Oracle® Solaris Tunable Parameters Reference Manual
Ce logiciel et la documentation qui l’accompagne sont protégés par les lois sur la propriété intellectuelle. Ils sont concédés sous licence et soumis à des restrictions d’utilisation et de divulgation. Sauf disposition de votre contrat de licence ou de la loi, vous ne pouvez pas copier, reproduire, traduire, diffuser, modifiier, breveter, transmettre, distribuer, exposer, exécuter, publier ou afficher le logiciel, même partiellement, sous quelque forme et par quelque procédé que ce soit. Par ailleurs, il est interdit de procéder à toute ingénierie inverse du logiciel, de le déasssembler ou de le décompiler, excepté à des fins d’interopérabilité avec des logiciels tiers ou tel que prescrit par la loi.

Les informations fournies dans ce document sont susceptibles de modification sans préavis. Par ailleurs, Oracle Corporation ne garantit pas qu’elles soient exemptes d’erreurs et vous invite, le cas échéant, à lui en faire part par écrit.

Si ce logiciel, ou la documentation qui l’accompagne, est concédé sous licence au Gouvernement des États-Unis, ou à toute entité qui délivre la licence de ce logiciel ou l’utilise pour le compte du Gouvernement des États-Unis, la notice suivante s’applique:

U.S. GOVERNMENT RIGHTS. Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle America, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

Ce logiciel ou matériel a été développé pour un usage général dans le cadre d’applications de gestion des informations. Ce logiciel ou matériel n’est pas conçu ni n’est destiné à être utilisé dans des applications à risque, notamment dans des applications pouvant causer des dommages corporels. Si vous utilisez ce logiciel ou matériel dans le cadre d’applications dangereuses, il est de votre responsabilité de prendre toutes les mesures de secours, de sauvegarde, de redondance et autres mesures nécessaires à son utilisation dans des conditions optimales de sécurité. Oracle Corporation et ses affiliés déclinent toute responsabilité quant aux dommages causés par l’utilisation de ce logiciel ou matériel pour ce type d’applications.

Oracle et Java sont des marques déposées d’Oracle Corporation et/ou de ses affiliés. Tout autre nom mentionné peut correspondre à des marques appartenant à d’autres propriétaires qu’Oracle.

Intel et Intel Xeon sont des marques ou des marques déposées d’Intel Corporation. Toutes les marques SPARC sont utilisées sous licence et sont des marques ou des marques déposées d’Advanced Micro Devices, Inc. AMD, Opteron, le logo AMD et le logo AMD Opteron sont des marques ou des marques déposées d'Advanced Micro Devices. UNIX est une marque déposée de The Open Group.

Ce logiciel ou matériel et la documentation qui l’accompagne peuvent fournir des informations ou des liens donnant accès à des contenus, des produits et des services émanant de tiers. Oracle Corporation et ses affiliés déclinent toute responsabilité ou garantie expresse quant aux contenus, produits ou services émanant de tiers. En aucun cas, Oracle Corporation et ses affiliés ne sauraient être tenus pour responsables des pertes subies, des coûts occasionnés ou des dommages causés par l’accès à des contenus, produits ou services tiers, ou à leur utilisation.
Contents

Preface ................................................................................................................................................... 13

1 Overview of Oracle Solaris System Tuning ............................................................... 17
  What’s New in Oracle Solaris System Tuning? .............................................................. 17
  Tuning an Oracle Solaris System .................................................................................... 18
  Tuning Format of Tunable Parameters Descriptions ..................................................... 19
  Tuning the Oracle Solaris Kernel ...................................................................................... 21
    /etc/system File ............................................................................................................ 21
    kmdb Command .......................................................................................................... 22
    mdb Command ............................................................................................................ 22
  Special Oracle Solaris tune and var Structures ............................................................ 23
  Viewing Oracle Solaris System Configuration Information ......................................... 23
    sysdef Command ....................................................................................................... 24
    kstat Utility ................................................................................................................ 24

2 Oracle Solaris Kernel Tunable Parameters ............................................................... 25
  Where to Find Tunable Parameter Information ........................................................... 25
  General Kernel and Memory Parameters ...................................................................... 26
    physmem ..................................................................................................................... 26
    zfs_arc_min ................................................................................................................ 26
    zfs_arc_max ................................................................................................................ 27
    default_stksize .......................................................................................................... 28
    lwp_default_stksize .................................................................................................... 29
    logevent_max_q_sz ...................................................................................................... 30
    segkpsize ................................................................................................................... 30
    noexec_user_stack .................................................................................................... 31
    fsflush and Related Parameters ................................................................................. 32
Contents

- fsflush ......................................................................................................................................... 32
- tune_t_fsflushr ................................................................. 33
- autoup ......................................................................................... 33
- dopageflush ........................................................................... 34
- doiflush ................................................................................... 35

Process-Sizing Parameters ................................................................................................................. 36
- maxusers ..................................................................................... 36
- reserved_procs ......................................................................... 37
- pidmax ......................................................................................... 37
- max_nprocs .................................................................................. 38
- maxuprc ......................................................................................... 39
- ngroups_max ............................................................................. 39

Paging-Related Parameters ................................................................................................................40
- lotsfree ....................................................................................... 41
- desfree ......................................................................................... 42
- minfree ......................................................................................... 43
- throttlefree ................................................................................ 44
- pageout_reserve ......................................................................... 45
- pages_pp_maximum ...................................................................... 46
- tune_t_minarmem ....................................................................... 47
- fastscan ....................................................................................... 47
- slowscan ....................................................................................... 48
- min_percent_cpu ......................................................................... 48
- handspreadpages ......................................................................... 49
- pages_beforePager .................................................................... 50
- maxpqio ....................................................................................... 50

Swapping-Related Parameters ........................................................................................................... 51
- swapfs_reserve ........................................................................... 51
- swapfs_minfree .......................................................................... 52

Kernel Memory Allocator .............................................................................................................. 52
- kmem_flags ................................................................................. 53
- kmem_stackinfo ......................................................................... 54

General Driver Parameters .............................................................................................................. 55
- moddebug ..................................................................................... 55
- ddi_msix_alloc_limit .............................................................. 56

Network Driver Parameters ........................................................................................................... 57
3  **NFS Tunable Parameters** ........................................................................................................83

Where to Find Tunable Parameter Information ........................................................................83
Tuning the NFS Environment ........................................................................................................83

NFS Module Parameters ................................................................................................................84
  nfs:nfs3_pathconf_disable_cache ..........................................................................................84
  nfs:nfs4_pathconf_disable_cache .......................................................................................84
  nfs:nfs_allow_preePOCH_time ..............................................................................................85
  nfs:nfs_cots_timeo ...............................................................................................................86
  nfs:nfs3_cots_timeo .............................................................................................................86
  nfs:nfs4_cots_timeo ...............................................................................................................87
  nfs:nfs_do_symlink_cache ......................................................................................................88
  nfs:nfs3_do_symlink_cache .................................................................................................88
  nfs:nfs4_do_symlink_cache ..................................................................................................89
  nfs:nfs_dynamic ...................................................................................................................89
  nfs:nfs3_dynamic ................................................................................................................90
  nfs:nfs_lookup_neg_cache .....................................................................................................90
  nfs:nfs3_lookup_neg_cache ...................................................................................................91
  nfs:nfs4_lookup_neg_cache ...................................................................................................92
  nfs:nfs_max_threads ............................................................................................................93
  nfs:nfs3_max_threads ...........................................................................................................94
  nfs:nfs4_max_threads ..........................................................................................................94
  nfs:nfs_nra ........................................................................................................................95
  nfs:nfs3_nra ........................................................................................................................96
  nfs:nfs4_nra ........................................................................................................................97
  nfs:nrnode ............................................................................................................................97
  nfs:nfs_shrinkreaddir ..........................................................................................................98
  nfs:nfs3_shrinkreaddir .........................................................................................................99
  nfs_write_error_interval .................................................................................................100
  nfs_write_error_to_cons_only ............................................................................................100
nfs:nfs_disable_rddir_cache ................................................................. 101
nfs:nfs3_bsize .................................................................................... 102
nfs:nfs4_bsize .................................................................................... 102
nfs:nfs_async_clusters ...................................................................... 103
nfs:nfs3_async_clusters ..................................................................... 104
nfs:nfs4_async_clusters ..................................................................... 105
nfs:nfs_async_timeout ...................................................................... 106
nfs:nacache ....................................................................................... 106
nfs:nfs3_jukebox_delay .................................................................... 107
nfs:nfs3_max_transfer_size ................................................................ 108
nfs:nfs4_max_transfer_size ................................................................ 109
nfs:nfs3_max_transfer_size_clts ........................................................ 110
nfs:nfs3_max_transfer_size_cots ........................................................ 110

rpcmod Module Parameters ................................................................ 111
rpcmod:clnt_max_conns .................................................................... 111
rpcmod:clnt_idle_timeout .................................................................. 112
rpcmod:svc_idle_timeout .................................................................. 112
rpcmod:svc_default_stksize .............................................................. 112
rpcmod:maxdupreqs ......................................................................... 113
rpcmod:cotsmaxdupreqs ................................................................... 114

4 Internet Protocol Suite Tunable Parameters .................................... 117
Where to Find Tunable Parameter Information .................................... 117
Overview of Tuning IP Suite Parameters .......................................... 117
IP Suite Parameter Validation ............................................................ 118
Internet Request for Comments (RFCs) ........................................... 118
IP Tunable Parameters ........................................................................ 118
icmp_err_interval and icmp_err_burst .............................................. 118
respond_to_echo_broadcast and respond_to_echo_multicast (ipv4 or ipv6) ................................................................. 119
send_redirects (ipv4 or ipv6) ............................................................... 119
forwarding (ipv4 or ipv6) ................................................................. 120
ttl ................................................................................................. 120
hoplimit (ipv6) ................................................................................. 120
addrs_per_if ................................................................................... 121
hostmodel (ipv4 or ipv6) ................................................................. 121
5 Network Cache and Accelerator Tunable Parameters ........................................... 157
Where to Find Tunable Parameters Information ............................................ 157
Tuning NCA Parameters .................................................................................. 157
nca:nca_conn_hash_size ............................................................................... 158
nca:nca_conn_req_max_q ........................................................................... 158
nca:nca_conn_req_max_q0 .......................................................................... 158
nca:nca_ppmax ............................................................................................ 159
Contents

nca:nca_vpmax ........................................................................................................................... 159
General System Tuning for the NCA ....................................................................................... 160
sq_max_size ............................................................................................................................... 160
ge:ge_intr_mode ...................................................................................................................... 161

6 System Facility Parameters .......................................................................................................................... 163
System Default Parameters .................................................................................................................. 164
  autofs .............................................................................................................................................. 164
cron .............................................................................................................................................. 164
devfsadm ....................................................................................................................................... 164
dhcpcagent ................................................................................................................................... 164
fs .................................................................................................................................................. 165
ftp .................................................................................................................................................. 165
inetinit ....................................................................................................................................... 165
init ............................................................................................................................................... 165
ipsec ............................................................................................................................................ 166
kbd .............................................................................................................................................. 166
keyserv ....................................................................................................................................... 166
login .......................................................................................................................................... 166
mpathd ....................................................................................................................................... 167
nfs .............................................................................................................................................. 167
nfslogd ...................................................................................................................................... 167
nss .............................................................................................................................................. 167
passwd ....................................................................................................................................... 167
su ............................................................................................................................................... 168
syslog ....................................................................................................................................... 168
tar .............................................................................................................................................. 168
telnetd ...................................................................................................................................... 168
utmpd ........................................................................................................................................ 168

A Tunable Parameters Change History ................................................................................................. 169
Kernel Parameters ....................................................................................................................... 169
  Paging-Related Parameters ........................................................................................................ 169
Process-Sizing Tunables ............................................................................................................. 169
  General Driver Parameter ...................................................................................................... 170
Network Driver Parameters ...................................................................................................... 170
General Kernel and Memory Parameters ................................................................................ 170
TCP/IP Tunable Parameters ....................................................................................................... 171
[tcp,sctp,udp] smallest_anon_port and [tcp,sctp,udp] largest_anon_port (Oracle Solaris 11) .............................................................................................................. 171
IP Parameter Name Changes (Oracle Solaris 11) ................................................................ 171
TCP Parameter Name Changes (Oracle Solaris 11) ............................................................... 172
UDP Parameter Name Changes (Oracle Solaris 11) ............................................................ 173
SCTP Parameter Name Changes (Oracle Solaris 11) ............................................................ 174
Parameters That Are Obsolete or Have Been Removed .......................................................... 175
rstchown ................................................................................................................................ 175
Obsolete TCP/IP Module Parameters .................................................................................... 176

B Revision History for This Manual .......................................................................................... 177
Current Version: Oracle Solaris 11 Release .............................................................................. 177
New or Changed Parameters in the Oracle Solaris Release .................................................. 177

Index .......................................................................................................................................... 179
Preface

The Oracle Solaris Tunable Parameters Reference Manual provides reference information about Oracle Solaris OS kernel and network tunable parameters. This manual does not provide tunable parameter information about desktop systems or Java environments.

This manual contains information for both SPARC based and x86 based systems.

Note - This Oracle Solaris release supports systems that use the SPARC and x86 families of processor architectures. The supported systems appear in the Oracle Solaris Hardware Compatibility List at http://www.oracle.com/webfolder/technetwork/hcl/index.html. This document cites any implementation differences between the platform types.

Who Should Use This Book

This book is intended for experienced Oracle Solaris system administrators who might need to change kernel tunable parameters in certain situations. For guidelines on changing Oracle Solaris tunable parameters, refer to “Tuning an Oracle Solaris System” on page 18.

How This Book Is Organized

The following table describes the chapters and appendixes in this book.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1, “Overview of Oracle Solaris System Tuning”</td>
<td>An overview of tuning an Oracle Solaris system. Also provides a description of the format used in the book to describe the kernel tunables.</td>
</tr>
<tr>
<td>Chapter 2, “Oracle Solaris Kernel Tunable Parameters”</td>
<td>A description of Oracle Solaris kernel tunables such as kernel memory, file system, process size, and paging parameters.</td>
</tr>
<tr>
<td>Chapter 3, “NFS Tunable Parameters”</td>
<td>A description of NFS tunables such as caching symbolic links, dynamic retransmission, and RPC security parameters.</td>
</tr>
<tr>
<td>Chapter 4, “Internet Protocol Suite Tunable Parameters”</td>
<td>A description of TCP/IP tunables such as IP forwarding, source routing, and buffer-sizing parameters.</td>
</tr>
</tbody>
</table>
Other Resources for Oracle Solaris Tuning Information

This table describes other resources for Oracle Solaris tuning information.

<table>
<thead>
<tr>
<th>Tuning Resource</th>
<th>For More Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online performance tuning information</td>
<td><a href="http://www.solarisinternals.com/si/index.php">http://www.solarisinternals.com/si/index.php</a></td>
</tr>
<tr>
<td>In-depth technical white papers</td>
<td><a href="http://www.oracle.com/technetwork/server-storage/solaris/overview/index.html">http://www.oracle.com/technetwork/server-storage/solaris/overview/index.html</a></td>
</tr>
</tbody>
</table>

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Typographic Conventions

The following table describes the typographic conventions that are used in this book.

<table>
<thead>
<tr>
<th>Typeface</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AaBbCc123</td>
<td>The names of commands, files, and directories, and onscreen computer output</td>
<td>Edit your .login file. Use ls -a to list all files. machine_name% you have mail.</td>
</tr>
</tbody>
</table>
TABLE P–1  Typographic Conventions  (Continued)

<table>
<thead>
<tr>
<th>Typeface</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AaBbCc123</td>
<td>What you type, contrasted with onscreen computer output</td>
<td>machine_name% su</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Password:</td>
</tr>
<tr>
<td>aabbcc123</td>
<td>Placeholder: replace with a real name or value</td>
<td>The command to remove a file is rm filename.</td>
</tr>
<tr>
<td>AaBbCc123</td>
<td>Book titles, new terms, and terms to be emphasized</td>
<td>Read Chapter 6 in the User's Guide. A cache is a copy that is stored locally. Do not save the file.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong> Some emphasized items appear bold online.</td>
</tr>
</tbody>
</table>

Shell Prompts in Command Examples

The following table shows the default UNIX system prompt and superuser prompt for shells that are included in the Oracle Solaris OS. Note that the default system prompt that is displayed in command examples varies, depending on the Oracle Solaris release.

TABLE P–2  Shell Prompts

<table>
<thead>
<tr>
<th>Shell</th>
<th>Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bash shell, Korn shell, and Bourne shell</td>
<td>$</td>
</tr>
<tr>
<td>Bash shell, Korn shell, and Bourne shell for superuser</td>
<td>#</td>
</tr>
<tr>
<td>C shell</td>
<td>machine_name%</td>
</tr>
<tr>
<td>C shell for superuser</td>
<td>machine_name#</td>
</tr>
</tbody>
</table>
Overview of Oracle Solaris System Tuning

This section provides overview information about the format of the tuning information in this manual. This section also describes the different ways to tune an Oracle Solaris system.

- “What’s New in Oracle Solaris System Tuning?” on page 17
- “Tuning an Oracle Solaris System” on page 18
- “Tuning Format of Tunable Parameters Descriptions” on page 19
- “Tuning the Oracle Solaris Kernel” on page 21
- “Special Oracle Solaris tune and var Structures” on page 23
- “Viewing Oracle Solaris System Configuration Information” on page 23
- “kstat Utility” on page 24

What's New in Oracle Solaris System Tuning?

This section describes new or changed parameters in the Oracle Solaris 11 release.

- Oracle Solaris 11: The rstchown parameter that was previously set in the /etc/system file is obsolete. If you set this parameter in the /etc/system file, the following error message is displayed:

  sorry, variable ‘rstchown’ is not defined in the ‘kernel’

  This parameter has been replaced by the ZFS rstchown file system property and a general file system mount option. For more information, see Oracle Solaris Administration: ZFS File Systems and mount(1M).

- Oracle Solaris 11: The following system configuration parameters that were previously configured by editing files in the /etc/default directory have changed to SMF services:
  - autos
  - init
  - kbd
  - nfs
For information about changing SMF service properties, see Chapter 6, “System Facility Parameters.”

- Oracle Solaris 11: The `ipadm` command replaces the `ndd` command for setting network properties. TCP, IP, UDP, and SCTP properties are set as follows:
  ```
  ipadm set-prop -p parameter ip|ipv4|ipv6|tcp|udp|sctp
  ```
  In addition, most of the network tunable names have changed slightly to correlate better with the `ipadm` format.

  For more information, see “Overview of Tuning IP Suite Parameters” on page 117.

- Oracle Solaris 11: This release includes the `disp_rechoose_interval` parameter. For more information, see “`disp_rechoose_interval`” on page 73.

- Oracle Solaris 11: This release includes the `ngroups_max` parameter description. For more information, see “`ngroups_max`” on page 39.

- Oracle Solaris 11: This release includes the `zfs_arc_min` and `zfs_arc_max` parameter descriptions. For more information, see “`zfs_arc_min`” on page 26 and “`zfs_arc_max`” on page 27.

  For additional information about tuning ZFS file systems, see the following site:

- Oracle Solaris 11: This release includes several `igb` and `ixgbe` network driver parameters. For more information, see “`igb Parameters`” on page 57 and “`ixgbe Parameters`” on page 58.

- Oracle Solaris 11: This release includes the `ddi_msix_alloc_limit` parameter that can be used to increase the number of MSI-X interrupts that a device instance can allocate. For more information, see “`ddi_msix_alloc_limit`” on page 56.

- Oracle Solaris 11: This release includes the `kmem_stackinfo` parameter, which can be enabled to monitor kernel thread stack usage. For more information, see “`kmem_stackinfo`” on page 54.

- Oracle Solaris 11: Memory locality group parameters are provided in this release. For more information about these parameters, see “Locality Group Parameters” on page 79.

### Tuning an Oracle Solaris System

The Oracle Solaris OS is a multi-threaded, scalable UNIX operating system that runs on SPARC and x86 processors. It is self-adjusting to system load and demands minimal tuning. In some cases, however, tuning is necessary. This book provides details about the officially supported kernel tuning options available for the Oracle Solaris OS.
The Solaris kernel is composed of a core portion, which is always loaded, and a number of loadable modules that are loaded as references are made to them. Many variables referred to in the kernel portion of this guide are in the core portion. However, a few variables are located in loadable modules.

A key consideration in system tuning is that setting system parameters (or system variables) is often the least effective action that can be done to improve performance. Changing the behavior of the application is generally the most effective tuning aid available. Adding more physical memory and balancing disk I/O patterns are also useful. In a few rare cases, changing one of the variables described in this guide will have a substantial effect on system performance.

Remember that one system’s /etc/system settings might not be applicable, either wholly or in part, to another system’s environment. Carefully consider the values in the file with respect to the environment in which they will be applied. Make sure that you understand the behavior of a system before attempting to apply changes to the system variables that are described here.

We recommend that you start with an empty /etc/system file when moving to a new Oracle Solaris release. As a first step, add only those tunables that are required by in-house or third-party applications. After baseline testing has been established, evaluate system performance to determine if additional tunable settings are required.

Caution – The tunable parameters described in this book can and do change from Oracle Solaris release to Oracle Solaris release. Publication of these tunable parameters does not preclude changes to the tunable parameters and their descriptions without notice.

Tuning Format of Tunable Parameters Descriptions

The format for the description of each tunable parameter is as follows:

- Parameter Name
- Description
- Data Type
- Default
- Range
- Units
- Dynamic?
- Validation
- Implicit
- When to Change
- Zone Configuration
- Commitment Level
- Change History

Parameter Name  Is the exact name that is typed in the /etc/system file, or found in the /etc/default/facility file.
Most parameters names are of the form `parameter` where the parameter name does not contain a colon (\( : \)). These names refer to variables in the core portion of the kernel. If the name does contain a colon, the characters to the left of the colon reference the name of a loadable module. The name of the parameter within the module consists of the characters to the right of the colon. For example:

```
module\.\_name: variable
```

<table>
<thead>
<tr>
<th>Description</th>
<th>Briefly describes what the parameter does or controls.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>Indicates the signed or unsigned short integer or long integer. A long integer is twice the width in bits as an integer. For example, an unsigned integer = 32 bits, an unsigned long integer = 64 bits.</td>
</tr>
<tr>
<td>Units</td>
<td>(Optional) Describes the unit type.</td>
</tr>
<tr>
<td>Default</td>
<td>What the system uses as the default value.</td>
</tr>
<tr>
<td>Range</td>
<td>Specifies the possible range allowed by system validation or the bounds of the data type.</td>
</tr>
<tr>
<td></td>
<td>- MAXINT – A shorthand description for the maximum value of a signed integer (2,147,483,647)</td>
</tr>
<tr>
<td></td>
<td>- MAXUINT – A shorthand description for the maximum value of an unsigned integer (4,294,967,295)</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes, if the parameter can be changed on a running system with the mdb or kmdb debugger. No, if the parameter is a boot time initialization only.</td>
</tr>
<tr>
<td>Validation</td>
<td>Checks that the system applies to the value of the variable either as specified in the <code>/etc/system</code> file or the default value, as well as when the validation is applied.</td>
</tr>
<tr>
<td>Implicit</td>
<td>(Optional) Provides unstated constraints that might exist on the parameter, especially in relation to other parameters.</td>
</tr>
<tr>
<td>When to Change</td>
<td>Explains why someone might want to change this value. Includes error messages or return codes.</td>
</tr>
<tr>
<td>Zone Configuration</td>
<td>Identifies whether the parameter can be set in an exclusive-IP zone or must be set in the global zone. None of the parameters can be set in shared-IP zones.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Identifies the stability of the interface. Many of the parameters in this manual are still evolving and are classified as unstable. For more information, see attributes(5).</td>
</tr>
<tr>
<td>Change History</td>
<td>(Optional) Contains a link to the Change History appendix, if applicable.</td>
</tr>
</tbody>
</table>
Tuning the Oracle Solaris Kernel

The following table describes the different ways tunable parameters can be applied.

<table>
<thead>
<tr>
<th>Apply Tunable Parameters in These Ways</th>
<th>For More Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modify the /etc/system file</td>
<td>“/etc/system File” on page 21</td>
</tr>
<tr>
<td>Use the kernel debugger (kmdb)</td>
<td>“kmdb Command” on page 22</td>
</tr>
<tr>
<td>Use the modular debugger (mdb)</td>
<td>“mdb Command” on page 22</td>
</tr>
<tr>
<td>Use the ipadm command to set TCP/IP parameters</td>
<td>Chapter 4, “Internet Protocol Suite Tunable Parameters”</td>
</tr>
<tr>
<td>Modify the /etc/default files</td>
<td>“Tuning NCA Parameters” on page 157</td>
</tr>
</tbody>
</table>

/etc/system File

The /etc/system file provides a static mechanism for adjusting the values of kernel parameters. Values specified in this file are read at boot time and are applied. Any changes that are made to the file are not applied to the operating system until the system is rebooted.

One pass is made to set all the values before the configuration parameters are calculated.

