It would be practical to optimize for hops; the fewer the better. Long hops with response times greater than a few milliseconds may be WAN hops; LAN hops are preferred. For performance considerations, the `traceroute` utility can be used from the client to help determine the best ordering of the name servers. This utility can also help debug configuration problems in network equipment. For example, a packet which travels across the country and back again may indicate a router or firewall is misconfigured.

Duplicate IP addresses can result in intermittent performance problems. Solaris™ Operating Environment logs duplicate IP addresses in the `/var/adm/messages` file. With the wide variety of clients available, and the ease at which untrained users can set IP addresses on their systems, duplicate IP address problems are becoming common. A better solution is to limit the exposure of the servers by segregating them from any uncontrolled client networks. Unfortunately the client side of this problem is more difficult to debug, as some clients do not bother the user with warning messages.

Hardware Problems

Hardware problems can also degrade performance. Traditionally, hardware problems tend to result in nonfunctional systems. As more fault resiliency is built into systems it is quite possible for the system to remain functional, but operating in a degraded mode.

CPU Failures

CPU failures in most Sun™ systems result in a Solaris Operating Environment panic. For Sun Enterprise™ servers however, the system will attempt to reboot while isolating and disabling the faulty CPU. This is done as part of Automatic System Recovery (ASR) in Sun Enterprise 3000-6500 servers and power on self test (HPOST) in Sun Enterprise 10000 servers. All processors should be operational for best performance.

The problem is that such reconfiguration may go unnoticed. The `psrinfo` command will show how many processors are online. The Sun Management Console and Configuration Service Tracker can be used to monitor the status of all hardware and create a report to a central monitoring station.

Disk Failures

Disk failures may have a negative impact on performance, however the failures may be masked by a RAID system. The performance may degrade while a hot spare disk is brought online, or a volume is relocated to an operational disk. These performance problems tend to disappear once the hot spare is fully operational. However, the performance of the system will likely be different after such an event. RAID systems should be checked to ensure that all is well. Error messages from the RAID systems should be collected by a monitoring system.

Disks themselves can change their performance characteristics as they wear. It is common for many disks to allocate one or two sectors per track to be available to hold data in case of media error. Any bad sector will be remapped to the spare sector on the track. However, if there are too many media errors, the sector may be slipped to a spare track. Typically the spare tracks are located at the end of the disk. A slipped sector might require a full seek to access the data. This can significantly change the performance of applications using the disk.

The Solaris Operating Environment command `iostat -E` shows errors encountered by the drive. Predictive failure errors are also logged. Disks should be checked periodically for increasing error rates.

SCSI Bus Speed

The SCSI protocol implements negotiation of the data transfer rate between hosts and devices. This allows older, slower SCSI devices to coexist with newer, faster devices. The faster SCSI transfer rates put tighter restrictions on bus length and noise on the bus. Long busses and noisy environments may result in lower negotiated bus speeds.

The Solaris Operating Environment does not explicitly notify the system administrator of the negotiated SCSI bus speeds. This data is available via the `prtconf -v` command. Unfortunately, this data is not in an easily readable form. For example:

```
# prtconf -v
...
  SUNW,fas, instance #1
            Driver properties:
...
                name <target2-TQ> length <4>
                    value <0x00000001>.
                name <target2-wide> length <4>
                    value <0x00000001>.
                name <target2-sync-speed> length <4>
                    value <0x00004e20>.
...
                name <scsi-options> length <4>
                    value <0x00001ff8>.
...
```

From this example the Qlogic FAS366 SCSI (fas) driver describes the driver properties per target. In this case target 2 has negotiated command tagged queing (TQ) and 16 bit (wide) transfers. The transfer speed (sync-speed) for target 2 is 0x4e20 KB/s or 20,000 KB/s.

Some third party vendors have offered settings which explicitly disable some SCSI bus options. This is accomplished by setting the `scsi-options` parameter in the `/etc/system` file or the host bus adaptor driver configuration file (see `driver.conf(4)` for more information). Unfortunately, these settings may also have the effect of limiting the negotiated SCSI bus options. This is especially the case where the options are expanded in a Solaris Operating Environment release and the `scsi-option` setting does not allow the new features to be enabled.

In the previous example, the `scsi-options` default is set to 0x1ff8. The meaning of each bit in the `scsi-options` is specified in the `/usr/include/sys/scsi/conf/autoconf.h` file as:

```
/*
 * Following are applicable to all interconnects
 */
#define SCSI_OPTIONS_LINK   0x10  /* Global linked commands */
#define SCSI_OPTIONS_TAG    0x80  /* Global tagged command support
*/

/*
 * Following are for parallel SCSI only
 */
#define SCSI_OPTIONS_DR   0x8  /* Global disconnect/reconnect  */
#define SCSI_OPTIONS_SYNC  0x20 /* Global synchronous xfer
capability */
#define SCSI_OPTIONS_PARITY  0x40    /* Global parity support */
#define SCSI_OPTIONS_FAST    0x100   /* Global FAST scsi support */
#define SCSI_OPTIONS_WIDE    0x200   /* Global WIDE scsi support */
#define SCSI_OPTIONS_FAST20  0x400   /* Global FAST20 scsi support
*/
#define SCSI_OPTIONS_FAST40  0x800   /* Global FAST40 scsi support
*/
#define SCSI_OPTIONS_FAST80  0x1000  /* Global FAST80 scsi support
*/
```

Thus the `scsi-options` flag setting of 0x1ff8 enables all of the specified SCSI options to be negotiated. This is the system default value for the Solaris 8 Operating Environment.