Example—Setting a Parameter in /etc/system

The following /etc/system entry sets the ZFS ARC maximum (zfs_arc_max) to 30 GB.

```
set zfs:zfs_arc_max = 0x780000000
```

Recovering From an Incorrect Value

Make a copy of the /etc/system file before modifying it so that you can easily recover from incorrect value. For example:

```
# cp /etc/system /etc/system.good
```

If a value specified in the /etc/system file causes the system to become unbootable, you can recover with the following command:

```
ok boot -a
```

This command causes the system to ask for the name of various files used in the boot process. Press the Return key to accept the default values until the name of the /etc/system file is requested. When the Name of system file [/etc/system]: prompt is displayed, type the name of the good /etc/system file or /dev/null:
Name of system file [/etc/system]: /etc/system.good

If /dev/null is specified, this path causes the system to attempt to read from /dev/null for its configuration information. Because this file is empty, the system uses the default values. After the system is booted, the /etc/system file can be corrected.

For more information on system recovery, see Oracle Solaris Administration: Common Tasks.

### kmdb Command

kmdb is an interactive kernel debugger with the same general syntax as mdb. An advantage of interactive kernel debugger is that you can set breakpoints. When a breakpoint is reached, you can examine data or step through the execution of kernel code.

kmdb can be loaded and unloaded on demand. You do not have to reboot the system to perform interactive kernel debugging, as was the case with kadb.

For more information, see kmdb(1).

### mdb Command

The modular debugger, mdb, is unique among Solaris debuggers because it is easily extensible. A programming API is available that allows compilation of modules to perform desired tasks within the context of the debugger.

mdb also includes a number of desirable usability features, including command-line editing, command history, built-in output pager, syntax checking, and command pipelining. mdb is the recommended post-mortem debugger for the kernel.

For more information, see mdb(1).

### Example—Using mdb to Display Information

Display a high-level view of a system's memory usage. For example:

```
# mdb -k
Loading modules: [ unix genunix specfs dtrace mac cpu.generic cpu_ms.AuthenticAMD.15 uppc pcplusmp scsi vhci zfs mpt sd ip hook neti arp usbassocmds kssl qtc qlm qlm_stmf stmf_sdb md lofs random idm fcp crypto cpc smbsrv nfs fcip sppp ufs logindmux ptm nsmb scu mpt_sas pmcss emlxs ] > ::memstat
Page Summary Pages MB %Tot
----------------- ----------------- ----
Kernel 160876 628 16%
ZFS File Data 303401 1185 30%
Anon 25335 98 2%
Exec and libs 1459 5 0%
```
For more information on using the modular debugger, see the Oracle Solaris Modular Debugger Guide.

When using either kmdb or mdb debugger, the module name prefix is not required. After a module is loaded, its symbols form a common name space with the core kernel symbols and any other previously loaded module symbols.

**Special Oracle Solaris tune and var Structures**

Oracle Solaris tunable parameters come in a variety of forms. The tune structure defined in the /usr/include/sys/tuneable.h file is the runtime representation of tune_t_fsflushr, tune_t_minarmem, and tune_t_fltrec. After the kernel is initialized, all references to these variables are found in the appropriate field of the tune structure.

The proper way to set parameters for this structure at boot time is to initialize the special parameter that corresponds to the desired field name. The system initialization process then loads these values into the tune structure.

A second structure into which various tunable parameters are placed is the var structure named v. You can find the definition of a var structure in the /usr/include/sys/var.h file. The runtime representation of variables such as autoup and bufhwm is stored here.

Do not change either the tune or v structure on a running system. Changing any field in these structures on a running system might cause the system to panic.

**Viewing Oracle Solaris System Configuration Information**

Several tools are available to examine system configuration information. Some tools require superuser privilege. Other tools can be run by a non-privileged user. Every structure and data item can be examined with the kernel debugger by using mdb on a running system or by booting under kmdb.

For more information, see mdb(1) or kadb(1M).
**sysdef Command**

The `sysdef` command provides the values of memory and process resource limits, and portions of the `tune` and `v` structures. For example, the `sysdef "Tunable Parameters"` section from an x86 system with 8 GB of memory is as follows:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>171614208</td>
<td>maximum memory allowed in buffer cache (bufhwm)</td>
</tr>
<tr>
<td>30000</td>
<td>maximum number of processes (v.v proc)</td>
</tr>
<tr>
<td>90</td>
<td>maximum global priority in sys class (MAXCLSYSPRI)</td>
</tr>
<tr>
<td>29995</td>
<td>maximum processes per user id (v.v maxup)</td>
</tr>
<tr>
<td>30</td>
<td>auto update time limit in seconds (NAUTOUP)</td>
</tr>
<tr>
<td>25</td>
<td>page stealing low water mark (GPGSLO)</td>
</tr>
<tr>
<td>1</td>
<td>fsflush run rate (FSFLUSHR)</td>
</tr>
<tr>
<td>25</td>
<td>minimum resident memory for avoiding deadlock (MINARMEM)</td>
</tr>
<tr>
<td>25</td>
<td>minimum swapable memory for avoiding deadlock (MINASMEM)</td>
</tr>
</tbody>
</table>

For more information, see `sysdef(1M)`.

**kstat Utility**

`kstats` are data structures maintained by various kernel subsystems and drivers. They provide a mechanism for exporting data from the kernel to user programs without requiring that the program read kernel memory or have superuser privilege. For more information, see `kstat(1M)` or `kstat(3KSTAT)`.
This chapter describes most of the Oracle Solaris kernel tunable parameters.

- “General Kernel and Memory Parameters” on page 26
- “fsflush and Related Parameters” on page 32
- “Process-Sizing Parameters” on page 36
- “Paging-Related Parameters” on page 40
- “Swapping-Related Parameters” on page 51
- “Kernel Memory Allocator” on page 52
- “General Driver Parameters” on page 55
- “Network Driver Parameters” on page 57
- “General I/O Parameters” on page 62
- “General File System Parameters” on page 64
- “TMPFS Parameters” on page 66
- “Pseudo Terminals” on page 67
- “STREAMS Parameters” on page 70
- “System V Message Queues” on page 71
- “System V Semaphores” on page 72
- “System V Shared Memory” on page 72
- “Scheduling” on page 73
- “Timers” on page 74
- “SPARC System Specific Parameters” on page 75
- “Locality Group Parameters” on page 79

Where to Find Tunable Parameter Information

<table>
<thead>
<tr>
<th>Tunable Parameter</th>
<th>For Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFS tunable parameters</td>
<td>Chapter 3, “NFS Tunable Parameters”</td>
</tr>
</tbody>
</table>
General Kernel and Memory Parameters

This section describes general kernel parameters that are related to physical memory and stack configuration.

**physmem**

Description: Modifies the system’s configuration of the number of physical pages of memory after the Oracle Solaris OS and firmware are accounted for.

Data Type: Unsigned long

Default: Number of usable pages of physical memory available on the system, not counting the memory where the core kernel and data are stored

Range: 1 to amount of physical memory on system

Units: Pages

Dynamic?: No

Validation: None

When to Change: Whenever you want to test the effect of running the system with less physical memory. Because this parameter does not take into account the memory used by the core kernel and data, as well as various other data structures allocated early in the startup process, the value of physmem should be less than the actual number of pages that represent the smaller amount of memory.

Commitment Level: Unstable

**zfs_arc_min**

Description: Determines the minimum size of the ZFS Adaptive Replacement Cache (ARC). See also "zfs_arc_max" on page 27.

Data Type: Unsigned Integer (64-bit)
### zfs_arc_max

**Description**
Determined the maximum size of the ZFS Adaptive Replacement Cache (ARC). See also "zfs_arc_min" on page 26.

**Data Type**
Unsigned Integer (64-bit)

**Default**
Three-fourths of memory on systems with less than 4 GB of memory
physmem minus 1 GB on systems with greater than 4 GB of memory

**Range**
64 MB to physmem

**Units**
Bytes

**Dynamic?**
No

**Validation**
Yes, the range is validated.

**When to Change**
If a future memory requirement is significantly large and well defined, you might consider reducing the value of this parameter to cap the ARC so that it does not compete with the memory requirement. For example, if you know that a future workload requires 20% of memory, it makes sense to cap the ARC such that it does not consume more than the remaining 80% of memory.

**Commitment Level**
Unstable

**Change History**
For information, see "zfs_arc_max (Oracle Solaris 11)" on page 170.
default_stksize

Description  Specifies the default stack size of all threads. No thread can be created with a stack size smaller than default_stksize. If default_stksize is set, it overrides lwp_default_stksize. See also “lwp_default_stksize” on page 29.

Data Type Integer

Default

- 3 x PAGESIZE on SPARC systems
- 5 x PAGESIZE on x64 systems

Range Minimum is the default values:

- 3 x PAGESIZE on SPARC systems
- 5 x PAGESIZE on x64 systems

Maximum is 32 times the default value.

Units Bytes in multiples of the value returned by the getpagesize parameter. For more information, see getpagesize(3).

Dynamic? Yes. Affects threads created after the variable is changed.

Validation Must be greater than or equal to 8192 and less than or equal to 262,144 (256 x 1024). Also must be a multiple of the system page size. If these conditions are not met, the following message is displayed:

Illegal stack size, Using N

The value of N is the default value of default_stksize.

When to Change When the system panics because it has run out of stack space. The best solution for this problem is to determine why the system is running out of space and then make a correction.

Increasing the default stack size means that almost every kernel thread will have a larger stack, resulting in increased kernel memory consumption for no good reason. Generally, that space will be unused. The increased consumption means other resources that are competing for the same pool of memory will have the amount of space available to them reduced, possibly decreasing the system’s ability to perform work. Among the side effects is a reduction in the number of threads that the kernel can create. This solution should be treated as no more than an interim workaround until the root cause is remedied.

Commitment Level Unstable
lwp_default_stksize

Description  Specifies the default value of the stack size to be used when a kernel thread is created, and when the calling routine does not provide an explicit size to be used.

Data Type  Integer

Default  ■ 24,576 for SPARC platforms
         ■ 20,480 for x64 platforms

Range  Minimum is the default values:
         ■ 3 x PAGESIZE on SPARC systems
         ■ 5 x PAGESIZE on x64 systems

Maximum is 32 times the default value.

Units  Bytes in multiples of the value returned by the getpagesize parameter. For more information, see getpagesize(3C).

Dynamic?  Yes. Affects threads created after the variable is changed.

Validation  Must be greater than or equal to 8192 and less than or equal to 262,144 (256 x 1024). Also must be a multiple of the system page size. If these conditions are not met, the following message is displayed:

Illegal stack size, Using N

The value of N is the default value of lwp_default_stksize.

When to Change  When the system panics because it has run out of stack space. The best solution for this problem is to determine why the system is running out of space and then make a correction.

Increasing the default stack size means that almost every kernel thread will have a larger stack, resulting in increased kernel memory consumption for no good reason. Generally, that space will be unused. The increased consumption means other resources that are competing for the same pool of memory will have the amount of space available to them reduced, possibly decreasing the system’s ability to perform work. Among the side effects is a reduction in the number of threads that the kernel can create. This solution should be treated as no more than an interim workaround until the root cause is remedied.

Commitment Level  Unstable
### logevent_max_q_sz

**Description**

Maximum number of system events allowed to be queued and waiting for delivery to the syseventd daemon. Once the size of the system event queue reaches this limit, no other system events are allowed on the queue.

**Data Type**

Integer

**Default**

5000

**Range**

0 to MAXINT

**Units**

System events

**Dynamic?**

Yes

**Validation**

The system event framework checks this value every time a system event is generated by ddi_log_sysevent and sysevent_post_event.

For more information, see ddi_log_sysevent(9F) and sysevent_post_event(3SYSEVENT).

**When to Change**

When error log messages indicate that a system event failed to be logged, generated, or posted.

**Commitment Level**

Unstable

### segkpsize

**Description**

Specifies the amount of kernel pageable memory available. This memory is used primarily for kernel thread stacks. Increasing this number allows either larger stacks for the same number of threads or more threads. A system running a 64-bit kernel uses a default stack size of 24 KB.

**Data Type**

Unsigned long

**Default**

2 GB

**Range**

512 MB to 24 GB

**Units**

8-KB pages

**Dynamic?**

No

**Validation**

Value is compared to minimum and maximum sizes (512 MB and 24 GB). If smaller than the minimum or larger than the maximum, it is reset to 2 GB. A message to that effect is displayed.
The actual size used in creation of the cache is the lesser of the value specified in segkpsize after the validation checking or 50 percent of physical memory.

**When to Change**
Required to support large numbers of processes on a system. The default size of 2 GB, assuming at least 1 GB of physical memory is present. This default size allows creation of 24-KB stacks for more than 87,000 kernel threads. The size of a stack is the same, whether the process is a 32-bit process or a 64-bit process. If more than this number is needed, segkpsize can be increased, assuming sufficient physical memory exists.

**Commitment Level**
Unstable

---

**noexec_user_stack**

**Description**
Enables the stack to be marked as nonexecutable, which helps make buffer-overflow attacks more difficult.

An Oracle Solaris system running a 64-bit kernel makes the stacks of all 64-bit applications nonexecutable by default. Setting this parameter is necessary to make 32-bit applications nonexecutable.

**Data Type**
Signed integer

**Default**
0 (disabled)

**Range**
0 (disabled) or 1 (enabled)

**Units**
Toggle (on/off)

**Dynamic?**
Yes. Does not affect currently running processes, only processes created after the value is set.

**Validation**
None

**When to Change**
Should be enabled at all times unless applications are deliberately placing executable code on the stack without using mprotect to make the stack executable. For more information, see mprotect(2).

**Commitment Level**
Unstable
fsflush and Related Parameters

This section describes fsflush and related tunables.

fsflush

The system daemon, fsflush, runs periodically to do three main tasks:

1. On every invocation, fsflush flushes dirty file system pages over a certain age to disk.
2. On every invocation, fsflush examines a portion of memory and causes modified pages to be written to their backing store. Pages are written if they are modified and if they do not meet one of the following conditions:
   - Pages are kernel page
   - Pages are free
   - Pages are locked
   - Pages are associated with a swap device
   - Pages are currently involved in an I/O operation

The net effect is to flush pages from files that are mapped with mmap with write permission and that have actually been changed.

Pages are flushed to backing store but left attached to the process using them. This will simplify page reclamation when the system runs low on memory by avoiding delay for writing the page to backing store before claiming it, if the page has not been modified since the flush.

3. fsflush writes file system metadata to disk. This write is done every n-th invocation, where n is computed from various configuration variables. See “tune_t_fsflushr” on page 33 and “autoup” on page 33 for details.

The following features are configurable:

- Frequency of invocation (tune_t_fsflushr)
- Whether memory scanning is executed (dopageflush)
- Whether file system data flushing occurs (doiflush)
- The frequency with which file system data flushing occurs (autoup)

For most systems, memory scanning and file system metadata synchronizing are the dominant activities for fsflush. Depending on system usage, memory scanning can be of little use or consume too much CPU time.
**tune_t_fsflushr**

Description: Specifies the number of seconds between `fsflush` invocations.

Data Type: Signed integer

Default: 1

Range: 1 to MAXINT

Units: Seconds

Dynamic?: No

Validation: If the value is less than or equal to zero, the value is reset to 1 and a warning message is displayed. This check is done only at boot time.

When to Change: See the `autoup` parameter.

Commitment Level: Unstable

---

**autoup**

Description: Along with `tune_t_fsflushr`, `autoup` controls the amount of memory examined for dirty pages in each invocation and frequency of file system synchronizing operations.

The value of `autoup` is also used to control whether a buffer is written out from the free list. Buffers marked with the `B_DELWRI` flag (which identifies file content pages that have changed) are written out whenever the buffer has been on the list for longer than `autoup` seconds. Increasing the value of `autoup` keeps the buffers in memory for a longer time.

Data Type: Signed integer

Default: 30

Range: 1 to MAXINT

Units: Seconds

Dynamic?: No

Validation: If `autoup` is less than or equal to zero, it is reset to 30 and a warning message is displayed. This check is done only at boot time.
Implicit

autoup should be an integer multiple of tune_t_fsflushr. At a minimum, autoup should be at least 6 times the value of tune_t_fsflushr. If not, excessive amounts of memory are scanned each time fsflush is invoked.

The total system pages multiplied by tune_t_fsflushr should be greater than or equal to autoup to cause memory to be checked if dopageflush is non-zero.

When to Change

Here are several potential situations for changing autoup, tune_t_fsflushr, or both:

- Systems with large amounts of memory – In this case, increasing autoup reduces the amount of memory scanned in each invocation of fsflush.
- Systems with minimal memory demand – Increasing both autoup and tune_t_fsflushr reduces the number of scans made. autoup should be increased also to maintain the current ratio of autoup / tune_t_fsflushr.
- Systems with large numbers of transient files (for example, mail servers or software build machines) – If large numbers of files are created and then deleted, fsflush might unnecessarily write data pages for those files to disk.

Commitment Level

Unstable

dopageflush

Description

Controls whether memory is examined for modified pages during fsflush invocations. In each invocation of fsflush, the number of physical memory pages in the system is determined. This number might have changed because of a dynamic reconfiguration operation. Each invocation scans by using this algorithm: total number of pages x tune_t_fsflushr / autoup pages

Data Type

Signed integer

Default

1 (enabled)

Range

0 (disabled) or 1 (enabled)

Units

Toggle (on/off)

Dynamic?

Yes

Validation

None
### When to Change

If the system page scanner rarely runs, which is indicated by a value of 0 in the `sr` column of `vmstat` output.

### Commitment Level

Unstable

---

**doiflush**

**Description**

Controls whether file system metadata syncs will be executed during `fsflush` invocations. This synchronization is done every Nth invocation of `fsflush` where \( N = \frac{\text{autoup}}{\text{tune}_t_{fsflushr}} \). Because this algorithm is integer division, if \( \text{tune}_t_{fsflushr} \) is greater than \( \text{autoup} \), a synchronization is done on every invocation of `fsflush` because the code checks to see if its iteration counter is greater than or equal to \( N \). Note that \( N \) is computed once on invocation of `fsflush`. Later changes to \( \text{tune}_t_{fsflushr} \) or \( \text{autoup} \) have no effect on the frequency of synchronization operations.

**Data Type**

Signed integer

**Default**

1 (enabled)

**Range**

0 (disabled) or 1 (enabled)

**Units**

Toggle (on/off)

**Dynamic?**

Yes

**Validation**

None

**When to Change**

When files are frequently modified over a period of time and the load caused by the flushing perturbs system behavior.

Files whose existence, and therefore consistency of state, does not matter if the system reboots are better kept in a TMPFS file system (for example, `/tmp`). Inode traffic can be reduced on systems by using the `mount -o atime` option. This option eliminates inode updates when the file is accessed.

For a system engaged in real-time processing, you might want to disable this option and use explicit application file synchronizing to achieve consistency.

**Commitment Level**

Unstable
Process-Sizing Parameters

Several parameters (or variables) are used to control the number of processes that are available on the system and the number of processes that an individual user can create. The foundation parameter is \texttt{maxusers}. This parameter drives the values assigned to \texttt{max_nprocs} and \texttt{maxuprc}.

\textbf{maxusers}

\textbf{Description} Originally, \texttt{maxusers} defined the number of logged in users the system could support. When a kernel was generated, various tables were sized based on this setting. Current Oracle Solaris releases do much of its sizing based on the amount of memory on the system. Thus, much of the past use of \texttt{maxusers} has changed. A number of subsystems that are still derived from \texttt{maxusers}:

- The maximum number of processes on the system
- The number of quota structures held in the system
- The size of the directory name look-up cache (DNLC)

\textbf{DataType} Signed integer

\textbf{Default} Lesser of the amount of memory in MB or 2048

\textbf{Range} 1 to 2048, based on physical memory if not set in the \texttt{/etc/system} file

1 to 4096, if set in the \texttt{/etc/system} file

\textbf{Units} Users

\textbf{Dynamic?} No. After computation of dependent parameters is done, \texttt{maxusers} is never referenced again.

\textbf{Validation} None

\textbf{When to Change} When the default number of user processes derived by the system is too low. This situation is evident when the following message displays on the system console:

\texttt{out of processes}

You might also change this parameter when the default number of processes is too high, as in these situations:

- Database servers that have a lot of memory and relatively few running processes can save system memory when the default value of \texttt{maxusers} is reduced.
- If file servers have a lot of memory and few running processes, you might reduce this value. However, you should explicitly set the size of the DNLC. See “ncsize” on page 64.
- If compute servers have a lot of memory and few running processes, you might reduce this value.

**Commitment Level**  Unstable

### reserved_procs

**Description**
Specifies the number of system process slots to be reserved in the process table for processes with a UID of root (0). For example, \texttt{fsflush} has a UID of root (0).

**Data Type**  Signed integer

**Default**  5

**Range**  5 to MAXINT

**Units**  Processes

**Dynamic?**  No. Not used after the initial parameter computation.

**Validation**  Any \texttt{/etc/system} setting is honored.

**Commitment Level**  Unstable

**When to Change**
Consider increasing to 10 + the normal number of UID 0 (root) processes on system. This setting provides some cushion should it be necessary to obtain a root shell when the system is otherwise unable to create user-level processes.

### pidmax

**Description**
Specifies the value of the largest possible process ID.

\texttt{pidmax} sets the value for the \texttt{maxpid} variable. Once \texttt{maxpid} is set, \texttt{pidmax} is ignored. \texttt{maxpid} is used elsewhere in the kernel to determine the maximum process ID and for validation checking.

Any attempts to set \texttt{maxpid} by adding an entry to the \texttt{/etc/system} file have no effect.

**Data Type**  Signed integer

**Default**  30,000
## max_nprocs

**Description**
Specifies the maximum number of processes that can be created on a system. Includes system processes and user processes. Any value specified in `/etc/system` is used in the computation of `maxuprc`.

This value is also used in determining the size of several other system data structures. Other data structures where this parameter plays a role are as follows:

- Determining the size of the directory name lookup cache (if `ncsize` is not specified)
- Verifying that the amount of memory used by configured system V semaphores does not exceed system limits
- Configuring Hardware Address Translation resources for x86 platforms.

**Data Type**
Signed integer

**Default**
10 + (16 x `maxusers`)

**Range**
266 to value of `maxpid`

**Dynamic?**
No

**Validation**
Yes. The value is compared to `maxpid` and set to `maxpid` if it is larger. On x86 platforms, an additional check is made against a platform-specific value. `max_nprocs` is set to the smallest value in the triplet (`max_nprocs`, `maxpid`, platform value). Both SPARC and x86 platforms use 65,534 as the platform value.
When to Change  
Changing this parameter is one of the steps necessary to enable support for more than 30,000 processes on a system.

Commitment Level  
Unstable

**maxuprc**

Description  
Specifies the maximum number of processes that can be created on a system by any one user.

Data Type  
Signed integer

Default  
max_nprocs - reserved_procs

Range  
1 to max_nprocs - reserved_procs

Units  
Processes

Dynamic?  
No

Validation  
Yes. This value is compared to max_nprocs - reserved_procs and set to the smaller of the two values.

When to Change  
When you want to specify a hard limit for the number of processes a user can create that is less than the default value of however many processes the system can create. Attempting to exceed this limit generates the following warning messages on the console or in the messages file:

```
out of per-user processes for uid N
```

Commitment Level  
Unstable

**ngroups_max**

Description  
Specifies the maximum number of supplemental groups per process.

Data Type  
Signed integer

Default  
16

Range  
0 to 1024

Units  
Groups

Dynamic?  
No

Validation  
No

When to Change  
When you want to increase the maximum number of groups.
Keep in mind that if a particular user is assigned to more than 16 groups, the user might experience problems with AUTH_SYS credentials in an NFS environment.

Commitment Level Unstable

Paging-Related Parameters

The Solaris OS uses a demand paged virtual memory system. As the system runs, pages are brought into memory as needed. When memory becomes occupied above a certain threshold and demand for memory continues, paging begins. Paging goes through several levels that are controlled by certain parameters.

The general paging algorithm is as follows:

- A memory deficit is noticed. The page scanner thread runs and begins to walk through memory. A two-step algorithm is employed:
  1. A page is marked as unused.
  2. If still unused after a time interval, the page is viewed as a subject for reclaim.

If the page has been modified, a request is made to the pageout thread to schedule the page for I/O. Also, the page scanner continues looking at memory. Pageout causes the page to be written to the page’s backing store and placed on the free list. When the page scanner scans memory, no distinction is made as to the origin of the page. The page might have come from a data file, or it might represent a page from an executable’s text, data, or stack.

- As memory pressure on the system increases, the algorithm becomes more aggressive in the pages it will consider as candidates for reclamation and in how frequently the paging algorithm runs. (For more information, see "fastscan" on page 47 and "slowscan" on page 48.) As available memory falls between the range \texttt{lotsfree} and \texttt{minfree}, the system linearly increases the amount of memory scanned in each invocation of the pageout thread from the value specified by \texttt{slowscan} to the value specified by \texttt{fastscan}. The system uses the \texttt{desfree} parameter to control a number of decisions about resource usage and behavior.

The system initially constrains itself to use no more than 4 percent of one CPU for pageout operations. As memory pressure increases, the amount of CPU time consumed in support of pageout operations linearly increases until a maximum of 80 percent of one CPU is consumed. The algorithm looks through some amount of memory between \texttt{slowscan} and \texttt{fastscan}, then stops when one of the following occurs:

- Enough pages have been found to satisfy the memory shortfall.
- The planned number of pages have been looked at.
- Too much time has elapsed.
If a memory shortfall is still present when pageout finishes its scan, another scan is scheduled for 1/4 second in the future.

The configuration mechanism of the paging subsystem was changed. Instead of depending on a set of predefined values for fastscan, slowscan, and handspreadpages, the system determines the appropriate settings for these parameters at boot time. Setting any of these parameters in the /etc/system file can cause the system to use less than optimal values.

**Caution** – Remove all tuning of the VM system from the /etc/system file. Run with the default settings and determine if it is necessary to adjust any of these parameters. Do not set either cachefree or priority_paging.

Dynamic reconfiguration (DR) for CPU and memory is supported. A system in a DR operation that involves the addition or deletion of memory recalculates values for the relevant parameters, unless the parameter has been explicitly set in /etc/system. In that case, the value specified in /etc/system is used, unless a constraint on the value of the variable has been violated. In this case, the value is reset.

**lotsfree**

**Description**
Serves as the initial trigger for system paging to begin. When this threshold is crossed, the page scanner wakes up to begin looking for memory pages to reclaim.

**Data Type**
Unsigned long

**Default**
The greater of 1/64th of physical memory or 512 KB

**Range**
The minimum value is 512 KB or 1/64th of physical memory, whichever is greater, expressed as pages using the page size returned by getpagesize. For more information, see getpagesize(3C).

The maximum value is the number of physical memory pages. The maximum value should be no more than 30 percent of physical memory. The system does not enforce this range, other than that described in the Validation section.

**Units**
Pages

**Dynamic?**
Yes, but dynamic changes are lost if a memory-based DR operation occurs.

**Validation**
If lotsfree is greater than the amount of physical memory, the value is reset to the default.
Implicit

The relationship of `lotsfree` being greater than `desfree`, which is greater than `minfree`, should be maintained at all times.

When to Change

When demand for pages is subject to sudden sharp spikes, the memory algorithm might be unable to keep up with demand. One workaround is to start reclaiming memory at an earlier time. This solution gives the paging system some additional margin.

A rule of thumb is to set this parameter to 2 times what the system needs to allocate in a few seconds. This parameter is workload dependent. A DBMS server can probably work fine with the default settings. However, you might need to adjust this parameter for a system doing heavy file system I/O.

For systems with relatively static workloads and large amounts of memory, lower this value. The minimum acceptable value is 512 KB, expressed as pages using the page size returned by `getpagesize`.

Commitment Level

Unstable

---

### `desfree`

**Description**

Specifies the preferred amount of memory to be free at all times on the system.

**Data Type**

Unsigned integer

**Default**

`lotsfree / 2`

**Range**

The minimum value is 256 KB or 1/128th of physical memory, whichever is greater, expressed as pages using the page size returned by `getpagesize`.

The maximum value is the number of physical memory pages. The maximum value should be no more than 15 percent of physical memory. The system does not enforce this range other than that described in the Validation section.

**Units**

Pages

**Dynamic?**

Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided in the `/etc/system` file or calculated from the new physical memory value.

**Validation**

If `desfree` is greater than `lotsfree`, `desfree` is set to `lotsfree / 2`. No message is displayed.
**Implicit**
The relationship of \( \text{lots free} \) being greater than \( \text{des free} \), which is greater than \( \text{min free} \), should be maintained at all times.

**Side Effects**
Several side effects can arise from increasing the value of this parameter. When the new value nears or exceeds the amount of available memory on the system, the following can occur:
- Asynchronous I/O requests are not processed, unless available memory exceeds \( \text{des free} \). Increasing the value of \( \text{des free} \) can result in rejection of requests that otherwise would succeed.
- NFS asynchronous writes are executed as synchronous writes.
- The swapper is awakened earlier, and the behavior of the swapper is biased towards more aggressive actions.
- The system might not preload (prefault) as many executable pages as possible into the system. This side effect results in applications potentially running slower than they otherwise would.

**When to Change**
For systems with relatively static workloads and large amounts of memory, lower this value. The minimum acceptable value is 256 KB, expressed as pages using the page size returned by `getpagesize`.

**Commitment Level**
Unstable

---

**minfree**

**Description**
Specifies the minimum acceptable memory level. When memory drops below this number, the system biases allocations toward allocations necessary to successfully complete pageout operations or to swap processes completely out of memory. Either allocation denies or blocks other allocation requests.

**Data Type**
Unsigned integer

**Default**
\( \text{des free} / 2 \)

**Range**
The minimum value is 128 KB or \( 1/256 \)th of physical memory, whichever is greater, expressed as pages using the page size returned by `getpagesize`.

The maximum value is the number of physical memory pages. The maximum value should be no more than 7.5 percent of physical memory. The system does not enforce this range other than that described in the Validation section.

**Units**
Pages
| **Dynamic?** | Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided in the `/etc/system` file or calculated from the new physical memory value. |
| **Validation** | If `minfree` is greater than `desfree`, `minfree` is set to `desfree / 2`. No message is displayed. |
| **Implicit** | The relationship of `lotsfree` being greater than `desfree`, which is greater than `minfree`, should be maintained at all times. |
| **When to Change** | The default value is generally adequate. For systems with relatively static workloads and large amounts of memory, lower this value. The minimum acceptable value is 128 KB, expressed as pages using the page size returned by `getpagesize`. |
| **Commitment Level** | Unstable |

### throttlefree

**Description**

Specifies the memory level at which blocking memory allocation requests are put to sleep, even if the memory is sufficient to satisfy the request.

**Data Type**

Unsigned integer

**Default**

`minfree`

**Range**

The minimum value is 128 KB or 1/256th of physical memory, whichever is greater, expressed as pages using the page size returned by `getpagesize`.

The maximum value is the number of physical memory pages. The maximum value should be no more than 4 percent of physical memory. The system does not enforce this range other than that described in the Validation section.

**Units**

Pages

**Dynamic?**

Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided in the `/etc/system` file or calculated from the new physical memory value.

**Validation**

If `throttlefree` is greater than `desfree`, `throttlefree` is set to `minfree`. No message is displayed.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default</th>
<th>Range</th>
<th>Units</th>
<th>Dynamic?</th>
<th>Validation</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implicit</strong></td>
<td>The relationship of lotsfree is greater than desfree, which is greater than minfree, should be maintained at all times.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>When to Change</strong></td>
<td>The default value is generally adequate. For systems with relatively static workloads and large amounts of memory, lower this value. The minimum acceptable value is 128 KB, expressed as pages using the page size returned by getpagesize. For more information, see getpagesize(3C).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Commitment Level</strong></td>
<td>Unstable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pageout_reserve</strong></td>
<td>Specifies the number of pages reserved for the exclusive use of the pageout or scheduler threads. When available memory is less than this value, nonblocking allocations are denied for any processes other than pageout or the scheduler. Pageout needs to have a small pool of memory for its use so it can allocate the data structures necessary to do the I/O for writing a page to its backing store.</td>
<td>throttlefree/2</td>
<td>The minimum value is 64 KB or 1/512th of physical memory, whichever is greater, expressed as pages using the page size returned by getpagesize(3C).</td>
<td>Pages</td>
<td>Yes</td>
<td>If pageout_reserve is greater than throttlefree/2, pageout_reserve is set to throttlefree/2. No message is displayed.</td>
<td>The relationship of lotsfree being greater than desfree, which is greater than minfree, should be maintained at all times.</td>
</tr>
</tbody>
</table>
When to Change

The default value is generally adequate. For systems with relatively static workloads and large amounts of memory, lower this value. The minimum acceptable value is 64 KB, expressed as pages using the page size returned by `getpagesize`.

Commitment Level
Unstable

**pages_pp_maximum**

Description
Defines the number of pages that must be unlocked. If a request to lock pages would force available memory below this value, that request is refused.

Data Type
Unsigned long

Default
The greater of `(tune_t_minarmem + 100 and [4% of memory available at boot time + 4 MB])`

Range
Minimum value enforced by the system is `tune_t_minarmem + 100`. The system does not enforce a maximum value.

Units
Pages

Dynamic?
Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided in the `/etc/system` file or was calculated from the new physical memory value.

Validation
If the value specified in the `/etc/system` file or the calculated default is less than `tune_t_minarmem + 100`, the value is reset to `tune_t_minarmem + 100`.

No message is displayed if the value from the `/etc/system` file is increased. Validation is done only at boot time and during dynamic reconfiguration operations that involve adding or deleting memory.

When to Change
When memory-locking requests fail or when attaching to a shared memory segment with the `SHARE_MMU` flag fails, yet the amount of memory available seems to be sufficient.

Excessively large values can cause memory locking requests (`mlock`, `mlockall`, and `memcntl`) to fail unnecessarily. For more information, see `mlock(3C)`, `mlockall(3C)`, and `memcntl(2)`.

Commitment Level
Unstable
### tune_t_minarmem

**Description**
Defines the minimum available resident (not swappable) memory to maintain necessary to avoid deadlock. Used to reserve a portion of memory for use by the core of the OS. Pages restricted in this way are not seen when the OS determines the maximum amount of memory available.

**Data Type**
Signed integer

**Default**
25

**Range**
1 to physical memory

**Units**
Pages

**Dynamic?**
No

**Validation**
None. Large values result in wasted physical memory.

**When to Change**
The default value is generally adequate. Consider increasing the default value if the system locks up and debugging information indicates that no memory was available.

**Commitment Level**
Unstable

### fastscan

**Description**
Defines the maximum number of pages per second that the system looks at when memory pressure is highest.

**Data Type**
Signed integer

**Default**
The `fastscan` default value is set in one of the following ways:

- The `fastscan` value set in the `/etc/system` file is used.
- The `maxfastscan` value set in the `/etc/system` file is used.
- If neither `fastscan` nor `maxfastscan` is set in the `/etc/system` file, `fastscan` is set to 64 MB when the system is booted. Then, after the system is booted for a few minutes, the `fastscan` value is set to the number of pages that the scanner can scan in one second using 10% of a CPU.

In all three cases, if the derived value is more than half the memory in the system, the `fastscan` value is capped at the value of half the memory in the system.

**Range**
64 MB to half the system’s physical memory
Paging-Related Parameters

### Units Pages

**Dynamic?** Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided by `/etc/system` or calculated from the new physical memory value.

**Validation** The maximum value is the lesser of 64 MB and 1/2 of physical memory.

**When to Change** When more aggressive scanning of memory is preferred during periods of memory shortfall, especially when the system is subject to periods of intense memory demand or when performing heavy file I/O.

**Commitment Level** Unstable

### `slowscan`

**Description** Defines the minimum number of pages per second that the system looks at when attempting to reclaim memory.

**Data Type** Signed integer

**Default** The smaller of 1/20th of physical memory in pages and 100.

**Range** 1 to `fastscan` / 2

**Units** Pages

**Dynamic?** Yes, unless dynamic reconfiguration operations that add or delete memory occur. At that point, the value is reset to the value provided in the `/etc/system` file or calculated from the new physical memory value.

**Validation** If `slowscan` is larger than `fastscan` / 2, `slowscan` is reset to `fastscan` / 2. No message is displayed.

**When to Change** When more aggressive scanning of memory is preferred during periods of memory shortfall, especially when the system is subject to periods of intense memory demand.

**Commitment Level** Unstable

### `min_percent_cpu`

**Description** Defines the minimum percentage of CPU that pageout can consume. This parameter is used as the starting point for determining the maximum amount of time that can be consumed by the page scanner.

**Data Type** Signed integer
Default 4
Range 1 to 80
Units Percentage
Dynamic? Yes
Validation None
When to Change Increasing this value on systems with multiple CPUs and lots of memory, which are subject to intense periods of memory demand, enables the pager to spend more time attempting to find memory.
Commitment Level Unstable

**handspreadpages**

Description The Oracle Solaris OS uses a two-handed clock algorithm to look for pages that are candidates for reclaiming when memory is low. The first hand of the clock walks through memory marking pages as unused. The second hand walks through memory some distance after the first hand, checking to see if the page is still marked as unused. If so, the page is subject to being reclaimed. The distance between the first hand and the second hand is handspreadpages.

Data Type Unsigned long
Default fastscan
Range 1 to maximum number of physical memory pages on the system
Units Pages
Dynamic? Yes. This parameter requires that the kernel reset_hands parameter also be set to a non-zero value. Once the new value of handspreadpages has been recognized, reset_hands is set to zero.
Validation The value is set to the lesser of either the amount of physical memory and the handspreadpages value.
When to Change When you want to increase the amount of time that pages are potentially resident before being reclaimed. Increasing this value increases the separation between the hands, and therefore, the amount of time before a page can be reclaimed.
Commitment Level Unstable
Pages Before Pager

**Description**
Defines part of a system threshold that immediately frees pages after an I/O completes instead of storing the pages for possible reuse. The threshold is `lots free + pages before pager`. The NFS environment also uses this threshold to curtail its asynchronous activities as memory pressure mounts.

**Data Type**
Signed integer

**Default**
200

**Range**
1 to amount of physical memory

**Units**
Pages

**Dynamic?**
No

**Validation**
None

**When to Change**
You might change this parameter when the majority of I/O is done for pages that are truly read or written once and never referenced again. Setting this variable to a larger amount of memory keeps adding pages to the free list.

You might also change this parameter when the system is subject to bursts of severe memory pressure. A larger value here helps maintain a larger cushion against the pressure.

**Commitment Level**
Unstable

MaxPGIO

**Description**
Defines the maximum number of page I/O requests that can be queued by the paging system. This number is divided by 4 to get the actual maximum number used by the paging system. This parameter is used to throttle the number of requests as well as to control process swapping.

**Data Type**
Signed integer

**Default**
40

**Range**
1 to a variable maximum that depends on the system architecture, but mainly by the I/O subsystem, such as the number of controllers, disks, and disk swap size

**Units**
I/Os
Swapping-Related Parameters

Swapping in the Oracle Solaris OS is accomplished by the swapfs pseudo file system. The combination of space on swap devices and physical memory is treated as the pool of space available to support the system for maintaining backing store for anonymous memory. The system attempts to allocate space from disk devices first, and then uses physical memory as backing store. When swapfs is forced to use system memory for backing store, limits are enforced to ensure that the system does not deadlock because of excessive consumption by swapfs.

**swapfs_reserve**

<table>
<thead>
<tr>
<th>Description</th>
<th>Defines the amount of system memory that is reserved for use by system (UID = 0) processes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>Unsigned long</td>
</tr>
<tr>
<td>Default</td>
<td>The smaller of 4 MB and 1/16th of physical memory</td>
</tr>
<tr>
<td>Range</td>
<td>The minimum value is 4 MB or 1/16th of physical memory, whichever is smaller, expressed as pages using the page size returned by getpagesize. The maximum value is the number of physical memory pages. The maximum value should be no more than 10 percent of physical memory. The system does not enforce this range, other than that described in the Validation section.</td>
</tr>
</tbody>
</table>
### Units Pages

**Dynamic?** No

**Validation** None

**When to Change** Generally not necessary. Only change when recommended by a software provider, or when system processes are terminating because of an inability to obtain swap space. A much better solution is to add physical memory or additional swap devices to the system.

**Commitment Level** Unstable

#### swapfs_minfree

**Description** Defines the desired amount of physical memory to be kept free for the rest of the system. Attempts to reserve memory for use as swap space by any process that causes the system’s perception of available memory to fall below this value are rejected. Pages reserved in this manner can only be used for locked-down allocations by the kernel or by user-level processes.

**Data Type** Unsigned long

**Default** The larger of 2 MB and 1/8th of physical memory

**Range** 1 to amount of physical memory

**Units** Pages

**Dynamic?** No

**Validation** None

**When to Change** When processes are failing because of an inability to obtain swap space, yet the system has memory available.

**Commitment Level** Unstable

### Kernel Memory Allocator

The Oracle Solaris kernel memory allocator distributes chunks of memory for use by clients inside the kernel. The allocator creates a number of caches of varying size for use by its clients. Clients can also request the allocator to create a cache for use by that client (for example, to allocate structures of a particular size). Statistics about each cache that the allocator manages can be seen by using the `kstat -c kmem_cache` command.
Occasionally, systems might panic because of memory corruption. The kernel memory allocator supports a debugging interface (a set of flags), that performs various integrity checks on the buffers. The kernel memory allocator also collects information on the allocators. The integrity checks provide the opportunity to detect errors closer to where they actually occurred. The collected information provides additional data for support people when they try to ascertain the reason for the panic.

Use of the flags incurs additional overhead and memory usage during system operations. The flags should only be used when a memory corruption problem is suspected.

**kmem_flags**

**Description**  
The Oracle Solaris kernel memory allocator has various debugging and test options.

Five supported flag settings are described here.

<table>
<thead>
<tr>
<th>Flag</th>
<th>Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUDIT</td>
<td>0x1</td>
<td>The allocator maintains a log that contains recent history of its activity. The number of items logged depends on whether CONTENTS is also set. The log is a fixed size. When space is exhausted, earlier records are reclaimed.</td>
</tr>
<tr>
<td>TEST</td>
<td>0x2</td>
<td>The allocator writes a pattern into freed memory and checks that the pattern is unchanged when the buffer is next allocated. If some portion of the buffer is changed, then the memory was probably used by a client that had previously allocated and freed the buffer. If an overwrite is identified, the system panics.</td>
</tr>
<tr>
<td>REDZONE</td>
<td>0x4</td>
<td>The allocator provides extra memory at the end of the requested buffer and inserts a special pattern into that memory. When the buffer is freed, the pattern is checked to see if data was written past the end of the buffer. If an overwrite is identified, the kernel panics.</td>
</tr>
<tr>
<td>CONTENTS</td>
<td>0x8</td>
<td>The allocator logs up to 256 bytes of buffer contents when the buffer is freed. This flag requires that AUDIT also be set. The numeric value of these flags can be logically added together and set by the <code>/etc/system</code> file.</td>
</tr>
</tbody>
</table>
### Flag Setting Description

<table>
<thead>
<tr>
<th>Flag</th>
<th>Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LITE</td>
<td>0x100</td>
<td>Does minimal integrity checking when a buffer is allocated and freed. When enabled, the allocator checks that the redzone has not been written into, that a freed buffer is not being freed again, and that the buffer being freed is the size that was allocated. Do not combine this flag with any other flags.</td>
</tr>
</tbody>
</table>

**Data Type** Signed integer  
**Default** 0 (disabled)  
**Range** 0 (disabled) or 1 - 15 or 256 (0x100)  
**Dynamic?** Yes. Changes made during runtime only affect new kernel memory caches. After system initialization, the creation of new caches is rare.  
**Validation** None  
**When to Change** When memory corruption is suspected  
**Commitment Level** Unstable

### kmem_stackinfo

**Description** If the `kmem_stackinfo` variable is enabled in the `/etc/system` file at kernel thread creation time, the kernel thread stack is filled with a specific pattern instead of filled with zeros. During kernel thread execution, this kernel thread stack pattern is progressively overwritten. A simple count from the stack top until the pattern is not found gives a high watermark value, which is the maximum kernel stack space used by a kernel thread. This mechanism allows the following features:
- Compute the percentage of kernel thread stack really used (a high watermark) for current kernel threads in the system
- When a kernel thread ends, the system logs the last kernel threads that have used the most of their kernel thread stacks before dying to a small circular memory buffer

**Data Type** Unsigned integer  
**Default** 0 (disabled)  
**Range** 0 (disabled) or 1 (enabled)  
**Dynamic?** Yes
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>When you want to monitor kernel thread stack usage. Keep in mind that when kmem_stackinfo is enabled, the performance of creating and deleting kthreads is decreased. For more information, see the Chapter 5, &quot;Built-In Commands,&quot; in Oracle Solaris Modular Debugger Guide.</td>
</tr>
<tr>
<td>Zone Configuration</td>
<td>This parameter must be set in the global zone.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
</tr>
</tbody>
</table>

**General Driver Parameters**

**moddebug**

**Description**  When this parameter is enabled, messages about various steps in the module loading process are displayed.

**Data Type**  Signed integer

**Default**  0 (messages off)

**Range**  Here are the most useful values:

- 0x80000000 – Prints [un] loading... message. For every module loaded, messages such as the following appear on the console and in the /var/adm/messages file:

  Apr 20 17:18:04 neo genunix: [ID 943528 kern.notice] load 'sched/TS_DPTBL' id 15
  loaded @ 0x7be1b2f8/0x19c8380 size 176/2096
  Apr 20 17:18:04 neo genunix: [ID 131579 kern.notice] installing TS_DPTBL, module id 15.

- 0x40000000 – Prints detailed error messages. For every module loaded, messages such as the following appear on the console and in the /var/adm/messages file:

  Apr 20 18:30:00 neo unix: Errno = 2
  Apr 20 18:30:00 neo unix: kobj open: vn_open of /platform/sun4v/kernel/exec/sparcv9/intpexec fails
  Apr 20 18:30:00 neo unix: Errno = 2
  Apr 20 18:30:00 neo unix: kobj open: '/kernel/exec/sparcv9/intpexec'
  Apr 20 18:30:00 neo unix: vp = 60015777600
  Apr 20 18:30:00 neo unix: kobj close: 0x60015777600
  Apr 20 18:30:00 neo unix: kobj open: vn_open of /platform/SUNW,Sun-Fire-T200/kernel/exec/sparcv9
  /intpexec fails,
  Apr 20 18:30:00 neo unix: Errno = 2
  Apr 20 18:30:00 neo unix: kobj_open: vn_open of /platform/sun4v/kernel/exec/sparcv9/intpexec fails
- 0x20000000 - Prints even more detailed messages. This value doesn’t print any additional information beyond what the 0x40000000 flag does during system boot. However, this value does print additional information about releasing the module when the module is unloaded.

These values can be added together to set the final value.

Dynamic? Yes
Validation None
When to Change When a module is either not loading as expected, or the system seems to hang while loading modules. Note that when 0x40000000 is set, system boot is slowed down considerably by the number of messages written to the console.

Commitment Level Unstable

**ddi_msix_alloc_limit**

Description x86 only: This parameter controls the number of Extended Message Signaled Interrupts (MSI-X) that a device instance can allocate. Due to an existing system limitation, the default value is 2. You can increase the number of MSI-X interrupts that a device instance can allocate by increasing the value of this parameter. This parameter can be set either by editing the `/etc/system` file or by setting it with `/mdb` before the device driver attach occurs.

Data Type Signed integer
Default 2
Range 1 to 16
Dynamic? Yes
Validation None
When to Change To increase the number of MSI-X interrupts that a device instance can allocate. However, if you increase the number of MSI-X interrupts that a device instance can allocate, adequate interrupts might not be available to satisfy all allocation requests. If this happens, some devices might stop functioning or the system might fail to boot. Reduce the value or remove the parameter in this case.

Commitment Level Unstable
**Network Driver Parameters**

**igb Parameters**

**mr_enable**
- **Description**: This parameter enables or disables multiple receive and transmit queues that are used by the igb network driver. This parameter can be set by editing the `/etc/driver/drv/igb.conf` file before the igb driver attach occurs.
- **Data Type**: Boolean
- **Default**: 1 (disable multiple queues)
- **Range**: 0 (enable multiple queues) or 1 (disable multiple queues)
- **Dynamic?**: No
- **Validation**: None
- **When to Change**: To enable or disable multiple receive and transmit queues that are used by the igb network driver.
- **Commitment Level**: Unstable

**intr_force**
- **Description**: This parameter is used to force an interrupt type, such as MSI, MSI-X, or legacy, that is used by the igb network driver. This parameter can be set by editing the `/etc/driver/drv/igb.conf` file before the igb driver attach occurs.
- **Data Type**: Unsigned integer
- **Default**: 0 (do not force an interrupt type)
- **Range**: 0 (do not force an interrupt type)
  - 1 (force MSI-X interrupt type)
  - 2 (force MSI interrupt type)
  - 3 (force legacy interrupt type)
- **Dynamic?**: No
- **Validation**: None
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Data Type</th>
<th>Default</th>
<th>Range</th>
<th>Dynamic?</th>
<th>Validation</th>
<th>When to Change</th>
<th>Commitment Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>tx_queue_number</td>
<td>This parameter controls the number of transmit queues that are used by the ixgbe network driver. You can increase the number of transmit queues by increasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/igb.conf file before the ixgbe driver attach occurs.</td>
<td>Unsigned integer</td>
<td>8</td>
<td>1 to 32</td>
<td>No</td>
<td>None</td>
<td>To change the number of transmit queues that are used by the ixgbe network driver.</td>
<td>Unstable</td>
</tr>
<tr>
<td>rx_queue_number</td>
<td>This parameter controls the number of receive queues that are used by the ixgbe network driver. You can increase the number of receive queues by increasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/igb.conf file before the ixgbe driver attach occurs.</td>
<td>Unsigned integer</td>
<td>8</td>
<td>1 to 64</td>
<td>No</td>
<td>None</td>
<td>To change the number of receive queues that are used by the ixgbe network driver.</td>
<td>Unstable</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**intr_throttling**

**Description**
This parameter controls the interrupt throttling rate of the ixgbe network driver. You can increase the rate of interrupt by decreasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/ixgbe.conf file before the ixgbe driver attach occurs.

**Data Type**
Unsigned integer

**Default**
200

**Range**
0 to 65535

**Dynamic?**
No

**Validation**
None

**When to Change**
To change the interrupt throttling rate that is used by the ixgbe network driver.

<table>
<thead>
<tr>
<th>Commitment Level</th>
<th>Unstable</th>
</tr>
</thead>
</table>

**rx_limit_per_intr**

**Description**
This parameter controls the maximum number of receive queue buffer descriptors per interrupt that are used by the ixgbe network driver. You can increase the number of receive queue buffer descriptors by increasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/ixgbe.conf file before the ixgbe driver attach occurs.