Network Equipment Problems

Faulty network equipment can have a negative impact on the performance of networked systems. An obvious problem would be a port which was not working. Any network that is expected to be operational should be connected to the network equipment at boot time. During the boot sequence, the Solaris Operating Environment attempts to plumb and bring up all interfaces with `/etc/hostname.*` entries. A disconnected or failed network will delay the boot process considerably.

Other hardware errors in network equipment may be less obvious.

The `netstat -i` command can be used to check the input and output error rates. Output errors could indicate faulty hardware or cables. The output error rate should be zero, or less than 0.25% for all output packets.

Software Patches

Performance improvements can be delivered as patches for existing software. Patches are created as bottlenecks are determined, understood, and fixed. A collection of patches for delivering performance improvements is shown in Table 1.

Patch-ID#	Description
106541-09	SunOS™ 5.7: Kernel update patch
107841-01	SunOS 5.7: /kernel/misc/rpcsec and /kernel/misc/sparcv9/rpcsec patch
107148-06	SunOS 5.7: /kernel/fs/cachefs patch
107473-04	SunOS 5.7: luxadm patch
107324-01	SunOS 5.7: Euro locales, user interface refresh is very slow
106145-15	SunOS 5.7: Creator 7 FFB Graphics Patch
106146-13	SunOS 5.7: M64 Graphics Patch
106148-10	SunOS 5.7: XFB Graphics Patch
107081-10	Motif 1.2.7 and 2.1.1: Runtime library patch for Solaris 7 Operating Environment
108131-06	OpenGL® 1.2: OpenGL Patch for Solaris 2.5.1/2.6/7 Operating Environment (32-bit)
108132-06	OpenGL 1.2: OpenGL Patch for Solaris 7 Operating Environment (64-bit)
107432-03	SunOS 5.7: CTL printing patch
107115-03	SunOS 5.7: LP patch

Patch-ID#	Description
107242-02	Java™ DMK 3.0: Patch
107245-02	Java DMK 3.0: patch (Runtime only)
107780-01	Java DMK 3.2: Patch
107782-01	Java DMK 3.2: patch (Runtime only)
108391-01	Sun StorEdge™ Component Manager 1.0: Component Manager patch
106871-01	Sun Quad FastEthernet™ 2.2: POINT PATCH: to fix interrupt distribution
106765-04	Sun Gigabit Ethernet 2.0: Patch for Solaris 7 Operating Environment ge driver
106738-04	SunFDDI™ SBus 6.0: Performance enhancements and NFS failure
106736-02	Solstice Backup™ 5.1.1: Product Patch
107356-04	Fortran 90 2.0: Patch for Fortran 90 (f90) 2.0 compiler
107377-04	Fortran 90 2.0: Patch for 64-bit Fortran 90 (f90) 2.0 compiler
104018-05	Solstice Site Manager™/Solstice SunNet Manager™/Solstice Domain Manager™ 2.3 Rev B: Multiple fixes patch
104844-06	Sun WorkShop™ IPE 4.0: Patch
107355-04	Sun WorkShop IPE 5.0: Patch for dbx
107358-04	Sun WorkShop IPE 5.0: Patch for 64-bit dbx
107311-09	C++ 5.0: Patch for C++ 5.0 compiler
107390-09	C++ 5.0: Patch for 64-bit C++ 5.0 compiler
107596-03	F77 5.0: Patch for FORTRAN 77 (f77) 5.0 compiler
108002-01	HPC 3.0: S3L fixes
108003-01	HPC 3.0: S3L 64bit fixes
106514-09	Solstice Internet Mail Server 3.5: Misc. fixes
108049-05	Solstice Internet Mail Server 4.0: Misc. fixes
106621-07	Sun Directory Services 3.1: Patch

The contents of Table 1 were compiled by searching the SunSolve™ patch database for SunOS 5.7 and by using the keywords *performance* or *slow*.

Conclusion

Static performance tuning complements dynamic performance tuning. Both are effective at detecting problems. The static performance tuning methods described here should be part of any system performance analysis. High performance cars with flat tires do not drive very fast.

References

Solaris Naming Administration Guide. In addition to name service administration, this guide also has troubleshooting information for NIS, NIS+, and DNS.

The *Sun™ Management Center* documentation is available via the web at http://www.sun.com/sunmanagementcenter/docs/index.html

The *Configuration and Service Tracker* is available from Sun Enterprise Services and is free of charge for systems covered under a SunSpectrum℠ program contract.

Author's Bio: Richard Elling

Richard Elling is a Senior Staff Engineer in the Enterprise Engineering group at Sun Microsystems in San Diego, California. Prior to the Enterprise Engineering group, Richard had been a Field Systems Engineer at Sun for five years. Richard was the Sun Worldwide Field Systems Engineer of the year in 1996. Prior to Sun, Richard was the Manager of Network Support for the College of Engineering at Auburn University. Richard was also a design engineer for a startup microelectronics company, and he worked for NASA doing electronic design and experiments integration for Space Shuttle missions.