**Data Type**
Unsigned integer

**Default**
256

**Range**
16 to 4096

**Dynamic?**
No

**Validation**
None

**When to Change**
To change the number of receive queue buffer descriptors that are handled per interrupt by the ixgbe network driver.

| Commitment Level | Unstable |
**tx_ring_size**

**Description**
This parameter controls the transmit queue size that is used by the ixgbe network driver. You can increase the transmit queue size by increasing the value of this parameter. This parameter can be set by editing the `/etc/driver/drv/ixgbe.conf` file before the ixgbe driver attach occurs.

**Data Type**
Unsigned integer

**Default**
1024

**Range**
64 to 4096

**Dynamic?**
No

**Validation**
None

**When to Change**
To change the transmit queue size that is used by the ixgbe network driver.

**Commitment Level**
Unstable

**rx_ring_size**

**Description**
This parameter controls the receive queue size that is used by the ixgbe network driver. You can increase the receive queue size by increasing the value of this parameter. This parameter can be set by editing the `/etc/driver/drv/ixgbe.conf` file before the ixgbe driver attach occurs.

**Data Type**
Unsigned integer

**Default**
1024

**Range**
64 to 4096

**Dynamic?**
No

**Validation**
None

**When to Change**
To change the receive queue size that is used by the ixgbe network driver.

**Commitment Level**
Unstable

**tx_copy_threshold**

**Description**
This parameter controls the transmit buffer copy threshold that is used by the ixgbe network driver. You can increase the transmit buffer copy
threshold by increasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/ixgbe.conf file before the ixgbe driver attach occurs.

**Data Type**  Unsigned integer  
**Default**  512  
**Range**  0 to 9126  
**Dynamic?**  No  
**Validation**  None  
**When to Change**  To change the transmit buffer copy threshold that is used by the ixgbe network driver.

**Commitment Level**  Unstable

### rx_copy_threshold

**Description**  This parameter controls the receive buffer copy threshold that is used by the ixgbe network driver. You can increase the receive buffer copy threshold by increasing the value of this parameter. This parameter can be set by editing the /etc/driver/drv/ixgbe.conf file before the ixgbe driver attach occurs.

**Data Type**  Unsigned integer  
**Default**  128  
**Range**  0 to 9126  
**Dynamic?**  No  
**Validation**  None  
**When to Change**  To change the receive buffer copy threshold that is used by the ixgbe network driver.

**Commitment Level**  Unstable
General I/O Parameters

**maxphys**

**Description**
Defines the maximum size of physical I/O requests. If a driver encounters a request larger than this size, the driver breaks the request into maxphys sized chunks. File systems can and do impose their own limit.

**Data Type**
Signed integer

**Default**
131,072 (sun4u or sun4v) or 57,344 (x86). The sd driver uses the value of 1,048,576 if the drive supports wide transfers. The ssd driver uses 1,048,576 by default.

**Range**
Machine-specific page size to MAXINT

**Units**
Bytes

**Dynamic?**
Yes, but many file systems load this value into a per-mount point data structure when the file system is mounted. A number of drivers load the value at the time a device is attached to a driver-specific data structure.

**Validation**
None

**When to Change**
When doing I/O to and from raw devices in large chunks. Note that a DBMS doing OLTP operations issues large numbers of small I/Os. Changing maxphys does not result in any performance improvement in that case.

**Commitment Level**
Unstable

**rlim_fd_max**

**Description**
Specifies the “hard” limit on file descriptors that a single process might have open. Overriding this limit requires superuser privilege.

**Data Type**
Signed integer

**Default**
65,536

**Range**
1 to MAXINT

**Units**
File descriptors

**Dynamic?**
No
Validation

None

When to Change

When the maximum number of open files for a process is not enough. Other limitations in system facilities can mean that a larger number of file descriptors is not as useful as it might be. For example:

- A 32-bit program using standard I/O is limited to 256 file descriptors. A 64-bit program using standard I/O can use up to 2 billion descriptors. Specifically, standard I/O refers to the `stdio(3C)` functions in `libc(3LIB)`.

- `select` is by default limited to 1024 descriptors per `fd_set`. For more information, see `select(3C)`. A 32-bit application code can be recompiled with a larger `fd_set` size (less than or equal to 65,536). A 64-bit application uses an `fd_set` size of 65,536, which cannot be changed.

An alternative to changing this on a system wide basis is to use the `plimit(1)` command. If a parent process has its limits changed by `plimit`, all children inherit the increased limit. This alternative is useful for daemons such as `inetd`.

Commitment Level

Unstable

**rlim_fd_cur**

Description

Defines the "soft" limit on file descriptors that a single process can have open. A process might adjust its file descriptor limit to any value up to the "hard" limit defined by `rlim_fd_max` by using the `setrlimit()` call or by issuing the `limit` command in whatever shell it is running. You do not require superuser privilege to adjust the limit to any value less than or equal to the hard limit.

Data Type

Signed integer

Default

256

Range

1 to `MAXINT`

Units

File descriptors

Dynamic?

No

Validation

Compared to `rlim_fd_max`. If `rlim_fd_cur` is greater than `rlim_fd_max`, `rlim_fd_cur` is reset to `rlim_fd_max`. 

Chapter 2 • Oracle Solaris Kernel Tunable Parameters 63
When to Change: When the default number of open files for a process is not enough. Increasing this value means only that it might not be necessary for a program to use setrlimit to increase the maximum number of file descriptors available to it.

Commitment Level: Unstable

**General File System Parameters**

**ncsize**

**Description**: Defines the number of entries in the directory name look-up cache (DNLC). This parameter is used by UFS, NFS, and ZFS to cache elements of pathname that have been resolved.

The DNLC also caches negative look-up information, which means it caches a name not found in the cache.

**Data Type**: Signed integer

**Default**: 
\[
(4 \times (v.v\_proc + maxusers) + 320) + (4 \times (v.v\_proc + maxusers) + 320) / 100
\]

**Range**: 0 to MAXINT

**Units**: DNLC entries

**Dynamic?**: No

**Validation**: None. Larger values cause the time it takes to unmount a file system to increase as the cache must be flushed of entries for that file system during the unmount process.

**When to Change**: You can use the kstat -n dnlcstats command to determine when entries have been removed from the DNLC because it was too small. The sum of the pick_heuristic and the pick_last parameters represents otherwise valid entries that were reclaimed because the cache was too small.

Excessive values of ncsize have an immediate impact on the system because the system allocates a set of data structures for the DNLC based on the value of ncsize. By default, a system allocates 64-byte structures for ncsize. The value has a further effect on UFS and NFS, unless ufs_ninode and nfs:ninode are explicitly set.

**Commitment Level**: Unstable
**dnlc_dir_enable**

Description: Enables large directory caching

---

**Note** – This parameter has no effect on NFS or ZFS file systems.

- **Data Type**: Unsigned integer
- **Default**: 1 (enabled)
- **Range**: 0 (disabled) or 1 (enabled)
- **Dynamic?**: Yes, but do not change this tunable dynamically. You can enable this parameter if it was originally disabled. Or, you can disable this parameter if it was originally enabled. However, enabling, disabling, and then enabling this parameter might lead to stale directory caches.
- **Validation**: No
- **When to Change**: Directory caching has no known problems. However, if problems occur, then set `dnlc_dir_enable` to 0 to disable caching.
- **Commitment Level**: Unstable

**dnlc_dir_min_size**

Description: Specifies the minimum number of entries cached for one directory.

---

**Note** – This parameter has no effect on NFS or ZFS file systems.

- **Data Type**: Unsigned integer
- **Default**: 40
- **Range**: 0 to MAXUINT (no maximum)
- **Units**: Entries
- **Dynamic?**: Yes, this parameter can be changed at any time.
- **Validation**: None
- **When to Change**: If performance problems occur with caching small directories, then increase `dnlc_dir_min_size`. Note that individual file systems might have their own range limits for caching directories.
### dnlc_dir_max_size

**Description**
Specifies the maximum number of entries cached for one directory.

**Note**
- This parameter has no effect on NFS or ZFS file systems.

**Data Type**
Unsigned integer

**Default**
MAXUINT (no maximum)

**Range**
0 to MAXUINT

**Dynamic?**
Yes, this parameter can be changed at any time.

**Validation**
None

**When to Change**
If performance problems occur with large directories, then decrease `dnlc_dir_max_size`.

**Commitment Level**
Unstable

---

### TMPFS Parameters

#### tmpfs:tmpfs_maxkmem

**Description**
Defines the maximum amount of kernel memory that TMPFS can use for its data structures (tmpnodes and directory entries).

**Data Type**
Unsigned long

**Default**
One page or 4 percent of physical memory, whichever is greater.

**Range**
Number of bytes in one page (8192 for sun4u or sun4v systems, 4096 for all other systems) to 25 percent of the available kernel memory at the time TMPFS was first used.

**Units**
Bytes

**Dynamic?**
Yes

**Validation**
None
When to Change

Increase if the following message is displayed on the console or written in the messages file:

```
tmp_memalloc: tmpfs over memory limit
```

The current amount of memory used by TMPFS for its data structures is held in the `tmp_kmemspace` field. This field can be examined with a kernel debugger.

Commitment Level

Unstable

**tmpfs:tmpfs_minfree**

Description

Defines the minimum amount of swap space that TMPFS leaves for the rest of the system.

Data Type

Signed long

Default

512

Range

0 to maximum swap space size

Units

Pages

Dynamic?

Yes

Validation

None

When to Change

To maintain a reasonable amount of swap space on systems with large amounts of TMPFS usage, you can increase this number. The limit has been reached when the console or messages file displays the following message:

```
fs-name: File system full, swap space limit exceeded
```

Commitment Level

Unstable

Pseudo Terminals

Pseudo terminals, ptys, are used for two purposes in Oracle Solaris software:

- Supporting remote logins by using the `telnet`, `rlogin`, or `rsh` commands
- Providing the interface through which the X Window system creates command interpreter windows

The default number of pseudo-terminals is sufficient for a desktop workstation. So, tuning focuses on the number of ptys available for remote logins.
The default number of ptys is now based on the amount of memory on the system. This default should be changed only to restrict or increase the number of users who can log in to the system.

Three related variables are used in the configuration process:

- **pt_cnt** – Default maximum number of ptys.
- **pt_pctofmem** – Percentage of kernel memory that can be dedicated to pty support structures. A value of zero means that no remote users can log in to the system.
- **pt_max_pty** – Hard maximum for number of ptys.

**pt_cnt** has a default value of zero, which tells the system to limit logins based on the amount of memory specified in **pt_pctofmem**, unless **pt_max_pty** is set. If **pt_cnt** is non-zero, ptys are allocated until this limit is reached. When that threshold is crossed, the system looks at **pt_max_pty**. If **pt_max_pty** has a non-zero value, it is compared to **pt_cnt**. The pty allocation is allowed if **pt_cnt** is less than **pt_max_pty**. If **pt_max_pty** is zero, **pt_cnt** is compared to the number of ptys supported based on **pt_pctofmem**. If **pt_cnt** is less than this value, the pty allocation is allowed. Note that the limit based on **pt_pctofmem** only comes into play if both **pt_cnt** and **ptms_pptymax** have default values of zero.

To put a hard limit on ptys that is different than the maximum derived from **pt_pctofmem**, set **pt_cnt** and **ptms_pptymax** in `/etc/system` to the preferred number of ptys. The setting of **ptms_pctofmem** is not relevant in this case.

To dedicate a different percentage of system memory to pty support and let the operating system manage the explicit limits, do the following:

- Do not set **pt_cnt** or **ptms_pptymax** in `/etc/system`.
- Set **pt_pctofmem** in `/etc/system` to the preferred percentage. For example, set **pt_pctofmem**=10 for a 10 percent setting.

Note that the memory is not actually allocated until it is used in support of a pty. Once memory is allocated, it remains allocated.

**pt_cnt**

**Description**

The number of available `/dev/pts` entries is dynamic up to a limit determined by the amount of physical memory available on the system. **pt_cnt** is one of three variables that determines the minimum number of logins that the system can accommodate. The default maximum number of `/dev/pts` devices the system can support is determined at boot time by computing the number of pty structures that can fit in a percentage of system memory (see **pt_pctofmem**). If **pt_cnt** is zero, the system allocates up to that maximum. If **pt_cnt** is non-zero, the system allocates to the greater of **pt_cnt** and the default maximum.
**pt_pctofmem**

**Description**
Specifies the maximum percentage of physical memory that can be consumed by data structures to support /dev/pts entries. A system consumes 176 bytes per /dev/pts entry.

**Data Type**
Unsigned integer

**Default**
5

**Range**
0 to 100

**Units**
Percentage

**Dynamic?**
No

**Validation**
None

**When to Change**
When you want to either restrict or increase the number of users who can log in to the system. A value of zero means that no remote users can log in to the system.

**Commitment Level**
Unstable

**pt_max_pty**

**Description**
Defines the maximum number of pty s the system offers

**Data Type**
Unsigned integer

**Default**
0 (Uses system-defined maximum)

**Range**
0 to MAXUINT
<table>
<thead>
<tr>
<th><strong>Units</strong></th>
<th>Logins/windows</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dynamic?</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Implicit</strong></td>
<td>Should be greater than or equal to <code>pt_cnt</code>. Value is not checked until the number of <code>ptys</code> allocated exceeds the value of <code>pt_cnt</code>.</td>
</tr>
<tr>
<td><strong>When to Change</strong></td>
<td>When you want to place an absolute ceiling on the number of logins supported, even if the system could handle more based on its current configuration values.</td>
</tr>
<tr>
<td><strong>Commitment Level</strong></td>
<td>Unstable</td>
</tr>
</tbody>
</table>

### STREAMS Parameters

#### nstrpush

- **Description**: Specifies the number of modules that can be inserted into (pushed onto) a STREAM.
- **Data Type**: Signed integer
- **Default**: 9
- **Range**: 9 to 16
- **Units**: Modules
- **Dynamic?**: Yes
- **Validation**: None
- **When to Change**: At the direction of your software vendor. No messages are displayed when a STREAM exceeds its permitted push count. A value of EINVAL is returned to the program that attempted the push.
- **Commitment Level**: Unstable

#### strmgszs

- **Description**: Specifies the maximum number of bytes that a single system call can pass to a STREAM to be placed in the data part of a message. Any write exceeding this size is broken into multiple messages. For more information, see `write(2)`.
System V Message Queues

System V message queues provide a message-passing interface that enables the exchange of messages by queues created in the kernel. Interfaces are provided in the Oracle Solaris environment to enqueue and dequeue messages. Messages can have a type associated with them. Enqueueing places messages at the end of a queue. Dequeueing removes the first message of a specific type from the queue or the first message if no type is specified.
For detailed information on tuning these system resources, see Chapter 6, "Resource Controls (Overview)," in Oracle Solaris Administration: Oracle Solaris Zones, Oracle Solaris 10 Zones, and Resource Management.

**System V Semaphores**

System V semaphores provide counting semaphores in the Oracle Solaris OS. A **semaphore** is a counter used to provide access to a shared data object for multiple processes. In addition to the standard set and release operations for semaphores, System V semaphores can have values that are incremented and decremented as needed (for example, to represent the number of resources available). System V semaphores also provide the ability to do operations on a group of semaphores simultaneously as well as to have the system undo the last operation by a process if the process dies.

**System V Shared Memory**

System V shared memory allows the creation of a segment by a process. Cooperating processes can attach to the memory segment (subject to access permissions on the segment) and gain access to the data contained in the segment. This capability is implemented as a loadable module. Entries in the `/etc/system` file must contain the `shmsys:` prefix.

A special kind of shared memory known as **intimate shared memory** (ISM) is used by DBMS vendors to maximize performance. When a shared memory segment is made into an ISM segment, the memory for the segment is locked. This feature enables a faster I/O path to be followed and improves memory usage. A number of kernel resources describing the segment are then shared between all processes that attach to the segment in ISM mode.

**segspt_minfree**

<table>
<thead>
<tr>
<th>Description</th>
<th>Identifies pages of system memory that cannot be allocated for ISM shared memory.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>Unsigned long</td>
</tr>
<tr>
<td>Default</td>
<td>5 percent of available system memory when the first ISM segment is created</td>
</tr>
<tr>
<td>Range</td>
<td>0 to 50 percent of physical memory</td>
</tr>
<tr>
<td>Units</td>
<td>Pages</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Validation

None. Values that are too small can cause the system to hang or performance to severely degrade when memory is consumed with ISM segments.

When to Change

On database servers with large amounts of physical memory using ISM, the value of this parameter can be decreased. If ISM segments are not used, this parameter has no effect. A maximum value of 128 MB (0x4000) is almost certainly sufficient on large memory machines.

Commitment Level

Unstable

Scheduling

disp_rechoose_interval

Description

Similar to the previous rechoose_interval parameter, this parameter specifies the amount of time before a process is deemed to have lost all affinity for the last CPU it ran on. However, this parameter is set in more granular time increments. This parameter should be used instead of the deprecated rechoose_interval parameter, but the rechoose_interval parameter is still accepted if it is set in the /etc/system file.

After this interval expires, any CPU is considered a candidate for scheduling a thread. This parameter does not apply to threads in the real-time class, but applies to threads in all other scheduling classes.

Use mdb if you want to change the value of this parameter by using the following steps:

1. Convert nanoseconds to unscaled time. For example, to convert a 5000000 nanosecond based value to unscaled time, use the following syntax:

   # mdb -kw
   .
   .
   > 0t5000000::time -u
   0xb6a444

2. Set disp_rechoose_interval to the unscaled time value. For example, provide the value that was returned in preceding step.

   > disp_rechoose_interval /Z 0xb6a444
disp_rechoose_interval: 0x447d998 = 0xb6a444
3. Verify that `disp_rechoose_interval` has been set to the right value. For example:

```bash
> disp_rechoose_interval::print
0xb6a444
```

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Signed integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>3</td>
</tr>
<tr>
<td>Range</td>
<td>0 to MAXINT</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>When caches are large, or when the system is running a critical process or a set of processes that seem to suffer from excessive cache misses not caused by data access patterns. Consider using the processor set capabilities or processor binding before changing this parameter. For more information, see <code>psrset(1M)</code> or <code>pbind(1M)</code>.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
</tr>
<tr>
<td>Change History</td>
<td>For information, see “<code>disp_rechoose_interval</code> (Oracle Solaris 11)” on page 170.</td>
</tr>
</tbody>
</table>

### Timers

#### hires_tick

**Description**

When set, this parameter causes the Oracle Solaris OS to use a system clock rate of 1000 instead of the default value of 100.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Signed integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0</td>
</tr>
<tr>
<td>Range</td>
<td>0 (disabled) or 1 (enabled)</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>When you want timeouts with a resolution of less than 10 milliseconds, and greater than or equal to 1 millisecond.</td>
</tr>
</tbody>
</table>
Commitment Level  Unstable

**timer_max**

Description  Specifies the number of POSIX timers available.
Data Type  Signed integer
Default  32
Range  0 to MAXINT
Dynamic?  No. Increasing the value can cause a system crash.
Validation  None
When to Change  When the default number of timers offered by the system is inadequate. Applications receive an EAGAIN error when executing timer_create system calls.

SPARC System Specific Parameters

The following parameters apply to sun4v and SPARC M-Series sun4u platforms.

**consistent_coloring**

Description  The ability to use different page placement policies on the UltraSPARC platform is available. A page placement policy attempts to allocate physical page addresses to maximize the use of the L2 cache. Whatever algorithm is chosen as the default algorithm, that algorithm can potentially provide less optimal results than another algorithm for a particular application set. This parameter changes the placement algorithm selected for all processes on the system.

Based on the size of the L2 cache, memory is divided into bins. The page placement code allocates a page from a bin when a page fault first occurs on an unmapped page. The page chosen depends on which of the three possible algorithms are used:
Page coloring – Various bits of the virtual address are used to determine the bin from which the page is selected. consistent_coloring is set to zero to use this algorithm. No per-process history exists for this algorithm.

Virtual addr=physical address – Consecutive pages in the program selects pages from consecutive bins. consistent_coloring is set to 1 to use this algorithm. No per-process history exists for this algorithm.

Bin-hopping – Consecutive pages in the program generally allocate pages from every other bin, but the algorithm occasionally skips more bins. consistent_coloring is set to 2 to use this algorithm. Each process starts at a randomly selected bin, and a per-process memory of the last bin allocated is kept.

Dynamic? Yes

Validation None. Values larger than 2 cause a number of WARNING: AS_2_BIN: bad consistent coloring value messages to appear on the console. The system hangs immediately thereafter. A power-cycle is required to recover.

When to Change When the primary workload of the system is a set of long-running high-performance computing (HPC) applications. Changing this value might provide better performance. File servers, database servers, and systems with a number of active processes (for example, compile or time sharing servers) do not benefit from changes.

Commitment Level Unstable

**tsb_alloc_hiwater_factor**

**Description**

Initializes tsb_alloc_hiwater to impose an upper limit on the amount of physical memory that can be allocated for translation storage buffers (TSBs) as follows:

\[
\text{tsb\_alloc\_hiwater} = \frac{\text{physical memory (bytes)}}{\text{tsb\_alloc\_hiwater\_factor}}
\]

When the memory that is allocated to TSBs is equal to the value of tsb_alloc_hiwater, the TSB memory allocation algorithm attempts to reclaim TSB memory as pages are unmapped.
Exercise caution when using this factor to increase the value of tsb_alloc_hiwater. To prevent system hangs, the resulting high water value must be considerably lower than the value of swapfs_minfree and segspt_minfree.

**default_tsb_size**

Description: Selects size of the initial translation storage buffers (TSBs) allocated to all processes.

Data Type: Integer

Default: Default is 0 (8 KB), which corresponds to 512 entries

Range: Possible values are:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8 KB</td>
</tr>
<tr>
<td>1</td>
<td>16 KB</td>
</tr>
<tr>
<td>3</td>
<td>32 KB</td>
</tr>
<tr>
<td>4</td>
<td>128 KB</td>
</tr>
<tr>
<td>5</td>
<td>256 KB</td>
</tr>
<tr>
<td>6</td>
<td>512 KB</td>
</tr>
</tbody>
</table>
### enable_tsb_rss_sizing

**Description**: Enables a resident set size (RSS) based TSB sizing heuristic.

**Data Type**: Boolean

**Default**: 1 (TSBs can be resized)

**Range**: 0 (TSBs remain at \(\text{tsb\_default\_size}\)) or 1 (TSBs can be resized)

If set to 0, then \(\text{tsb\_rss\_factor}\) is ignored.

**Dynamic?**: Yes

**Validation**: Yes

**When to Change**: Can be set to 0 to prevent growth of the TSBs. Under most circumstances, this parameter should be left at the default setting.

**Commitment Level**: Unstable

### tsb_rss_factor

**Description**: Controls the RSS to TSB span ratio of the RSS sizing heuristic. This factor divided by 512 yields the percentage of the TSB span which must be resident in memory before the TSB is considered as a candidate for resizing.

**Data Type**: Integer
Default 384, resulting in a value of 75%. Thus, when the TSB is 3/4 full, its size will be increased. Note that some virtual addresses typically map to the same slot in the TSB. Therefore, conflicts can occur before the TSB is at 100% full.

Range 0 to 512

Dynamic? Yes

Validation None

When to Change If the system is experiencing an excessive number of traps due to TSB misses, for example, due to virtual address conflicts in the TSB, you might consider decreasing this value toward 0.

For example, changing \texttt{tsb\_rss\_factor} to 256 (effectively, 50%) instead of 384 (effectively, 75%) might help eliminate virtual address conflicts in the TSB in some cases, but will use more kernel memory, particularly on a heavily loaded system.

TSB activity can be monitored with the \texttt{trapstat -T} command.

Commitment Level Unstable

### Locality Group Parameters

This section provides generic memory tunables, which apply to any SPARC or x86 system that uses a Non-Uniform Memory Architecture (NUMA).

#### lpg\_alloc\_prefer

**Description** Controls a heuristic for allocation of large memory pages when the requested page size is not immediately available in the local memory group, but could be satisfied from a remote memory group.

By default, the Oracle Solaris OS allocates a remote large page if local free memory is fragmented, but remote free memory is not. Setting this parameter to 1 indicates that additional effort should be spent attempting to allocate larger memory pages locally, potentially moving smaller pages around to coalesce larger pages in the local memory group.

**Data Type** Boolean
Default 0 (Prefer remote allocation if local free memory is fragmented and remote free memory is not)

Range 0 (Prefer remote allocation if local free memory is fragmented and remote free memory is not)

1 (Prefer local allocation whenever possible, even if local free memory is fragmented and remote free memory is not)

Dynamic? No

Validation None

When to Change This parameter might be set to 1 if long-running programs on the system tend to allocate memory that is accessed by a single program, or if memory that is accessed by a group of programs is known to be running in the same locality group (lgroup). In these circumstances, the extra cost of page coalesce operations can be amortized over the long run of the programs.

This parameter might be left at the default value (0) if multiple programs tend to share memory across different locality groups, or if pages tend to be used for short periods of time. In these circumstances, quick allocation of the requested size tends to be more important than allocation in a particular location.

Page locations and sizes might be observed by using the NUMA observability tools, available at http://hub.opensolaris.org/bin/view/Main/. TLB miss activity might be observed by using the trapstat -T command.

Commitment Level Uncommitted

**lgrp_mem_default_policy**

Description This variable reflects the default memory allocation policy used by the Oracle Solaris OS. This variable is an integer, and its value should correspond to one of the policies listed in the sys/lgrp.h file.

Data Type Integer

Default 1, LGRP_MEM_POLICY_NEXT indicating that memory allocation defaults to the home lgroup of the thread performing the memory allocation.

Range Possible values are:
<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>LGRP_MEM_POLICY_DEFAULT</td>
<td>use system default policy</td>
</tr>
<tr>
<td>1</td>
<td>LGRP_MEM_POLICY_NEXT</td>
<td>next to allocating thread’s home lgroup</td>
</tr>
<tr>
<td>2</td>
<td>LGRP_MEM_POLICY_RANDOM_PROC</td>
<td>randomly across process</td>
</tr>
<tr>
<td>3</td>
<td>LGRP_MEM_POLICY_RANDOM_PSET</td>
<td>randomly across processor set</td>
</tr>
<tr>
<td>4</td>
<td>LGRP_MEM_POLICY_RANDOM</td>
<td>randomly across all lgroups</td>
</tr>
<tr>
<td>5</td>
<td>LGRP_MEM_POLICY_ROUNDROBIN</td>
<td>round robin across all lgroups</td>
</tr>
<tr>
<td>6</td>
<td>LGRP_MEM_POLICY_NEXT_CPU</td>
<td>near next CPU to touch memory</td>
</tr>
</tbody>
</table>

**Dynamic?** No  
**Validation** None  
**When to Change** For applications that are sensitive to memory latencies due to allocations that occur from remote versus local memory on systems that use NUMA.  
**Commitment Level** Uncommitted  

**lgrp_mem_pset_aware**  
**Description** If a process is running within a user processor set, this variable determines whether randomly placed memory for the process is selected from among all the lgroups in the system or only from those lgroups that are spanned by the processors in the processor set.  
For more information about creating processor sets, see `psrset(1M)`.  
**Data Type** Boolean  
**Default** 0, the Oracle Solaris OS selects memory from all the lgroups in the system  
**Range**  
- 0, the Oracle Solaris OS selects memory from all the lgroups in the system (default)  
- 1, try selecting memory only from those lgroups that are spanned by the processors in the processor set. If the first attempt fails, memory can be allocated in any lgroup.
<table>
<thead>
<tr>
<th>Dynamic?</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>Setting this value to a value of one (1) might lead to more reproducible performance when processor sets are used to isolate applications from one another.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Uncommitted</td>
</tr>
</tbody>
</table>
This section describes the NFS tunable parameters.

- “Tuning the NFS Environment” on page 83
- “NFS Module Parameters” on page 84
- “rpcmod Module Parameters” on page 111

Where to Find Tunable Parameter Information

<table>
<thead>
<tr>
<th>Tunable Parameter</th>
<th>For Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Solaris kernel tunables</td>
<td>Chapter 2, “Oracle Solaris Kernel Tunable Parameters”</td>
</tr>
<tr>
<td>Internet Protocol Suite tunable parameters</td>
<td>Chapter 4, “Internet Protocol Suite Tunable Parameters”</td>
</tr>
<tr>
<td>Network Cache and Accelerator (NCA) tunable parameters</td>
<td>Chapter 5, “Network Cache and Accelerator Tunable Parameters”</td>
</tr>
</tbody>
</table>

Tuning the NFS Environment

You can define NFS parameters in the `/etc/system` file, which is read during the boot process. Each parameter includes the name of its associated kernel module. For more information, see “Tuning an Oracle Solaris System” on page 18.
Caution – The names of the parameters, the modules that they reside in, and the default values can change between releases. Check the documentation for the version of the active SunOS release before making changes or applying values from previous releases.

NFS Module Parameters

This section describes parameters related to the NFS kernel module.

nfs:nfs3_pathconf_disable_cache

Description Controls the caching of `pathconf` information for NFS Version 3 mounted file systems.

Data Type Integer (32-bit)

Default 0 (caching enabled)

Range 0 (caching enabled) or 1 (caching disabled)

Units Boolean values

Dynamic? Yes

Validation None

When to Change The `pathconf` information is cached on a per file basis. However, if the server can change the information for a specific file dynamically, use this parameter to disable caching. There is no mechanism for the client to validate its cache entry.

Commitment Level Unstable

nfs:nfs4_pathconf_disable_cache

Description Controls the caching of `pathconf` information for NFS Version 4 mounted file systems.

Data Type Integer (32-bit)

Default 0 (caching enabled)

Range 0 (caching enabled) or 1 (caching disabled)

Units Boolean values

Dynamic? Yes
**Validation**  None

**When to Change**  The `pathconf` information is cached on a per file basis. However, if the server can change the information for a specific file dynamically, use this parameter to disable caching. There is no mechanism for the client to validate its cache entry.

**Commitment Level**  Unstable

### nfs:nfs_allow_preepoch_time

**Description**  Controls whether files with incorrect or negative time stamps should be made visible on the client.

Historically, neither the NFS client nor the NFS server would do any range checking on the file times being returned. The over-the-wire timestamp values are unsigned and 32-bits long. So, all values have been legal.

The timestamp values on the 64-bit Solaris kernel are signed and 64-bits long. It is impossible to determine whether a time field represents a full 32-bit time or a negative time, that is, a time prior to January 1, 1970.

It is impossible to determine whether to sign extend a time value when converting from 32 bits to 64 bits. The time value should be sign extended if the time value is truly a negative number. However, the time value should not be sign extended if it does truly represent a full 32-bit time value. This problem is resolved by simply disallowing full 32-bit time values.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0 (32-bit time stamps disabled)</td>
</tr>
<tr>
<td>Range</td>
<td>0 (32-bit time stamps disabled) or 1 (32-bit time stamps enabled)</td>
</tr>
<tr>
<td>Units</td>
<td>Boolean values</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>Even during normal operation, it is possible for the timestamp values on some files to be set very far in the future or very far in the past. If access to these files is preferred using NFS mounted file systems, set this parameter to 1 to allow the timestamp values to be passed through unchecked.</td>
</tr>
</tbody>
</table>
### Commitment Level
Unstable

### nfs:nfs\_cots\_timeo

**Description**
Controls the default RPC timeout for NFS version 2 mounted file systems using connection-oriented transports such as TCP for the transport protocol.

**Data Type**
Signed integer (32-bit)

**Default**
600 (60 seconds)

**Range**
0 to \(2^{31} - 1\)

**Units**
10th of seconds

**Dynamic?**
Yes, but the RPC timeout for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**
None

**When to Change**
TCP does a good job ensuring requests and responses are delivered appropriately. However, if the round-trip times are very large in a particularly slow network, the NFS version 2 client might time out prematurely.

Increase this parameter to prevent the client from timing out incorrectly. The range of values is very large, so increasing this value too much might result in situations where a retransmission is not detected for long periods of time.

### nfs:nfs3\_cots\_timeo

**Description**
Controls the default RPC timeout for NFS version 3 mounted file systems using connection-oriented transports such as TCP for the transport protocol.

**Data Type**
Signed integer (32-bit)

**Default**
600 (60 seconds)

**Range**
0 to \(2^{31} - 1\)

**Units**
10th of seconds
### nfs:nfs4_cots_timeo

**Description**
Controls the default RPC timeout for NFS version 4 mounted file systems using connection-oriented transports such as TCP for the transport protocol.

The NFS Version 4 protocol specification disallows retransmission over the same TCP connection. Thus, this parameter primarily controls how quickly the client responds to certain events, such as detecting a forced unmount operation or detecting how quickly the server fails over to a new server.

**Data Type**
Signed integer (32-bit)

**Default**
600 (60 seconds)

**Range**
0 to $2^{31} - 1$

**Units**
10th of seconds

**Dynamic?**
Yes, but this parameter is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**
None

**When to Change**
TCP does a good job ensuring requests and responses are delivered appropriately. However, if the round-trip times are very large in a particularly slow network, the NFS version 4 client might time out prematurely.
Increase this parameter to prevent the client from timing out incorrectly. The range of values is very large, so increasing this value too much might result in situations where a retransmission is not detected for long periods of time.

**Commitment Level**

Unstable

### nfs:nfs_do_symlink_cache

**Description**

Controls whether the contents of symbolic link files are cached for NFS version 2 mounted file systems.

**Data Type**

Integer (32-bit)

**Default**

1 (caching enabled)

**Range**

0 (caching disabled) or 1 (caching enabled)

**Units**

Boolean values

**Dynamic?**

Yes

**Validation**

None

**When to Change**

If a server changes the contents of a symbolic link file without updating the modification timestamp on the file or if the granularity of the timestamp is too large, then changes to the contents of the symbolic link file might not be visible on the client for extended periods. In this case, use this parameter to disable the caching of symbolic link contents. Doing so makes the changes immediately visible to applications running on the client.

**Commitment Level**

Unstable

### nfs:nfs3_do_symlink_cache

**Description**

Controls whether the contents of symbolic link files are cached for NFS version 3 mounted file systems.

**Data Type**

Integer (32-bit)

**Default**

1 (caching enabled)

**Range**

0 (caching disabled) or 1 (caching enabled)

**Units**

Boolean values

**Dynamic?**

Yes
Validation None
When to Change If a server changes the contents of a symbolic link file without updating the modification timestamp on the file or if the granularity of the timestamp is too large, then changes to the contents of the symbolic link file might not be visible on the client for extended periods. In this case, use this parameter to disable the caching of symbolic link contents. Doing so makes the changes immediately visible to applications running on the client.

Commitment Level Unstable

**nfs:nfs4_do_symlink_cache**

Description Controls whether the contents of symbolic link files are cached for NFS version 4 mounted file systems.

Data Type Integer (32-bit)
Default 1 (caching enabled)
Range 0 (caching disabled) or 1 (caching enabled)
Units Boolean values
Dynamic? Yes
Validation None
When to Change If a server changes the contents of a symbolic link file without updating the modification timestamp on the file or if the granularity of the timestamp is too large, then changes to the contents of the symbolic link file might not be visible on the client for extended periods. In this case, use this parameter to disable the caching of symbolic link contents. Doing so makes the changes immediately visible to applications running on the client.

Commitment Level Unstable

**nfs:nfs_dynamic**

Description Controls whether a feature known as *dynamic retransmission* is enabled for NFS version 2 mounted file systems using connectionless transports such as UDP. This feature attempts to reduce retransmissions by monitoring server response times and then adjusting RPC timeouts and read- and write- transfer sizes.
### nfs:nfs3_dynamic

**Description**
Controls whether a feature known as *dynamic retransmission* is enabled for NFS version 3 mounted file systems using connectionless transports such as UDP. This feature attempts to reduce retransmissions by monitoring server response times and then adjusting RPC timeouts and read- and write- transfer sizes.

**Data Type**
Integer (32-bit)

**Default**
1 (enabled)

**Range**
0 (disabled) or 1 (enabled)

**Dynamic?**
Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**
None

**When to Change**
Do not change this parameter.

**Commitment Level**
Unstable

### nfs:nfs_lookup_neg_cache

**Description**
Controls whether a negative name cache is used for NFS version 2 mounted file systems. This negative name cache records file names that
were looked up, but not found. The cache is used to avoid over-the-network look-up requests made for file names that are already known to not exist.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>1 (enabled)</td>
</tr>
<tr>
<td>Range</td>
<td>0 (disabled) or 1 (enabled)</td>
</tr>
<tr>
<td>Units</td>
<td>Boolean values</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>For the cache to perform correctly, negative entries must be strictly verified before they are used. This consistency mechanism is relaxed slightly for read-only mounted file systems. It is assumed that the file system on the server is not changing or is changing very slowly, and that it is okay for such changes to propagate slowly to the client. The consistency mechanism becomes the normal attribute cache mechanism in this case. If file systems are mounted read-only on the client, but are expected to change on the server and these changes need to be seen immediately by the client, use this parameter to disable the negative cache. If you disable the <code>nfs:nfs_disable_rddir_cache</code> parameter, you should probably also disable this parameter. For more information, see “nfs:nfs_disable_rddir_cache” on page 101.</td>
</tr>
</tbody>
</table>

Commitment Level Unstable

**nfs:nfs3_lookup_neg_cache**

Description Controls whether a negative name cache is used for NFS version 3 mounted file systems. This negative name cache records file names that were looked up, but were not found. The cache is used to avoid over-the-network look-up requests made for file names that are already known to not exist.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>1 (enabled)</td>
</tr>
<tr>
<td>Range</td>
<td>0 (disabled) or 1 (enabled)</td>
</tr>
<tr>
<td>Units</td>
<td>Boolean values</td>
</tr>
</tbody>
</table>
nfs:nfs4_lookup_neg_cache

Description: Controls whether a negative name cache is used for NFS version 4 mounted file systems. This negative name cache records file names that were looked up, but were not found. The cache is used to avoid over-the-network look-up requests made for file names that are already known to not exist.

Data Type: Integer (32-bit)

Default: 1 (enabled)

Range: 0 (disabled) or 1 (enabled)

Units: Boolean values

Dynamic? Yes

Validation None

When to Change: For the cache to perform correctly, negative entries must be strictly verified before they are used. This consistency mechanism is relaxed slightly for read-only mounted file systems. It is assumed that the file system on the server is not changing or is changing very slowly, and that it is okay for such changes to propagate slowly to the client. The consistency mechanism becomes the normal attribute cache mechanism in this case.

If file systems are mounted read-only on the client, but are expected to change on the server and these changes need to be seen immediately by the client, use this parameter to disable the negative cache.

If you disable the nfs:nfs_disable_rddir_cache parameter, you should probably also disable this parameter. For more information, see “nfs:nfs_disable_rddir_cache” on page 101.

Commitment Level Unstable
that it is okay for such changes to propagate slowly to the client. The
consistency mechanism becomes the normal attribute cache
mechanism in this case.

If file systems are mounted read-only on the client, but are expected to
change on the server and these changes need to be seen immediately by
the client, use this parameter to disable the negative cache.

If you disable the `nfs:nfs_disable_rddir_cache` parameter, you
should probably also disable this parameter. For more information, see
“`nfs:nfs_disable_rddir_cache`” on page 101.

Commitment Level  Unstable

### `nfs:nfs_max_threads`

<table>
<thead>
<tr>
<th>Description</th>
<th>Controls the number of kernel threads that perform asynchronous I/O for the NFS version 2 client. Because NFS is based on RPC and RPC is inherently synchronous, separate execution contexts are required to perform NFS operations that are asynchronous from the calling thread. The operations that can be executed asynchronously are read for read-ahead, readdir for readdir read-ahead, write for putpage and pageio operations, commit, and inactive for cleanup operations that the client performs when it stops using a file.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>Integer (16-bit)</td>
</tr>
<tr>
<td>Default</td>
<td>8</td>
</tr>
<tr>
<td>Range</td>
<td>0 to $2^{15}$ - 1</td>
</tr>
<tr>
<td>Units</td>
<td>Threads</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>To increase or reduce the number of simultaneous I/O operations that are outstanding at any given time. For example, for a very low bandwidth network, you might want to decrease this value so that the NFS client does not overload the network. Alternately, if the network is very high bandwidth, and the client and server have sufficient</td>
</tr>
</tbody>
</table>
resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources.

Commitment Level Unstable

### nfs:nfs3_max_threads

**Description**
Controls the number of kernel threads that perform asynchronous I/O for the NFS version 3 client. Because NFS is based on RPC and RPC is inherently synchronous, separate execution contexts are required to perform NFS operations that are asynchronous from the calling thread.

The operations that can be executed asynchronously are read for read-ahead, readdir for readdir read-ahead, write for putpage and pageio requests, and commit.

**Data Type**  
Integer (16-bit)

**Default**  
8

**Range**  
0 to $2^{15} - 1$

**Units**  
Threads

**Dynamic?**  
Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**  
None

**When to Change**  
To increase or reduce the number of simultaneous I/O operations that are outstanding at any given time. For example, for a very low bandwidth network, you might want to decrease this value so that the NFS client does not overload the network. Alternately, if the network is very high bandwidth, and the client and server have sufficient resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources.

Commitment Level Unstable

### nfs:nfs4_max_threads

**Description**
Controls the number of kernel threads that perform asynchronous I/O for the NFS version 4 client. Because NFS is based on RPC and RPC is
inherently synchronous, separate execution contexts are required to perform NFS operations that are asynchronous from the calling thread.

The operations that can be executed asynchronously are read for read-ahead, write-behind, directory read-ahead, and cleanup operations that the client performs when it stops using a file.

| Data Type | Integer (16-bit) |
| Default | 8 |
| Range | 0 to \(2^{15} - 1\) |
| Units | Threads |
| Dynamic? | Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter. |
| Validation | None |
| When to Change | To increase or reduce the number of simultaneous I/O operations that are outstanding at any given time. For example, for a very low bandwidth network, you might want to decrease this value so that the NFS client does not overload the network. Alternately, if the network is very high bandwidth, and the client and server have sufficient resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources. |
| Commitment Level | Unstable |

**nfs:nfs_nra**

Description

Controls the number of read-ahead operations that are queued by the NFS version 2 client when sequential access to a file is discovered. These read-ahead operations increase concurrency and read throughput. Each read-ahead request is generally for one logical block of file data.

| Data Type | Integer (32-bit) |
| Default | 4 |
| Range | 0 to \(2^{31} - 1\) |
| Units | Logical blocks |
| Dynamic? | Yes |
Validation None
When to Change To increase or reduce the number of read-ahead requests that are outstanding for a specific file at any given time. For example, for a very low bandwidth network or on a low memory client, you might want to decrease this value so that the NFS client does not overload the network or the system memory. Alternately, if the network is very high bandwidth, and the client and server have sufficient resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources.

Commitment Level Unstable

**nfs:nfs3_nra**

Description Controls the number of read-ahead operations that are queued by the NFS version 3 client when sequential access to a file is discovered. These read-ahead operations increase concurrency and read throughput. Each read-ahead request is generally for one logical block of file data.

Data Type Integer (32-bit)
Default 4
Range 0 to $2^{31} - 1$
Units Logical blocks. (See "nfs:nfs3_bsize" on page 102.)
Dynamic? Yes
Validation None
When to Change To increase or reduce the number of read-ahead requests that are outstanding for a specific file at any given time. For example, for a very low bandwidth network or on a low memory client, you might want to decrease this value so that the NFS client does not overload the network or the system memory. Alternately, if the network is very high bandwidth and the client and server have sufficient resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources.

Commitment Level Unstable
### nfs:nfs4_nra

**Description**

Controls the number of read-ahead operations that are queued by the NFS version 4 client when sequential access to a file is discovered. These read-ahead operations increase concurrency and read throughput. Each read-ahead request is generally for one logical block of file data.

**Data Type**

Integer (32-bit)

**Default**

4

**Range**

0 to $2^{31} - 1$

**Units**

Logical blocks. (See "nfs:nfs4_bsize" on page 102.)

**Dynamic?**

Yes

**Validation**

None

**When to Change**

To increase or reduce the number of read-ahead requests that are outstanding for a specific file at any given time. For example, for a very low bandwidth network or on a low memory client, you might want to decrease this value so that the NFS client does not overload the network or the system memory. Alternatively, if the network is very high bandwidth, and the client and server have sufficient resources, you might want to increase this value. Doing so can more effectively utilize the available network bandwidth, and the client and server resources.

**Commitment Level**

Unstable

### nfs:nrnode

**Description**

Controls the size of the $r$node cache on the NFS client.

The $r$node, used by both NFS version 2, 3, and 4 clients, is the central data structure that describes a file on the NFS client. The $r$node contains the file handle that identifies the file on the server. The $r$node also contains pointers to various caches used by the NFS client to avoid network calls to the server. Each $r$node has a one-to-one association with a $v$node. The $v$node caches file data.

The NFS client attempts to maintain a minimum number of $r$nodes to attempt to avoid destroying cached data and metadata. When an $r$node is reused or freed, the cached data and metadata must be destroyed.

**Data Type**

Integer (32-bit)
**Default**
The default setting of this parameter is 0, which means that the value of `rnode` should be set to the value of the `ncsize` parameter. Actually, any non-positive value of `rnode` results in `rnode` being set to the value of `ncsize`.

**Range**
1 to $2^{31} - 1$

**Units**
rnodes

**Dynamic?**
No. This value can only be changed by adding or changing the parameter in the `/etc/system` file, and then rebooting the system.

**Validation**
The system enforces a maximum value such that the `rnode` cache can only consume 25 percent of available memory.

**When to Change**
Because `rnodes` are created and destroyed dynamically, the system tends to settle upon a `rnode-size` cache, automatically adjusting the size of the cache as memory pressure on the system increases or as more files are simultaneously accessed. However, in certain situations, you could set the value of `rnode` if the mix of files being accessed can be predicted in advance. For example, if the NFS client is accessing a few very large files, you could set the value of `rnode` to a small number so that system memory can cache file data instead of `rnodes`. Alternately, if the client is accessing many small files, you could increase the value of `rnode` to optimize for storing file metadata to reduce the number of network calls for metadata.

Although it is not recommended, the `rnode` cache can be effectively disabled by setting the value of `rnode` to 1. This value instructs the client to only cache 1 `rnode`, which means that it is reused frequently.

**Commitment Level**
Unstable

**nfs:nfs_shrinkreaddir**

**Description**
Some older NFS servers might incorrectly handle NFS version 2 `READDR` requests for more than 1024 bytes of directory information. This problem is due to a bug in the server implementation. However, this parameter contains a workaround in the NFS version 2 client.

When this parameter is enabled, the client does not generate a `READDR` request for larger than 1024 bytes of directory information. If this parameter is disabled, then the over-the-wire size is set to the lesser of either the size passed in by using the `getdents` system call or by using `NFS_MAXDATA`, which is 8192 bytes. For more information, see `getdents(2)`. 
**DataType** Integer (32-bit)
**Default** 0 (disabled)
**Range** 0 (disabled) or 1 (enabled)
**Units** Boolean values
**Dynamic?** Yes
**Validation** None
**When to Change** Examine the value of this parameter if an older NFS version 2 only server is used and interoperability problems occur when the server tries to read directories. Enabling this parameter might cause a slight decrease in performance for applications that read directories.

**Commitment Level** Unstable

### nfs:nfs3_shrinkreaddir

**Description** Some older NFS servers might incorrectly handle NFS version 3 READDIR requests for more than 1024 bytes of directory information. This problem is due to a bug in the server implementation. However, this parameter contains a workaround in the NFS version 3 client.

When this parameter is enabled, the client does not generate a READDIR request for larger than 1024 bytes of directory information. If this parameter is disabled, then the over-the-wire size is set to the minimum of either the size passed in by using the getdents system call or by using MAXBSIZE, which is 8192 bytes. For more information, see `getdents(2)`.

**Data Type** Integer (32-bit)
**Default** 0 (disabled)
**Range** 0 (disabled) or 1 (enabled)
**Units** Boolean values
**Dynamic?** Yes
**Validation** None
**When to Change** Examine the value of this parameter if an older NFS version 3 only server is used and interoperability problems occur when the server tries to read directories. Enabling this parameter might cause a slight decrease in performance for applications that read directories.
Commitment Level Unstable

**nfs:nfs_write_error_interval**

Description Controls the time duration in between logging ENOSPC and EDQUOT write errors received by the NFS client. This parameter affects NFS version 2, 3, and 4 clients.

Data Type Long integer (64-bit)

Default 5 seconds

Range 0 to $2^{63} - 1$

Units Seconds

Dynamic? Yes

Validation None

When to Change Increase or decrease the value of this parameter in response to the volume of messages being logged by the client. Typically, you might want to increase the value of this parameter to decrease the number of out of space messages being printed when a full file system on a server is being actively used.

Commitment Level Unstable

**nfs:nfs_write_error_to_cons_only**

Description Controls whether NFS write errors are logged to the system console and syslog or to the system console only. This parameter affects messages for NFS version 2, 3, and 4 clients.

Data Type Integer (32-bit)

Default 0 (system console and syslog)

Range 0 (system console and syslog) or 1 (system console)

Units Boolean values

Dynamic? Yes

Validation None

When to Change Examine the value of this parameter to avoid filling up the file system containing the messages logged by the syslogd daemon. When this
parameter is enabled, messages are printed on the system console only and are not copied to the syslog messages file.

Commitment Level  Unstable

**nfs:nfs_disable_rddir_cache**

**Description**  Controls the use of a cache to hold responses from READDIR and READDIRPLUS requests. This cache avoids over-the-wire calls to the server to retrieve directory information.

**Data Type**  Integer (32-bit)

**Default**  0 (caching enabled)

**Range**  0 (caching enabled) or 1 (caching disabled)

**Units**  Boolean values

**Dynamic?**  Yes

**Validation**  None

**When to Change**  Examine the value of this parameter if interoperability problems develop due to a server that does not update the modification time on a directory when a file or directory is created in it or removed from it. The symptoms are that new names do not appear in directory listings after they have been added to the directory or that old names do not disappear after they have been removed from the directory.

This parameter controls the caching for NFS version 2, 3, and 4 mounted file systems. This parameter applies to all NFS mounted file systems, so caching cannot be disabled or enabled on a per file system basis.

If you disable this parameter, you should also disable the following parameters to prevent bad entries in the DNLC negative cache:

- “nfs:nfs_lookup_neg_cache” on page 90
- “nfs:nfs3_lookup_neg_cache” on page 91
- “nfs:nfs4_lookup_neg_cache” on page 92

Commitment Level  Unstable
**nfs:nfs3_bsize**

**Description**
Controls the logical block size used by the NFS version 3 client. This block size represents the amount of data that the client attempts to read from or write to the server when it needs to do an I/O.

**Data Type**
Unsigned integer (32-bit)

**Default**
32,768 (32 KB)

**Range**
0 to $2^{31} - 1$

**Units**
Bytes

**Dynamic?**
Yes, but the block size for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**
None. Setting this parameter too low or too high might cause the system to malfunction. Do not set this parameter to anything less than PAGESIZE for the specific platform. Do not set this parameter too high because it might cause the system to hang while waiting for memory allocations to be granted.

**When to Change**
Examine the value of this parameter when attempting to change the maximum data transfer size. Change this parameter in conjunction with the nfs:nfs3_max_transfer_size parameter. If larger transfers are preferred, increase both parameters. If smaller transfers are preferred, then just reducing this parameter should suffice.

**Commitment Level**
Unstable

---

**nfs:nfs4_bsize**

**Description**
Controls the logical block size used by the NFS version 4 client. This block size represents the amount of data that the client attempts to read from or write to the server when it needs to do an I/O.

**Data Type**
Unsigned integer (32-bit)

**Default**
32,768 (32 KB)

**Range**
0 to $2^{31} - 1$

**Units**
Bytes

**Dynamic?**
Yes, but the block size for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.
Validation

None. Setting this parameter too low or too high might cause the system to malfunction. Do not set this parameter to anything less than PAGESIZE for the specific platform. Do not set this parameter too high because it might cause the system to hang while waiting for memory allocations to be granted.

When to Change

Examine the value of this parameter when attempting to change the maximum data transfer size. Change this parameter in conjunction with the nfs:nfs4_max_transfer_size parameter. If larger transfers are preferred, increase both parameters. If smaller transfers are preferred, then just reducing this parameter should suffice.

Commitment Level

Unstable

nfs:nfs_async_clusters

Description

Controls the mix of asynchronous requests that are generated by the NFS version 2 client. The four types of asynchronous requests are read-ahead, putpage, pageio, and readdir-ahead. The client attempts to round-robin between these different request types to attempt to be fair and not starve one request type in favor of another.

However, the functionality in some NFS version 2 servers such as write gathering depends upon certain behaviors of existing NFS Version 2 clients. In particular, this functionality depends upon the client sending out multiple WRITE requests at about the same time. If one request is taken out of the queue at a time, the client would be defeating this server functionality designed to enhance performance for the client.

Thus, use this parameter to control the number of requests of each request type that are sent out before changing types.

Data Type

Unsigned integer (32-bit)

Default

1

Range

0 to $2^{31} - 1$

Units

Asynchronous requests

Dynamic?

Yes, but the cluster setting for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.
Validation: None. However, setting the value of this parameter to 0 causes all of the queued requests of a particular request type to be processed before moving on to the next type. This effectively disables the fairness portion of the algorithm.

When to Change: To increase the number of each type of asynchronous request that is generated before switching to the next type. Doing so might help with server functionality that depends upon clusters of requests coming from the client.

Commitment Level: Unstable

**nfs:nfs3_async_clusters**

Description: Controls the mix of asynchronous requests that are generated by the NFS version 3 client. The five types of asynchronous requests are read-ahead, putpage, pageio, readdir-ahead, and commit. The client attempts to round-robin between these different request types to attempt to be fair and not starve one request type in favor of another.

However, the functionality in some NFS version 3 servers such as write gathering depends upon certain behaviors of existing NFS version 3 clients. In particular, this functionality depends upon the client sending out multiple WRITE requests at about the same time. If one request is taken out of the queue at a time, the client would be defeating this server functionality designed to enhance performance for the client.

Thus, use this parameter to control the number of requests of each request type that are sent out before changing types.

**Data Type**: Unsigned integer (32-bit)

**Default**: 1

**Range**: 0 to \(2^{31} - 1\)

**Units**: Asynchronous requests

**Dynamic?**: Yes, but the cluster setting for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation**: None. However, setting the value of this parameter to 0 causes all of the queued requests of a particular request type to be processed before moving on to the next type. This value effectively disables the fairness portion of the algorithm.
When to Change: To increase the number of each type of asynchronous operation that is generated before switching to the next type. Doing so might help with server functionality that depends upon clusters of operations coming from the client.

Commitment Level: Unstable

**nfs:nfs4_async_clusters**

**Description:** Controls the mix of asynchronous requests that are generated by the NFS version 4 client. The six types of asynchronous requests are read-ahead, putpage, pageio, readdir-ahead, commit, and inactive. The client attempts to round-robin between these different request types to attempt to be fair and not starve one request type in favor of another.

However, the functionality in some NFS version 4 servers such as write gathering depends upon certain behaviors of existing NFS version 4 clients. In particular, this functionality depends upon the client sending out multiple WRITE requests at about the same time. If one request is taken out of the queue at a time, the client would be defeating this server functionality designed to enhance performance for the client.

Thus, use this parameter to control the number of requests of each request type that are sent out before changing types.

**Data Type:** Unsigned integer (32-bit)

**Default:** 1

**Range:** 0 to \(2^{31} - 1\)

**Units:** Asynchronous requests

**Dynamic?:** Yes, but the cluster setting for a file system is set when the file system is mounted. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation:** None. However, setting the value of this parameter to 0 causes all of the queued requests of a particular request type to be processed before moving on to the next type. This effectively disables the fairness portion of the algorithm.

**When to Change:** To increase the number of each type of asynchronous request that is generated before switching to the next type. Doing so might help with server functionality that depends upon clusters of requests coming from the client.
Commitment Level Unstable

**nfs:nfs_async_timeout**

Description Controls the duration of time that threads, which execute asynchronous I/O requests, sleep with nothing to do before exiting. When there are no more requests to execute, each thread goes to sleep. If no new requests come in before this timer expires, the thread wakes up and exits. If a request does arrive, a thread is woken up to execute requests until there are none again. Then, the thread goes back to sleep waiting for another request to arrive, or for the timer to expire.

- **Data Type**: Integer (32-bit)
- **Default**: 6000 (1 minute expressed as 60 sec * 100Hz)
- **Range**: 0 to $2^{31} - 1$
- **Units**: Hz. (Typically, the clock runs at 100Hz.)
- **Dynamic?**: Yes
- **Validation**: None. However, setting this parameter to a non positive value causes these threads exit as soon as there are no requests in the queue for them to process.

When to Change If the behavior of applications in the system is known precisely and the rate of asynchronous I/O requests can be predicted, it might be possible to tune this parameter to optimize performance slightly in either of the following ways:

- By making the threads expire more quickly, thus freeing up kernel resources more quickly
- By making the threads expire more slowly, thus avoiding thread create and destroy overhead

Commitment Level Unstable

**nfs:nacache**

Description Tunes the number of hash queues that access the file access cache on the NFS client. The file access cache stores file access rights that users have with respect to files that they are trying to access. The cache itself is dynamically allocated. However, the hash queues used to index into the cache are statically allocated. The algorithm assumes that there is one access cache entry per active file and four of these access cache
entries per hash bucket. Thus, by default, the value of this parameter is set to the value of the \texttt{nrnode} parameter.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>The default setting of this parameter is 0. This value means that the value of \texttt{nacache} should be set to the value of the \texttt{nrnode} parameter.</td>
</tr>
<tr>
<td>Range</td>
<td>1 to $2^{31} - 1$</td>
</tr>
<tr>
<td>Units</td>
<td>Access cache entries</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>No. This value can only be changed by adding or changing the parameter in the /etc/system file, and then rebooting system.</td>
</tr>
<tr>
<td>Validation</td>
<td>None. However, setting this parameter to a negative value will probably cause the system to try to allocate a very large set of hash queues. While trying to do so, the system is likely to hang.</td>
</tr>
<tr>
<td>When to Change</td>
<td>Examine the value of this parameter if the basic assumption of one access cache entry per file would be violated. This violation could occur for systems in a timesharing mode where multiple users are accessing the same file at about the same time. In this case, it might be helpful to increase the expected size of the access cache so that the hashed access to the cache stays efficient.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
</tr>
</tbody>
</table>

**\texttt{nfs:nfs3\_jukebox\_delay}**

Description: Controls the duration of time that the NFS version 3 client waits to transmit a new request after receiving the \texttt{NFS3ERR\_JUKEBOX} error from a previous request. The \texttt{NFS3ERR\_JUKEBOX} error is generally returned from the server when the file is temporarily unavailable for some reason. This error is generally associated with hierarchical storage, and CD or tape jukeboxes.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Long integer (64-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>1000 (10 seconds expressed as 10 sec * 100Hz)</td>
</tr>
<tr>
<td>Range</td>
<td>0 to $2^{63} - 1$ on 64-bit platforms</td>
</tr>
<tr>
<td>Units</td>
<td>Hz. (Typically, the clock runs at 100Hz.)</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
</tbody>
</table>
When to Change: Examine the value of this parameter and perhaps adjust it to match the behaviors exhibited by the server. Increase this value if the delays in making the file available are long in order to reduce network overhead due to repeated retransmissions. Decrease this value to reduce the delay in discovering that the file has become available.

Commitment Level: Unstable

**nfs:nfs3_max_transfer_size**

Description: Controls the maximum size of the data portion of an NFS version 3 READ, WRITE, READDIR, or READDIRPLUS request. This parameter controls both the maximum size of the request that the server returns as well as the maximum size of the request that the client generates.

Data Type: Integer (32-bit)

Default: 1,048,576 (1 Mbyte)

Range: 0 to $2^{31} - 1$

Units: Bytes

Dynamic?: Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.

Validation: None. However, setting the maximum transfer size on the server to 0 is likely to cause clients to malfunction or just decide not to attempt to talk to the server.

There is also a limit on the maximum transfer size when using NFS over the UDP transport. UDP has a hard limit of 64 KB per datagram. This 64 KB must include the RPC header as well as other NFS information, in addition to the data portion of the request. Setting the limit too high might result in errors from UDP and communication problems between the client and the server.

When to Change: To tune the size of data transmitted over the network. In general, the nfs:nfs3_bsize parameter should also be updated to reflect changes in this parameter.

For example, when you attempt to increase the transfer size beyond 32 KB, update nfs:nfs3_bsize to reflect the increased value. Otherwise, no change in the over-the-wire request size is observed. For more information, see "nfs:nfs3_bsize" on page 102.
If you want to use a smaller transfer size than the default transfer size, use the `mount` command’s `-wsize` or `-rsize` option on a per-file system basis.

Commitment Level Unstable

**nfs:nfs4_max_transfer_size**

**Description** Controls the maximum size of the data portion of an NFS version 4 READ, WRITE, READDIR, or READDIRPLUS request. This parameter controls both the maximum size of the request that the server returns as well as the maximum size of the request that the client generates.

**Data Type** Integer (32-bit)

**Default** 32,768 (32 KB)

**Range** 0 to $2^{31}$ - 1

**Units** Bytes

**Dynamic?** Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.

**Validation** None. However, setting the maximum transfer size on the server to 0 is likely to cause clients to malfunction or just decide not to attempt to talk to the server.

There is also a limit on the maximum transfer size when using NFS over the UDP transport. For more information on the maximum for UDP, see “nfs:nfs3_max_transfer_size” on page 108.

**When to Change** To tune the size of data transmitted over the network. In general, the `nfs:nfs4_bsize` parameter should also be updated to reflect changes in this parameter.

For example, when you attempt to increase the transfer size beyond 32 KB, update `nfs:nfs4_bsize` to reflect the increased value. Otherwise, no change in the over-the-wire request size is observed. For more information, see “nfs:nfs4_bsize” on page 102.

If you want to use a smaller transfer size than the default transfer size, use the `mount` command’s `-wsize` or `-rsize` option on a per-file system basis.

Commitment Level Unstable
**nfs:nfs3_max_transfer_size_clts**

Description: Controls the maximum size of the data portion of an NFS version 3 READ, WRITE, READDIR, or READDIRPLUS request over UDP. This parameter controls both the maximum size of the request that the server returns as well as the maximum size of the request that the client generates.

Data Type: Integer (32-bit)

Default: 32,768 (32 KB)

Range: 0 to $2^{31}$ - 1

Units: Bytes

Dynamic?: Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.

Validation: None. However, setting the maximum transfer size on the server to 0 is likely to cause clients to malfunction or just decide not to attempt to talk to the server.

When to Change: Do not change this parameter.

Commitment Level: Unstable

**nfs:nfs3_max_transfer_size_cots**

Description: Controls the maximum size of the data portion of an NFS version 3 READ, WRITE, READDIR, or READDIRPLUS request over TCP. This parameter controls both the maximum size of the request that the server returns as well as the maximum size of the request that the client generates.

Data Type: Integer (32-bit)

Default: 1,048,576 bytes

Range: 0 to $2^{31}$ - 1

Units: Bytes

Dynamic?: Yes, but this parameter is set per file system at mount time. To affect a particular file system, unmount and mount the file system after changing this parameter.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>rpcmod:clnt_max_conns</code></td>
<td>Controls the number of TCP connections that the NFS client uses when communicating with each NFS server. The kernel RPC is constructed so that it can multiplex RPCs over a single connection. However, multiple connections can be used, if preferred.</td>
</tr>
</tbody>
</table>

### Validation
None. However, setting the maximum transfer size on the server to 0 is likely to cause clients to malfunction or just decide not to attempt to talk to the server.

### When to Change
Do not change this parameter unless transfer sizes larger than 1 Mbyte are preferred.

### Commitment Level
Unstable
### rpcmod:clnt_idle_timeout

- **Description**: Controls the duration of time on the client that a connection between the client and server is allowed to remain idle before being closed.
- **Data Type**: Long integer (64-bit)
- **Default**: 300,000 milliseconds (5 minutes)
- **Range**: 0 to $2^{63} - 1$
- **Units**: Milliseconds
- **Dynamic?**: Yes
- **Validation**: None
- **When to Change**: Use this parameter to change the time that idle connections are allowed to exist on the client before being closed. You might want to close connections at a faster rate to avoid consuming system resources.
- **Commitment Level**: Unstable

### rpcmod:svc_idle_timeout

- **Description**: Controls the duration of time on the server that a connection between the client and server is allowed to remain idle before being closed.
- **Data Type**: Long integer (64-bit)
- **Default**: 360,000 milliseconds (6 minutes)
- **Range**: 0 to $2^{63} - 1$
- **Units**: Milliseconds
- **Dynamic?**: Yes
- **Validation**: None
- **When to Change**: Use this parameter to change the time that idle connections are allowed to exist on the server before being closed. You might want to close connections at a faster rate to avoid consuming system resources.
- **Commitment Level**: Unstable

### rpcmod:svc_default_stksize

- **Description**: Sets the size of the kernel stack for kernel RPC service threads.
### rpcmod:stack

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>The default value is 0. This value means that the stack size is set to the system default.</td>
</tr>
<tr>
<td>Range</td>
<td>0 to (2^{31} - 1)</td>
</tr>
<tr>
<td>Units</td>
<td>Bytes</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>Yes, for all new threads that are allocated. The stack size is set when the thread is created. Therefore, changes to this parameter do not affect existing threads but are applied to all new threads that are allocated.</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
<tr>
<td>When to Change</td>
<td>Very deep call depths can cause the stack to overflow and cause red zone faults. The combination of a fairly deep call depth for the transport, coupled with a deep call depth for the local file system, can cause NFS service threads to overflow their stacks. Set this parameter to a multiple of the hardware page size on the platform.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
</tr>
</tbody>
</table>

### rpcmod:maxdupreqs

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Integer (32-bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>8192</td>
</tr>
<tr>
<td>Range</td>
<td>1 to (2^{31} - 1)</td>
</tr>
<tr>
<td>Units</td>
<td>Requests</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>The cache is dynamically sized, but the hash queues that provide fast access to the cache are statically sized. Making the cache very large might result in long search times to find entries in the cache. Do not set the value of this parameter to 0. This value prevents the NFS server from handling non idempotent requests.</td>
</tr>
<tr>
<td>Validation</td>
<td>None</td>
</tr>
</tbody>
</table>
When to Change  Examine the value of this parameter if false failures are encountered by NFS clients. For example, if an attempt to create a directory fails, but the directory is actually created, perhaps that retransmitted `MKDIR` request was not detected by the server.

The size of the cache should match the load on the server. The cache records non-idempotent requests and so only needs to track a portion of the total requests. The cache does need to hold the information long enough to be able to detect a retransmission by the client. Typically, the client timeout for connectionless transports is relatively short, starting around 1 second and increasing to about 20 seconds.

Commitment Level  Unstable

**rpcmod:cotsmaxdupreqs**

**Description**  Controls the size of the duplicate request cache that detects RPC-level retransmissions on connection-oriented transports. This cache is indexed by the client network address and the RPC procedure number, program number, version number, and transaction ID. This cache avoids processing retransmitted requests that might not be idempotent.

**Data Type**  Integer (32–bit)

**Default**  8192

**Range**  1 to $2^{31} - 1$

**Units**  Requests

**Dynamic?**  Yes

**Validation**  The cache is dynamically sized, but the hash queues that provide fast access to the cache are statically sized. Making the cache very large might result in long search times to find entries in the cache.

Do not set the value of this parameter to 0. It prevents the NFS server from handling non-idempotent requests.

**When to Change**  Examine the value of this parameter if false failures are encountered by NFS clients. For example, if an attempt to create a directory fails, but the directory is actually created, it is possible that a retransmitted `MKDIR` request was not detected by the server.

The size of the cache should match the load on the server. The cache records non-idempotent requests and so only needs to track a portion
of the total requests. It does need to hold the information long enough to be able to detect a retransmission on the part of the client. Typically, the client timeout for connection oriented transports is very long, about 1 minute. Thus, entries need to stay in the cache for fairly long times.

Commitment Level    Unstable
This chapter describes various Internet Protocol suite properties.

- “IP Tunable Parameters” on page 118
- “TCP Tunable Parameters” on page 124
- “UDP Tunable Parameters” on page 141
- “IPQoS Tunable Parameter” on page 144
- “SCTP Tunable Parameters” on page 145
- “Per-Route Metrics” on page 156

**Where to Find Tunable Parameter Information**

<table>
<thead>
<tr>
<th>Tunable Parameter</th>
<th>For Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solaris kernel tunables</td>
<td>Chapter 2, “Oracle Solaris Kernel Tunable Parameters”</td>
</tr>
<tr>
<td>NFS tunable parameters</td>
<td>Chapter 3, “NFS Tunable Parameters”</td>
</tr>
<tr>
<td>Network Cache and Accelerator (NCA) tunable parameters</td>
<td>Chapter 5, “Network Cache and Accelerator Tunable Parameters”</td>
</tr>
</tbody>
</table>

**Overview of Tuning IP Suite Parameters**

You can set all of the tuning parameters described in this chapter by using the `ipadm` command except for the following parameters:

- “ipcl_conn_hash_size” on page 136
- "ip_squeue_worker_wait" on page 137
- “ip_squeue_fanout” on page 122

These parameters can only be set in the `/etc/system` file.
Use the following syntax to set TCP/IP parameters by using the `ipadm` command:

```
# ipadm set-prop -p parameter ip|ipv4|ipv6|tcp|udp|sctp
```

For example:

```
# ipadm set-prop -p extra_priv_ports=1047 tcp
# ipadm show-prop -p extra_priv_ports tcp
```

<table>
<thead>
<tr>
<th>PROTO PROPERTY</th>
<th>PERM CURRENT</th>
<th>PERSISTENT</th>
<th>DEFAULT</th>
<th>POSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcp</td>
<td>extra_priv_ports</td>
<td>rw</td>
<td>1047</td>
<td>1047</td>
</tr>
</tbody>
</table>

For more information, see `ipadm(1M)`.

Use the following syntax to set TCP/IP parameters by using the `ndd` command:

```
# ndd -set driver parameter value
```

For more information, see `ndd(1M)`.

### IP Suite Parameter Validation

All parameters described in this section are checked to verify that they fall in the parameter range. The parameter's range is provided with the description for each parameter.

### Internet Request for Comments (RFCs)

Internet protocol and standard specifications are described in RFC documents. You can get copies of RFCs from `ftp://ftp.rfc-editor.org/in-notes`. Browse RFC topics by viewing the `rfc-index.txt` file at this site.

### IP Tunable Parameters

#### `_icmp_err_interval` and `_icmp_err_burst`

**Description**

Controls the rate of IP in generating ICMP error messages. IP generates only up to `_icmp_err_burst` IP error messages in any `_icmp_err_interval`.

The `_icmp_err_interval` parameter protects IP from denial of service attacks. Setting this parameter to 0 disables rate limiting. It does not disable the generation of error messages.

**Default**

100 milliseconds for `_icmp_err_interval`

10 error messages for `_icmp_err_burst`
**icmp_err_interval**

Range: 0 – 99,999 milliseconds for `icmp_err_interval`

Dynamic?: Yes

When to Change: If you need a higher error message generation rate for diagnostic purposes.

Commitment Level: Unstable

Change History: For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

---

**respond_to_echo_broadcast**

Description: Controls whether IP responds to a broadcast ICMPv4 echo request or a ICMPv6 echo request.

Default: 1 (enabled)

Range: 0 (disabled) or 1 (enabled)

Dynamic?: Yes

When to Change: If you do not want this behavior for security reasons, disable it.

Commitment Level: Unstable

Change History: For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

---

**send_redirects (ipv4 or ipv6)**

Description: Controls whether IPv4 or IPv6 sends out ICMPv4 or ICMPv6 redirect messages.

Default: 1 (enabled)

Range: 0 (disabled) or 1 (enabled)

Dynamic?: Yes

When to Change: If you do not want this behavior for security reasons, disable it.

Commitment Level: Unstable
Change History  For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

**forwarding (ipv4 or ipv6)**

Description  Controls whether IPv4 or IPv6 forwards packets with source IPv4 routing options or IPv6 routing headers.

Default  0 (disabled)

Range  0 (disabled) or 1 (enabled)

Dynamic?  Yes

When to Change  Keep this parameter disabled to prevent denial of service attacks.

Commitment Level  Unstable

Change History  For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

**ttl**

Description  Controls the time to live (TTL) value in the IPv4 header for the outbound IPv4 packets on an IP association.

Default  255

Range  1 to 255

Dynamic?  Yes

When to Change  Generally, you do not need to change this value.

Commitment Level  Unstable

Change History  For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

**hoplimit (ipv6)**

Description  Sets the value of the hop limit in the IPv6 header for the outbound IPv6 packets on an IP association.

Default  255

Range  0 to 255
Dynamic? Yes
When to Change Generally, you do not need to change this value.
Commitment Level Unstable
Change History For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

_addrs_per_if

Description Defines the maximum number of logical IP interfaces associated with a real interface.

Default 256
Range 1 to 8192
Dynamic? Yes
When to Change Do not change the value. If more logical interfaces are required, you might consider increasing the value. However, recognize that this change might have a negative impact on IP’s performance.
Commitment Level Unstable
Change History For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

hostmodel (ipv4 or ipv6)

Description Controls send and receive behavior for IPv4 or IPv6 packets on a multi-homed system. This property can have the following values: weak, strong, and src-priority. The default value is weak.

Default weak
Range weak, strong, or src-priority

- weak
  - Outgoing packets - The source address of the packet going out need not match the address configured on the outgoing interface.
  - Incoming packets - The destination address of the incoming packet need not match the address configured on the incoming interface.
**strong**
- Outgoing packets - The source address of the packet going out must match the address configured on the outgoing interface.
- Incoming packets - The destination address of the incoming packet must match the address configured on the incoming interface.

**src-priority**
- Outgoing packets - If multiple routes for the IP destination in the packet are available, the system prefers routes where the IP source address in the packet is configured on the outgoing interface.
  
  If no such route is available, the system falls back to selecting the best route, as with the weak ES case.
- Incoming packets - The destination address of the incoming packet must be configured on any one of the host’s interface.

**Dynamic?** Yes
**When to Change** If a machine has interfaces that cross strict networking domains (for example, a firewall or a VPN node), set this parameter to strong.
**Commitment Level** Unstable
**Change History** For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

### ip_queue_fanout

**Description** Determines the mode of associating TCP/IP connections with squeues.

A value of 0 associates a new TCP/IP connection with the CPU that creates the connection. A value of 1 associates the connection with multiple squeues that belong to different CPUs.

**Default** 0
**Range** 0 or 1
**Dynamic?** Yes
**When to Change** Consider setting this parameter to 1 to spread the load across all CPUs in certain situations. For example, when the number of CPUs exceed...
the number of NICs, and one CPU is not capable of handling the network load of a single NIC, change this parameter to 1.

This property can only be set in the /etc/system file.

Zone Configuration
This parameter can only be set in the global zone.

Commitment Level
Unstable

**IP Tunable Parameters With Additional Cautions**

Changing the following parameters is not recommended.

### `_pathmtu_interval`

**Description**
Specifies the interval in milliseconds when IP flushes the path maximum transfer unit (PMTU) discovery information, and tries to rediscover PMTU.

Refer to RFC 1191 on PMTU discovery.

**Default**
10 minutes

**Range**
5 seconds to 277 hours

**Dynamic?**
Yes

**When to Change**
Do not change this value.

**Commitment Level**
Unstable

**Change History**
For information, see “IP Parameter Name Changes (Oracle Solaris 11)” on page 171.

### `_icmp_return_data_bytes (ipv4 or ipv6)`

**Description**
When IPv4 or IPv6 sends an ICMPv4 or ICMPv6 error message, it includes the IP header of the packet that caused the error message. This parameter controls how many extra bytes of the packet beyond the IPv4 or IPv6 header are included in the ICMPv4 or ICMPv6 error message.

**Default**
64 for IPv4

1280 for IPv6

**Range**
8–6636 for IPv4

8–1280 for IPv6
TCP Tunable Parameters

**_deferred_ack_interval_**

| Description | Specifies the time-out value for the TCP-delayed acknowledgment (ACK) timer for hosts that are not directly connected. |
| Default | 100 milliseconds |
| Range | 1 millisecond to 1 minute |
| Dynamic? | Yes |
| When to Change | Do not increase this value to more than 500 milliseconds. Increase the value under the following circumstances: |
| | - Slow network links (less than 57.6 Kbps) with greater than 512 bytes maximum segment size (MSS) |
| | - The interval for receiving more than one TCP segment is short |
| Commitment Level | Unstable |
| Change History | For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172 |

**_local_dack_interval_**

<p>| Description | Specifies the time-out value for TCP-delayed acknowledgment (ACK) timer for hosts that are directly connected. |
| Default | |
| Range | |
| Dynamic? | |
| When to Change | |
| Commitment Level | |
| Change History | |</p>
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default</th>
<th>Range</th>
<th>Dynamic?</th>
<th>When to Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>_deferred_acks_max</td>
<td>2</td>
<td>0 to 16</td>
<td>Yes</td>
<td>Do not change the value. In some circumstances, when the network traffic becomes very bursty because of the delayed ACK effect, decrease the value. Do not decrease this value below 2.</td>
</tr>
</tbody>
</table>

**Description**

Specifies the maximum number of TCP segments received from remote destinations (not directly connected) before an acknowledgment (ACK) is generated. TCP segments are measured in units of maximum segment size (MSS) for individual connections. If set to 0 or 1, no ACKs are delayed, assuming all segments are 1 MSS long. The actual number is dynamically calculated for each connection. The value is the default maximum.
_local_dacks_max

Description: Specifies the maximum number of TCP segments received from directly connected destinations before an acknowledgment (ACK) is generated. TCP segments are measured in units of maximum segment size (MSS) for individual connections. If set to 0 or 1, it means no ACKs are delayed, assuming all segments are 1 MSS long. The actual number is dynamically calculated for each connection. The value is the default maximum.

Default: 8
Range: 0 to 16
Dynamic?: Yes
When to Change: Do not change the value. In some circumstances, when the network traffic becomes very bursty because of the delayed ACK effect, decrease the value. Do not decrease this value below 2.
Commitment Level: Unstable
Change History: For information, see "TCP Parameter Name Changes (Oracle Solaris 11)" on page 172.

_wscale_always

Description: When this parameter is enabled, which is the default setting, TCP always sends a SYN segment with the window scale option, even if the window scale option value is 0. Note that if TCP receives a SYN segment with the window scale option, even if the parameter is disabled, TCP responds with a SYN segment with the window scale option. In addition, the option value is set according to the receive window size.

Refer to RFC 1323 for the window scale option.

Default: 1 (enabled)
Range: 0 (disabled) or 1 (enabled)
Dynamic?: Yes
When to Change: If there is an interoperability problem with an old TCP stack that does not support the window scale option, disable this parameter.
Commitment Level: Unstable
_tstamp_always

Description: If set to 1, TCP always sends a SYN segment with the timestamp option. Note that if TCP receives a SYN segment with the timestamp option, TCP responds with a SYN segment with the timestamp option even if the parameter is set to 0.

Default: 0 (disabled)
Range: 0 (disabled) or 1 (enabled)
Dynamic?: Yes
When to Change: If getting an accurate measurement of round-trip time (RTT) and TCP sequence number wraparound is a problem, enable this parameter.

Refer to RFC 1323 for more reasons to enable this option.

Commitment Level: Unstable
Change History: For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

send_buf

Description: Defines the default send window size in bytes. Refer to “Per-Route Metrics” on page 156 for a discussion of setting a different value on a per-route basis. See also “max_buf” on page 128.

Default: 49,152
Range: 4096 to the current value of “max_buf” on page 128
Dynamic?: Yes
When to Change: An application can use setsockopt(3XNET) SO_SNDBUF to change the individual connection’s send buffer.

Commitment Level: Unstable
Change History: For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.
**recv_buf**

Description: Defines the default receive window size in bytes. Refer to "Per-Route Metrics" on page 156 for a discussion of setting a different value on a per-route basis. See also “max_buf” on page 128 and "_recv_hiwat_minmss" on page 141.

Default: 128,000

Range: 2048 to the current value of "max_buf" on page 128

Dynamic?: Yes

When to Change: An application can use `setsockopt(3XNET) SO_RCVBUF` to change the individual connection's receive buffer.

Commitment Level: Unstable

Change History: For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

**max_buf**

Description: Defines the maximum send and receive buffer size in bytes. This parameter controls how large the send and receive buffers are set to by an application that uses `setsockopt(3XNET)`.

Default: 1,048,576

Range: 128,000 to 1,073,741,824

Dynamic?: Yes

When to Change: If TCP connections are being made in a high-speed network environment, increase the value to match the network link speed.

Commitment Level: Unstable

Change History: For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

**_cwnd_max**

Description: Defines the maximum value of the TCP congestion window (cwnd) in bytes.
For more information on the TCP congestion window, refer to RFC 1122 and RFC 2581.

**Default** 1,048,576  
**Range** 128 to 1,073,741,824  
**Dynamic?** Yes  
**When to Change** Even if an application uses `setsockopt(3XNET)` to change the window size to a value higher than `_cwnd_max`, the actual window used can never grow beyond `_cwnd_max`. Thus, `_max_buf` should be greater than `_cwnd_max`.

**Commitment Level** Unstable  
**Change History** For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

### `_slow_start_initial`

**Description** Defines the maximum initial congestion window (cwnd) size in the maximum segment size (MSS) of a TCP connection.

Refer to RFC 2414 on how the initial congestion window size is calculated.

**Default** 4  
**Range** 1 to 4  
**Dynamic?** Yes  
**When to Change** Do not change the value.

If the initial cwnd size causes network congestion under special circumstances, decrease the value.

**Commitment Level** Unstable  
**Change History** For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

### `_slow_start_after_idle`

**Description** The congestion window size in the maximum segment size (MSS) of a TCP connection after it has been idled (no segment received) for a period of one retransmission timeout (RTO).
Refer to RFC 2414 on how the initial congestion window size is calculated.

- **Default:** 4
- **Range:** 1 to 16,384
- **Dynamic?:** Yes
- **When to Change:** For more information, see "slow_start_initial" on page 129.
- **Commitment Level:** Unstable
- **Change History:** For information, see "TCP Parameter Name Changes (Oracle Solaris 11)" on page 172.

### sack

- **Description:** If set to 2, TCP always sends a SYN segment with the selective acknowledgment (SACK) permitted option. If TCP receives a SYN segment with a SACK-permitted option and this parameter is set to 1, TCP responds with a SACK-permitted option. If the parameter is set to 0, TCP does not send a SACK-permitted option, regardless of whether the incoming segment contains the SACK permitted option. Refer to RFC 2018 for information on the SACK option.
- **Default:** 2 (active enabled)
- **Range:** 0 (disabled), 1 (passive enabled), or 2 (active enabled)
- **Dynamic?:** Yes
- **When to Change:** SACK processing can improve TCP retransmission performance so it should be actively enabled. Sometimes, the other side can be confused with the SACK option actively enabled. If this confusion occurs, set the value to 1 so that SACK processing is enabled only when incoming connections allow SACK processing.
- **Commitment Level:** Unstable
- **Change History:** For information, see "TCP Parameter Name Changes (Oracle Solaris 11)" on page 172.
**_rev_src_routes_**

**Description**  
If set to 0, TCP does not reverse the IP source routing option for incoming connections for security reasons. If set to 1, TCP does the normal reverse source routing.

**Default**  
0 (disabled)

**Range**  
0 (disabled) or 1 (enabled)

**Dynamic?**  
Yes

**When to Change**  
If IP source routing is needed for diagnostic purposes, enable it.

**Commitment Level**  
Unstable

**Change History**  
For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

**_time_wait_interval_**

**Description**  
Specifies the time in milliseconds that a TCP connection stays in TIME-WAIT state.

For more information, refer to RFC 1122, 4.2.2.13.

**Default**  
60,000 (60 seconds)

**Range**  
1 second to 10 minutes

**Dynamic?**  
Yes

**When to Change**  
Do not set the value lower than 60 seconds.

For information on changing this parameter, refer to RFC 1122, 4.2.2.13.

**Commitment Level**  
Unstable

**Change History**  
For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

**ecn**

**Description**  
Controls Explicit Congestion Notification (ECN) support.

If this parameter is set to 0, TCP does not negotiate with a peer that supports the ECN mechanism.
If this parameter is set to 1 when initiating a connection, TCP does not tell a peer that it supports ECN mechanism.

However, TCP tells a peer that it supports ECN mechanism when accepting a new incoming connection request if the peer indicates that it supports ECN mechanism in the SYN segment.

If this parameter is set to 2, in addition to negotiating with a peer on the ECN mechanism when accepting connections, TCP indicates in the outgoing SYN segment that it supports the ECN mechanism when TCP makes active outgoing connections.

Refer to RFC 3168 for information on ECN.

**Default**
1 (passive enabled)

**Range**
0 (disabled), 1 (passive enabled), or 2 (active enabled)

**Dynamic?** Yes

**When to Change**
ECN can help TCP better handle congestion control. However, there are existing TCP implementations, firewalls, NATs, and other network devices that are confused by this mechanism. These devices do not comply to the IETF standard.

Because of these devices, the default value of this parameter is set to 1. In rare cases, passive enabling can still cause problems. Set the parameter to 0 only if absolutely necessary.

**Commitment Level** Unstable

**Change History** For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

---

**_conn_req_max_q**

**Description** Specifies the default maximum number of pending TCP connections for a TCP listener waiting to be accepted by accept(3SOCKET). See also “_conn_req_max_q0” on page 133.

**Default** 128

**Range** 1 to 4,294,967,295

**Dynamic?** Yes
When to Change | For applications such as web servers that might receive several connection requests, the default value might be increased to match the incoming rate.

Do not increase the parameter to a very large value. The pending TCP connections can consume excessive memory. Also, if an application cannot handle that many connection requests fast enough because the number of pending TCP connections is too large, new incoming requests might be denied.

Note that increasing _conn_req_max_q does not mean that applications can have that many pending TCP connections. Applications can use listen(3SOCKET) to change the maximum number of pending TCP connections for each socket. This parameter is the maximum an application can use listen() to set the number to. Thus, even if this parameter is set to a very large value, the actual maximum number for a socket might be much less than _conn_req_max_q, depending on the value used in listen().

Commitment Level | Unstable
Change History | For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

<table>
<thead>
<tr>
<th>_conn_req_max_q0</th>
</tr>
</thead>
</table>
| **Description** | Specifies the default maximum number of incomplete (three-way handshake not yet finished) pending TCP connections for a TCP listener. For more information on TCP three-way handshake, refer to RFC 793. See also _"conn_req_max_q" on page 132.
| **Default** | 1024
| **Range** | 0 to 4,294,967,295
| **Dynamic?** | Yes
| **When to Change** | For applications such as web servers that might receive excessive connection requests, you can increase the default value to match the incoming rate.

The following explains the relationship between _conn_req_max_q0 and the maximum number of pending connections for each socket.
When a connection request is received, TCP first checks if the number of pending TCP connections (three-way handshake is done) waiting to be accepted exceeds the maximum \( N \) for the listener. If the connections are excessive, the request is denied. If the number of connections is allowable, then TCP checks if the number of incomplete pending TCP connections exceeds the sum of \( N \) and \_conn\_req\_max\_q0. If it does not, the request is accepted. Otherwise, the oldest incomplete pending TCP request is dropped.

Commitment Level  Unstable
Change History  For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

\_conn\_req\_min

Description  Specifies the default minimum value for the maximum number of pending TCP connection requests for a listener waiting to be accepted. This is the lowest maximum value of \texttt{listen(3SOCKET)} that an application can use.

Default  1
Range  1 to 1024
Dynamic?  Yes
When to Change  This parameter can be a solution for applications that use \texttt{listen(3SOCKET)} to set the maximum number of pending TCP connections to a value too low. Increase the value to match the incoming connection request rate.

Commitment Level  Unstable
Change History  For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

\_rst\_sent\_rate\_enabled

Description  If this parameter is set to 1, the maximum rate of sending a RST segment is controlled by the \texttt{ipmadm} parameter, \_rst\_sent\_rate. If this parameter is set to 0, no rate control when sending a RST segment is available.

Default  1 (enabled)
TCP Tunable Parameters

RST Sent Rate

Description
Sets the maximum number of RST segments that TCP can send out per second.

Default
40

Range
0 to 4,294,967,295

Dynamic?
Yes

When to Change
In a TCP environment, there might be a legitimate reason to generate more RSTs than the default value allows. In this case, increase the default value of this parameter.

Commitment Level
Unstable

Change History
For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

Smallest Anon Port

Description
This parameter controls the smallest port number TCP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.

Unit
Port number

Default
32,768

Range
1,024 to 65,535

Dynamic?
Yes
TCP Tunable Parameters

When to Change When a larger ephemeral port range is required.
Commitment Level Unstable
Change History For information, see “[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)” on page 171.

largest_anon_port
Description This parameter controls the largest port number TCP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.
Unit Port number
Default 65,535
Range 32,768 to 65,535
Dynamic? Yes
When to Change When a larger ephemeral port range is required.
Commitment Level Unstable
Change History For information, see “[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)” on page 171.

TCP/IP Parameters Set in the /etc/system File
The following parameters can be set only in the /etc/system file. After the file is modified, reboot the system.

For example, the following entry sets the ipcl_conn_hash_size parameter:

```
set ip:ipcl_conn_hash_size=value
```

ipcl_conn_hash_size
Description Controls the size of the connection hash table used by IP. The default value of 0 means that the system automatically sizes an appropriate value for this parameter at boot time, depending on the available memory.
**ip_squeue_worker_wait**

Data Type: Unsigned integer  
Default: 0  
Range: 0 to 82,500  
Dynamic?: No. The parameter can only be changed at boot time.  
When to Change: If the system consistently has tens of thousands of TCP connections, the value can be increased accordingly. Increasing the hash table size means that more memory is wired down, thereby reducing available memory to user applications.  
Commitment Level: Unstable

**Description**
Governs the maximum delay in waking up a worker thread to process TCP/IP packets that are enqueued on an squeue. An squeue is a serialization queue that is used by the TCP/IP kernel code to process TCP/IP packets.

```
Default: 10 milliseconds
Range: 0 – 50 milliseconds
Dynamic?: Yes
When to Change: Consider tuning this parameter if latency is an issue, and network traffic is light. For example, if the machine serves mostly interactive network traffic.

The default value usually works best on a network file server, a web server, or any server that has substantial network traffic.
```

Zone Configuration: This parameter can only be set in the global zone.
Commitment Level: Unstable

---

**TCP Parameters With Additional Cautions**

Changing the following parameters is not recommended.

**_keepalive_interval**

Description: This ipadm parameter sets a probe interval that is first sent out after a TCP connection is idle on a system-wide basis.
Solaris supports the TCP keep-alive mechanism as described in RFC 1122. This mechanism is enabled by setting the SO_KEEPALIVE socket option on a TCP socket.

If SO_KEEPALIVE is enabled for a socket, the first keep-alive probe is sent out after a TCP connection is idle for two hours, the default value of the tcp_keepalive_interval parameter. If the peer does not respond to the probe after eight minutes, the TCP connection is aborted. For more information, refer to "_rexmit_interval_initial" on page 139.

You can also use the TCP_KEEPALIVE_THRESHOLD socket option on individual applications to override the default interval so that each application can have its own interval on each socket. The option value is an unsigned integer in milliseconds. See also tcp(7P).

| Default | 2 hours |
| Range   | 10 seconds to 10 days |
| Units   | Unsigned integer (milliseconds) |
| Dynamic? | Yes |
| When to Change | Do not change the value. Lowering it may cause unnecessary network traffic and might also increase the chance of premature termination of the connection because of a transient network problem. |
| Commitment Level | Unstable |
| Change History | For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172 |

_ip_abort_interval

Description Specifies the default total retransmission timeout value for a TCP connection. For a given TCP connection, if TCP has been retransmitting for _ip_abort_interval period of time and it has not received any acknowledgment from the other endpoint during this period, TCP closes this connection.

For TCP retransmission timeout (RTO) calculation, refer to RFC 1122, 4.2.3. See also "_rexmit_interval_max" on page 139.

| Default | 5 minutes |
| Range   | 500 milliseconds to 1193 hours |
| Dynamic? | Yes |
When to Change | Do not change this value. See "\_rexmit\_interval\_max" on page 139 for exceptions.
Commitment Level | Unstable
Change History | For information, see "TCP Parameter Name Changes (Oracle Solaris 11)" on page 172.

### \_rexmit\_interval\_initial

| Description | Specifies the default initial retransmission timeout (RTO) value for a TCP connection. Refer to "Per-Route Metrics" on page 156 for a discussion of setting a different value on a per-route basis. |
| Default | 1000 milliseconds |
| Range | 1 millisecond to 20000 milliseconds |
| Dynamic? | Yes |
| When to Change | Do not change this value. Lowering the value can result in unnecessary retransmissions. |
| Commitment Level | Unstable |
| Change History | For information, see "TCP Parameter Name Changes (Oracle Solaris 11)" on page 172. |

### \_rexmit\_interval\_max

| Description | Defines the default maximum retransmission timeout value (RTO). The calculated RTO for all TCP connections cannot exceed this value. See also "\_ip\_abort\_interval" on page 138. |
| Default | 6000 milliseconds |
| Range | 1 millisecond to 7200000 milliseconds |
| Dynamic? | Yes |
| When to Change | Do not change the value in a normal network environment. If, in some special circumstances, the round-trip time (RTT) for a connection is about 10 seconds, you can increase this value. If you change this value, you should also change the \_ip\_abort\_interval parameter. Change the value of \_ip\_abort\_interval to at least four times the value of \_rexmit\_interval\_max. |
| Commitment Level | Unstable |
### _rexmit_interval_min

| Description | Specifies the default minimum retransmission time out (RTO) value. The calculated RTO for all TCP connections cannot be lower than this value. See also “_rexmit_interval_max” on page 139. |
| Default | 200 milliseconds |
| Range | 1 millisecond to 7200000 milliseconds |
| Dynamic? | Yes |
| When to Change | Do not change the value in a normal network environment. TCP's RTO calculation should cope with most RTT fluctuations. If, in some very special circumstances, the round-trip time (RTT) for a connection is about 10 seconds, increase this value. If you change this value, you should change the _rexmit_interval_max parameter. Change the value of _rexmit_interval_max to at least eight times the value of _rexmit_interval_min. |
| Commitment Level | Unstable |
| Change History | For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172. |

### _rexmit_interval_extra

| Description | Specifies a constant added to the calculated retransmission time out value (RTO). |
| Default | 0 milliseconds |
| Range | 0 to 7200000 milliseconds |
| Dynamic? | Yes |
| When to Change | Do not change the value. When the RTO calculation fails to obtain a good value for a connection, you can change this value to avoid unnecessary retransmissions. |
| Commitment Level | Unstable |
| Change History | For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172. |
_tstamp_if_wscale
Description If this parameter is set to 1, and the window scale option is enabled for a connection, TCP also enables the timestamp option for that connection.
Default 1 (enabled)
Range 0 (disabled) or 1 (enabled)
Dynamic? Yes
When to Change Do not change this value. In general, when TCP is used in high-speed network, protection against sequence number wraparound is essential. Thus, you need the timestamp option.
Commitment Level Unstable
Change History For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

_recv_hiwat_minmss
Description Controls the default minimum receive window size. The minimum is _recv_hiwat_minmss times the size of maximum segment size (MSS) of a connection.
Default 8
Range 1 to 65,536
Dynamic? Yes
When to Change Do not change the value. If changing it is necessary, do not change the value lower than 4.
Commitment Level Unstable
Change History For information, see “TCP Parameter Name Changes (Oracle Solaris 11)” on page 172.

UDP Tunable Parameters

send_buf
Description Defines the default send buffer size for a UDP socket. For more information, see “max_buf” on page 142.
### max_buf

**Description**

Defines the maximum send and receive buffer size for a UDP socket. It controls how large the send and receive buffers are set to by an application that uses `getsockopt(3SOCKET)`.

**Default**

2,097,152

**Range**

65,536 to 1,073,741,824

**Dynamic?**

Yes

**When to Change**

Increase the value of this parameter to match the network link speed if associations are being made in a high-speed network environment.

**Commitment Level**

Unstable

**Change History**

For information, see “UDP Parameter Name Changes (Oracle Solaris 11)” on page 173.

---

### recv_buf

**Description**

Defines the default receive buffer size for a UDP socket. For more information, see “max_buf” on page 142.

**Default**

57,344 bytes

**Range**

128 to the current value of “max_buf” on page 142

**Dynamic?**

Yes

**Change History**

For information, see “UDP Parameter Name Changes (Oracle Solaris 11)” on page 173.

---

### max_buf

**Description**

Defines the maximum send and receive buffer size for a UDP socket. It controls how large the send and receive buffers are set to by an application that uses `getsockopt(3SOCKET)`.

**Default**

2,097,152

**Range**

65,536 to 1,073,741,824

**Dynamic?**

Yes

**When to Change**

Increase the value of this parameter to match the network link speed if associations are being made in a high-speed network environment.

**Commitment Level**

Unstable

**Change History**

For information, see “UDP Parameter Name Changes (Oracle Solaris 11)” on page 173.
smallest_anon_port

Description This parameter controls the smallest port number UDP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.

Unit Port number

Default 32,768

Range 1,024 to 65,535

Dynamic? Yes

When to Change When a larger ephemeral port range is required.

Commitment Level Unstable

Change History For information, see “[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)” on page 171.

largest_anon_port

Description This parameter controls the largest port number UDP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.

Unit Port number

Default 65,535

Range 32,768 to 65,535

Dynamic? Yes

When to Change When a larger ephemeral port range is required.
Commitment Level  Unstable
Change History  For information, see "[tcp, sctp, udp]_smallest_anon_port and [tcp, sctp, udp]_largest_anon_port (Oracle Solaris 11)" on page 171.

**IPQoS Tunable Parameter**

### _policy_mask_

**Description**
Enables or disables IPQoS processing in any of the following callout positions: forward outbound, forward inbound, local outbound, and local inbound. This parameter is a bitmask as follows:

<table>
<thead>
<tr>
<th>Not Used</th>
<th>Not Used</th>
<th>Not Used</th>
<th>Forward Outbound</th>
<th>Forward Inbound</th>
<th>Local Outbound</th>
<th>Local Inbound</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A 1 in any of the position masks or disables IPQoS processing in that particular callout position. For example, a value of 0x01 disables IPQoS processing for all the local inbound packets.

**Default**
The default value is 0, meaning that IPQoS processing is enabled in all the callout positions.

**Range**
0 (0x00) to 15 (0x0F). A value of 15 indicates that IPQoS processing is disabled in all the callout positions.

**Dynamic?**
Yes

**When to Change**
If you want to enable or disable IPQoS processing in any of the callout positions.

**Commitment Level**
Unstable
**SCTP Tunable Parameters**

**_max_init_retr**

Description: Controls the maximum number of attempts an SCTP endpoint should make at resending an INIT chunk. The SCTP endpoint can use the SCTP initiation structure to override this value.

Default: 8
Range: 0 to 128
Dynamic?: Yes
When to Change: The number of INIT retransmissions depend on "_pa_max_retr" on page 145. Ideally, _max_init_retr should be less than or equal to _pa_max_retr.
Commitment Level: Unstable
Change History: For information, see "SCTP Parameter Name Changes (Oracle Solaris 11)" on page 174.

**_pa_max_retr**

Description: Controls the maximum number of retransmissions (over all paths) for an SCTP association. The SCTP association is aborted when this number is exceeded.

Default: 10
Range: 1 to 128
Dynamic?: Yes
When to Change: The maximum number of retransmissions over all paths depend on the number of paths and the maximum number of retransmission over each path. Ideally, sctp_pa_max_retr should be set to the sum of "_pp_max_retr" on page 146 over all available paths. For example, if there are 3 paths to the destination and the maximum number of retransmissions over each of the 3 paths is 5, then _pa_max_retr should be set to less than or equal to 15. (See the Note in Section 8.2, RFC 2960.)
Commitment Level: Unstable
Change History: For information, see "SCTP Parameter Name Changes (Oracle Solaris 11)" on page 174.
### _pp_max_retr_

**Description**
Controls the maximum number of retransmissions over a specific path. When this number is exceeded for a path, the path (destination) is considered unreachable.

**Default**
5

**Range**
1 to 128

**Dynamic?**
Yes

**When to Change**
Do not change this value to less than 5.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### _cwnd_max_

**Description**
Controls the maximum value of the congestion window for an SCTP association.

**Default**
1,048,576

**Range**
128 to 1,073,741,824

**Dynamic?**
Yes

**When to Change**
Even if an application uses `setsockopt(3XNET)` to change the window size to a value higher than _cwnd_max_, the actual window used can never grow beyond _cwnd_max_. Thus, “max_buf” on page 151 should be greater than _cwnd_max_.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### _ipv4_ttl_

**Description**
Controls the time to live (TTL) value in the IP version 4 header for the outbound IPv4 packets on an SCTP association.

**Default**
64

**Range**
1 to 255
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default</th>
<th>Range</th>
<th>Dynamic?</th>
<th>When to Change</th>
<th>Commitment Level</th>
<th>Change History</th>
</tr>
</thead>
<tbody>
<tr>
<td>_ipv6_hoplimit</td>
<td>Sets the value of the hop limit in the IPv6 header for the outbound IPv6 packets on an SCTP association.</td>
<td>60</td>
<td>0 to 255</td>
<td>Yes</td>
<td>Generally, you do not need to change this value.</td>
<td>Unstable</td>
<td>For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.</td>
</tr>
<tr>
<td>_heartbeat_interval</td>
<td>Computes the interval between HEARTBEAT chunks to an idle destination, that is allowed to heartbeat.</td>
<td>30 seconds</td>
<td>0 to 86,400 seconds</td>
<td>Yes</td>
<td>Refer to RFC 2960, section 8.3.</td>
<td>Unstable</td>
<td>For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.</td>
</tr>
</tbody>
</table>
### _new_secret_interval_

**Description**
Determines when a new secret needs to be generated. The generated secret is used to compute the MAC for a cookie.

**Default**
2 minutes

**Range**
0 to 1,440 minutes

**Dynamic?**
Yes

**When to Change**
Refer to RFC 2960, section 5.1.3.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### _initial_mtu_

**Description**
Determines the initial maximum send size for an SCTP packet including the length of the IP header.

**Default**
1500 bytes

**Range**
68 to 65,535

**Dynamic?**
Yes

**When to Change**
Increase this parameter if the underlying link supports frame sizes that are greater than 1500 bytes.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### _deferred_ack_interval_

**Description**
Sets the time-out value for SCTP delayed acknowledgment (ACK) timer in milliseconds.

**Default**
100 milliseconds

**Range**
1 to 60,000 milliseconds

**Dynamic?**
Yes

**When to Change**
Refer to RFC 2960, section 6.2.

**Commitment Level**
Unstable
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default</th>
<th>Range</th>
<th>Dynamic?</th>
<th>When to Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>_ignore_path_mtu</strong></td>
<td>Enables or disables path MTU discovery.</td>
<td>0 (disabled)</td>
<td>0 (disabled) or 1 (enabled)</td>
<td>Yes</td>
<td>Enable this parameter if you want to ignore MTU changes along the path. However, doing so might result in IP fragmentation if the path MTU decreases.</td>
</tr>
<tr>
<td><strong>_initial_ssthresh</strong></td>
<td>Sets the initial slow start threshold for a destination address of the peer.</td>
<td>1,048,576</td>
<td>1024 to 4,294,967,295</td>
<td>Yes</td>
<td>Refer to RFC 2960, section 7.2.1.</td>
</tr>
<tr>
<td><strong>send_buf</strong></td>
<td>Defines the default send buffer size in bytes. See also “max_buf” on page 151.</td>
<td>102,400</td>
<td>8,192 to the current value of “max_buf” on page 151</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
### SCT Tunable Parameters

**When to Change**

An application can use `setsockopt(3XNET) SO_SNDBUF` to change the individual connection’s send buffer.

**Commitment Level**

Unstable

**Change History**

For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

---

#### xmit_lowat

**Description**

Controls the lower limit on the send window size.

**Default**

8,192

**Range**

8,192 to 1,073,741,824

**Dynamic?**

Yes

**When to Change**

Generally, you do not need to change this value. This parameter sets the minimum size required in the send buffer for the socket to be marked writable. If required, consider changing this parameter in accordance with “send_buf” on page 149.

**Commitment Level**

Unstable

**Change History**

For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

---

#### recv_buf

**Description**

Defines the default receive buffer size in bytes. See also “max_buf” on page 151.

**Default**

102,400

**Range**

8,192 to the current value of “max_buf” on page 151

**Dynamic?**

Yes

**When to Change**

An application can use `setsockopt(3XNET) SO_RCVBUF` to change the individual connection’s receive buffer.

**Commitment Level**

Unstable

**Change History**

For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.
**max_buf**

Description Controls the maximum send and receive buffer size in bytes. It controls how large the send and receive buffers are set to by an application that uses `getsockopt` (3SOCKET).

Default 1,048,576

Range 102,400 to 1,073,741,824

Dynamic? Yes

When to Change Increase the value of this parameter to match the network link speed if associations are being made in a high-speed network environment.

Commitment Level Unstable

Change History For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

**_rto_min**

Description Sets the lower bound for the retransmission timeout (RTO) in milliseconds for all the destination addresses of the peer.

Default 1,000

Range 500 to 60,000

Dynamic? Yes

When to Change Refer to RFC 2960, section 6.3.1.

Commitment Level Unstable

Change History For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

**_rto_max**

Description Controls the upper bound for the retransmission timeout (RTO) in milliseconds for all the destination addresses of the peer.

Default 60,000

Range 1,000 to 60,000,000

Dynamic? Yes

When to Change Refer to RFC 2960, section 6.3.1.
### Commitment Level
Unstable

### Change History
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### `_rto_initial`
**Description**
Controls the initial retransmission timeout (RTO) in milliseconds for all the destination addresses of the peer.

<table>
<thead>
<tr>
<th>Default</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,000</td>
<td>1,000 to 60,000,000</td>
</tr>
</tbody>
</table>

**Dynamic?** Yes

**When to Change**
Refer to RFC 2960, section 6.3.1.

### Commitment Level
Unstable

### Change History
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### `_cookie_life`
**Description**
Sets the lifespan of a cookie in milliseconds.

<table>
<thead>
<tr>
<th>Default</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>60,000</td>
<td>10 to 60,000,000</td>
</tr>
</tbody>
</table>

**Dynamic?** Yes

**When to Change**
Generally, you do not need to change this value. This parameter might be changed in accordance with “`_rto_max`” on page 151.

### Commitment Level
Unstable

### Change History
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

### `_max_in_streams`
**Description**
Controls the maximum number of inbound streams permitted for an SCTP association.

<table>
<thead>
<tr>
<th>Default</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1 to 65,535</td>
</tr>
</tbody>
</table>
### _initial_out_streams_

**Description**
Controls the maximum number of outbound streams permitted for an SCTP association.

**Default**
32

**Range**
1 to 65,535

**Dynamic?**
Yes

**When to Change**
Refer to RFC 2960, section 5.1.1.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

---

### _shutack_wait_bound_

**Description**
Controls the maximum time, in milliseconds, to wait for a SHUTDOWN ACK after having sent a SHUTDOWN chunk.

**Default**
60,000

**Range**
0 to 300,000

**Dynamic?**
Yes

**When to Change**
Generally, you do not need to change this value. This parameter might be changed in accordance with “_rto_max_” on page 151.

**Commitment Level**
Unstable

**Change History**
For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

---

### _maxburst_

**Description**
Sets the limit on the number of segments to be sent in a burst.
SCTP Tunable Parameters

_default_enabled

Description Enables or disables SCTP dynamic address reconfiguration.
Default 0 (disabled)
Range 0 (disabled) or 1 (enabled)
Dynamic? Yes
When to Change The parameter can be enabled if dynamic address reconfiguration is needed. Due to security implications, enable this parameter only for testing purposes.
Commitment Level Unstable
Change History For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.

_prsctp_enabled

Description Enables or disables the partial reliability extension (RFC 3758) to SCTP.
Default 1 (enabled)
Range 0 (disabled) or 1 (enabled)
Dynamic? Yes
When to Change Disable this parameter if partial reliability is not supported in your SCTP environment.
Commitment Level Unstable
Change History For information, see “SCTP Parameter Name Changes (Oracle Solaris 11)” on page 174.
**smallest_anon_port**

Description: This parameter controls the smallest port number SCTP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.

Unit: Port number

Default: 32,768

Range: 1,024 to 65,535

Dynamic?: Yes

When to Change: When a larger ephemeral port range is required.

Commitment Level: Unstable

Change History: For information, see “[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)” on page 171.

**largest_anon_port**

Description: This parameter controls the largest port number SCTP can select as an ephemeral port. An application can use an ephemeral port when it creates a connection with a specified protocol and it does not specify a port number. Ephemeral ports are not associated with a specific application. When the connection is closed, the port number can be reused by a different application.

Unit: Port number

Default: 65,535

Range: 32,768 — 65,535

Dynamic?: Yes

When to Change: When a larger ephemeral port range is required.

Commitment Level: Unstable

Change History: For information, see “[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)” on page 171.
Per-Route Metrics

You can use per-route metrics to associate some properties with IPv4 and IPv6 routing table entries.

For example, a system has two different network interfaces, a fast Ethernet interface and a gigabit Ethernet interface. The system default `recv_maxbuf` is 128,000 bytes. This default is sufficient for the fast Ethernet interface, but may not be sufficient for the gigabit Ethernet interface.

Instead of increasing the system’s default for `recv_maxbuf`, you can associate a different default TCP receive window size to the gigabit Ethernet interface routing entry. By making this association, all TCP connections going through the route will have the increased receive window size.

For example, the following is in the routing table (`netstat -rn`), assuming IPv4:

```
192.123.123.0 192.123.123.4 U 1 4 hme0
192.123.124.0 192.123.124.4 U 1 4 ge0
default 192.123.123.1 UG 1 8
```

In this example, do the following:

```
# route change -net 192.123.124.0 -recvpipe x
```

Then, all connections going to the 192.123.124.0 network, which is on the ge0 link, use the receive buffer size `x`, instead of the default 128,000 receive window size.

If the destination is in the a.b.c.d network, and no specific routing entry exists for that network, you can add a prefix route to that network and change the metric. For example:

```
# route add -net a.b.c.d 192.123.123.1 -netmask w.x.y.z
# route change -net a.b.c.d -recvpipe y
```

Note that the prefix route’s gateway is the default router. Then, all connections going to that network use the receive buffer size `y`. If you have more than one interface, use the `-ifp` argument to specify which interface to use. This way, you can control which interface to use for specific destinations. To verify the metric, use the `route(1M)` get command.
This chapter describes some of the Network Cache and Accelerator (NCA) tunable parameters.

- "nca:nca_conn_hash_size" on page 158
- "nca:nca_conn_req_max_q" on page 158
- "nca:nca_conn_req_max_q0" on page 158
- "nca:nca_pmax" on page 159
- "nca:nca_vpmax" on page 159
- "sq_max_size" on page 160
- "ge:ge_intr_mode" on page 161

Where to Find Tunable Parameters Information

<table>
<thead>
<tr>
<th>Tunable Parameter</th>
<th>For Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Solaris kernel tunables</td>
<td>Chapter 2, &quot;Oracle Solaris Kernel Tunable Parameters&quot;</td>
</tr>
<tr>
<td>NFS tunable parameters</td>
<td>Chapter 3, &quot;NFS Tunable Parameters&quot;</td>
</tr>
<tr>
<td>Internet Protocol Suite tunable_params</td>
<td>Chapter 4, &quot;Internet Protocol Suite Tunable Parameters&quot;</td>
</tr>
</tbody>
</table>

Tuning NCA Parameters

Setting these parameters is appropriate on a system that is a dedicated web server. These parameters allocate more memory for caching pages. You can set all of the tuning parameters described in this chapter in the /etc/system file.

For information on adding tunable parameters to the /etc/system file, see “Tuning the Oracle Solaris Kernel” on page 21.
### nca:nca_conn_hash_size

**Description**
Controls the hash table size in the NCA module for all TCP connections, adjusted to the nearest prime number.

**Default**
383 hash table entries

**Range**
0 to 201,326,557

**Dynamic?**
No

**When to Change**
When the NCA’s TCP hash table is too small to keep track of the incoming TCP connections. This situation causes many TCP connections to be grouped together in the same hashtable entry. This situation is indicated when NCA is receiving many TCP connections, and system performance decreases.

**Commitment Level**
Unstable

### nca:nca_conn_req_max_q

**Description**
Defines the maximum number of pending TCP connections for NCA to listen on.

**Default**
256 connections

**Range**
0 to 4,294,967,295

**Dynamic?**
No

**When to Change**
When NCA closes a connection immediately after it is established because it already has too many established TCP connections. If NCA is receiving many TCP connections and can handle a larger load, but is refusing any more connections, increase this parameter. Doing so allows NCA to handle more simultaneous TCP connections.

**Commitment Level**
Unstable

### nca:nca_conn_req_max_q0

**Description**
Defines the maximum number of incomplete (three-way handshake not yet finished) pending TCP connections for NCA to listen on.

**Default**
1024 connections

**Range**
0 to 4,294,967,295

**Dynamic?**
No
When to Change

When NCA refuses to accept any more TCP connections because it already has too many pending TCP connections. If NCA is receiving many TCP connections and can handle a larger load, but is refusing any more connections, increase this parameter. Doing so allows NCA to handle more simultaneous TCP connections.

Commitment Level

Unstable

**nca:nca_ppmax**

Description

Specifies the maximum amount of physical memory (in pages) used by NCA for caching the pages. This value should not be more than 75 percent of total memory.

Default

25 percent of physical memory

Range

1 percent to maximum amount of physical memory

Dynamic?

No

When to Change

When using NCA on a system with more than 512 MB of memory. If a system has a lot of physical memory that is not being used, increase this parameter. Then, NCA will efficiently use this memory to cache new objects. As a result, system performance will increase.

This parameter should be increased in conjunction with nca_vpmx, unless you have a system with more physical memory than virtual memory (a 32-bit kernel that has greater than 4 GB memory). Use `pagesize(1)` to determine your system’s page size.

Commitment Level

Unstable

**nca:nca_vpmx**

Description

Specifies the maximum amount of virtual memory (in pages) used by NCA for caching pages. This value should not be more than 75 percent of the total memory.

Default

25 percent of virtual memory

Range

1 percent to maximum amount of virtual memory

Dynamic?

No

When to Change

When using NCA on a system with more than 512 MB of memory. If a system has a lot of virtual memory that is not being used, increase this...
General System Tuning for the NCA

In addition to setting the NCA parameters, you can do some general system tuning to benefit NCA performance. If you are using gigabit Ethernet (ge driver), you should set the interface in interrupt mode for better results.

For example, a system with 4 GB of memory that is booted under 64-bit kernel should have the following parameters set in the /etc/system file. Use pagesize to determine your system’s page size.

```
set sq_max_size=0
set ge:ge_intr_mode=1
set nca:nca_conn_hash_size=82500
set nca:nca_conn_req_max_q=100000
set nca:nca_conn_req_max_q0=100000
set nca:nca_ppmax=393216
set nca:nca_vpmax=393216
```

**sq_max_size**

Description: Sets the depth of the syncq (number of messages) before a destination STREAMS queue generates a QFULL message.

Default: 10000 messages

Range: 0 (unlimited) to MAXINT

Dynamic?: No

When to Change: When NCA is running on a system with a lot of memory, increase this parameter to allow drivers to queue more packets of data. If a server is under heavy load, increase this parameter so that modules and drivers can process more data without dropping packets or getting backlogged.

Commitment Level: Unstable
### ge:ge_intr_mode

<table>
<thead>
<tr>
<th>Description</th>
<th>Enables the <code>ge</code> driver to send packets directly to the upper communication layers rather than queue the packets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0 (queue packets to upper layers)</td>
</tr>
<tr>
<td>Range</td>
<td>0 (enable) or 1 (disable)</td>
</tr>
<tr>
<td>Dynamic?</td>
<td>No</td>
</tr>
<tr>
<td>When to Change</td>
<td>When NCA is enabled, set this parameter to 1 so that the packet is delivered to NCA in interrupt mode for faster processing.</td>
</tr>
<tr>
<td>Commitment Level</td>
<td>Unstable</td>
</tr>
</tbody>
</table>
This chapter describes most of the parameters default values for various system facilities.

- "autofs" on page 164
- "cron" on page 164
- "devfsadm" on page 164
- "dhcpagent" on page 164
- "fs" on page 165
- "ftp" on page 165
- "inetinit" on page 165
- "init" on page 165
- "ipsec" on page 166
- "kbd" on page 166
- "keyserv" on page 166
- "login" on page 167
- "mpathd" on page 167
- "nfs" on page 167
- "nfslogd" on page 167
- "nss" on page 167
- "passwd" on page 167
- "su" on page 168
- "syslog" on page 168
- "tar" on page 168
- "telnetd" on page 168
- "utmpd" on page 168
System Default Parameters

The functioning of various system facilities is governed by a set of values that are read by each facility on startup. The values for each facility might be stored in a file for the facility located in the /etc/default directory, or in properties of a service instance in the Service Management Facility (SMF) configuration repository. For more information on SMF services and properties, see “Managing SMF Services” in Oracle Solaris Administration: Common Tasks.

For information about setting power management properties, see Chapter 16, “Managing the System Console, Terminal Devices, and Power Services (Tasks),” in Oracle Solaris Administration: Common Tasks.

autofs

You can display or configure SMF autofs properties by using the sharectl command. For example:

```
# sharectl get autofs
timeout=600
automount verbose=false
automountd verbose=false
nobrowse=false
trace=0
environment=
# sharectl set -p timeout=200 autofs
```

For details, see sharectl(1M).

cron

This facility enables you to disable or enable cron logging.

devfsadm

This file is not currently used.

dhcpagent

Client usage of DHCP is provided by the dhcpagent daemon. When ipadm is used to create a DHCP address object, or when ipadm identifies an interface that has been configured to receive its network configuration from DHCP, dhcpagent is started to manage an address on that interface.
For more information, see the /etc/default/dhcpagent information in the FILES section of dhcpagent(1M).

**fs**

File system administrative commands have a generic and file system-specific portion. If the file system type is not explicitly specified with the -F option, a default is applied. The value is specified in this file. For more information, see the Description section of default_fs(4).

**ftp**

This facility enables you to set the ls command behavior to the RFC 959 NLST command. The default ls behavior is the same as in the previous Solaris release.

For details, see ftp(4).

**inetinit**

This facility enables you to configure TCP sequence numbers and to enable or disable support for 6to4 relay routers.

**init**

System initialization properties are now part of the following SMF service:

```
svc:/system/environment:init
```

You can display and configure system initialization properties, such as TZ and LANG, by using similar syntax:

```
# svcCfg -s svc:/system/environment:init
```

```
svc:/system/environment:init> setprop
Usage: setprop pg/name = [type:] value
       setprop pg/name = [type:] ([value...])
```

Set the pg/name property of the currently selected entity. Values may be enclosed in double-quotes. Value lists may span multiple lines.

```
svc:/system/environment:init> listprop
umask application
umask/value authorization astring solaris.smf.value.environment
umask/umask astring 022
upgrade astring
upgrade/skip_init Upgrade boolean 0e2
```
For more information, see the FILES section of init(1M).

**ipsec**

This facility enables you to configure parameters, such as IKE daemon debugging information and the ikeadm privilege level.

**kbd**

Keyboard configuration properties are now part of the following SMF service:

```
svc:/system/keymap:default
```

You display and configure keyboard properties by using similar syntax:

```
# svcCfg -s svc:/system/keymap:default
svc:/system/keymap:default> setprop
Usage: setprop pg/name = [type:] value
      setprop pg/name = [type:] ([value...])
```

Set the pg/name property of the currently selected entity. Values may be enclosed in double-quotes. Value lists may span multiple lines.

```
svc:/system/keymap:default> listprop
general framework
  general/complete astring
  general/enabled boolean false
keymap system
  keymap/console_beeper_freq integer 900
  keymap/kbd_beeper_freq integer 2000
  keymap/keyboard_abort astring enable
  keymap/keyclick boolean false
```

For more information, see kbd(1).

**keyserv**

For details, see the /etc/default/keyserv information in the FILES section of keyserv(1M).
login
For details, see the /etc/default/login information in the FILES section of login(1).

mpathd
This facility enables you to set in.mpathd configuration parameters.
For details, see in.mpathd(1M).

nfs
You can display or configure SMF NFS properties by using the sharectl command. For example:

```
# sharectl get nfs
servers=1024
lockd_listen_backlog=32
lockd_servers=1024
lockd_retransmit_timeout=5
grace_period=90
serverVersMin=2
serverVersMax=4
clientVersMin=2
clientVersMax=4
server_delegation=on
nfsmapid_domain=
```

```
# sharectl set -p grace_period=60 nfs
```
For details, see nfs(4).

nfslogd
For details, see the Description section of nfslogd(1M).

nss
This facility enables you to configure initgroups (3C) lookup parameters.
For details, see nss(4).

passwd
For details, see the /etc/default/passwd information in the FILES section of passwd(1).
su
For details, see the `/etc/default/su` information in the FILES section of `su(1M)`.

details, see the `/etc/default/su` information in the FILES section of `su(1M)`.

syslog
For details, see the `/etc/default/syslogd` information in the FILES section of `syslogd(1M)`.

details, see the `/etc/default/syslogd` information in the FILES section of `syslogd(1M)`.

tar
For a description of the `-f` function modifier, see `tar(1)`.

If the `TAPE` environment variable is not present and the value of one of the arguments is a number and `-f` is not specified, the number matching the archiveN string is looked up in the `/etc/default/tar` file. The value of the archiveN string is used as the output device with the blocking and size specifications from the file.

For example:

```
% tar -c 2 /tmp/
```

This command writes the output to the device specified as `archive2` in the `/etc/default/tar` file.

details, see the `/etc/default/tar` file.

telnetd
This file identifies the default `BANNER` that is displayed upon a telnet connection.

telnetd
This file identifies the default `BANNER` that is displayed upon a telnet connection.

utmpd
The `utmpd` daemon monitors `/var/adm/utmp` (and `/var/adm/utmp` in earlier Solaris versions) to ensure that `utmp` entries inserted by non-root processes by `pututxline(3C)` are cleaned up on process termination.

Two entries in `/etc/default/utmpd` are supported:

- `SCAN_PERIOD` – The number of seconds that `utmpd` sleeps between checks of `/proc` to see if monitored processes are still alive. The default is 300.
- `MAX_FDS` – The maximum number of processes that `utmpd` attempts to monitor. The default value is 4096 and should never need to be changed.
This chapter describes the change history of specific tunable parameters. If a parameter is in this section, it has changed from a previous release. Parameters whose functionality has been removed are listed also.

- “Kernel Parameters” on page 169
- “TCP/IP Tunable Parameters” on page 171
- “Parameters That Are Obsolete or Have Been Removed” on page 175

Kernel Parameters

Paging-Related Parameters

fastscan (Oracle Solaris 11)
The default value of fastscan was clarified. For more information, see “fastscan” on page 47.

Process-Sizing Tunables

ngroups_max (Oracle Solaris 11)
This parameter was undocumented in previous Solaris releases. In this Solaris release, the default maximum has been increased to 1024 groups. For more information, see “ngroups_max” on page 39.
General Driver Parameter

**ddi_msix_alloc_limit (Oracle Solaris 11)**
This parameter is newly documented. For more information, see “ddi_msix_alloc_limit” on page 56.

Network Driver Parameters

**igb Parameters (Oracle Solaris 11)**
The igb network driver parameters are provided in the Oracle Solaris 11 release. For more information, see “igb Parameters” on page 57.

**ixgbe Parameters (Oracle Solaris 11)**
The ixgbe network driver parameters are provided in the Oracle Solaris 11 release. For more information, see “ixgbe Parameters” on page 58.

General Kernel and Memory Parameters

**zfs_arc_min (Oracle Solaris 11)**
This parameter description is newly documented. For more information, see “zfs_arc_min” on page 26.

**zfs_arc_max (Oracle Solaris 11)**
This parameter description is newly documented. For more information, see “zfs_arc_max” on page 27.

**disp_rechoose_interval (Oracle Solaris 11)**
This parameter is new in the Oracle Solaris 11 release. For more information, see “disp_rechoose_interval” on page 73.
TCP/IP Tunable Parameters

[tcp,sctp,udp]_smallest_anon_port and [tcp,sctp,udp]_largest_anon_port (Oracle Solaris 11)

These parameters are newly documented in the Oracle Solaris 11 release.

- “smallest_anon_port” on page 155
- “largest_anon_port” on page 155
- “smallest_anon_port” on page 135
- “largest_anon_port” on page 136
- “smallest_anon_port” on page 143
- “largest_anon_port” on page 143

IP Parameter Name Changes (Oracle Solaris 11)

In the Oracle Solaris 11 release, the following IP parameters have been renamed to IP properties.

You can set an IP property by using syntax similar to the following:

```
# ipadm set-prop -p icmp_err_interval=100 ip
```

You can display IP property information by using syntax similar to the following:

```
# ipadm show-prop -p icmp_err_interval ip
```

<table>
<thead>
<tr>
<th>PROTO PROPERTY</th>
<th>PERM CURRENT</th>
<th>PERSISTENT</th>
<th>DEFAULT</th>
<th>POSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip _icmp_err_interval</td>
<td>rw 100</td>
<td>100</td>
<td>100</td>
<td>0-99999</td>
</tr>
</tbody>
</table>

TABLE A-1 IP Parameter Name Changes

<table>
<thead>
<tr>
<th>Previous IP Parameter Name</th>
<th>IP Property Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip_addrs_per_if</td>
<td>_addr_per_if</td>
</tr>
<tr>
<td>ip_forwarding</td>
<td>forwarding (IPv4)</td>
</tr>
<tr>
<td>ip6_forwarding</td>
<td>forwarding (IPv6)</td>
</tr>
<tr>
<td>ip_forward_src_routed</td>
<td>_forward_src_routed (IPv4)</td>
</tr>
<tr>
<td>ip6_forward_src_routed</td>
<td>_forward_src_routed (IPv6)</td>
</tr>
<tr>
<td>ip_icmp_err_interval</td>
<td>_icmp_err_interval</td>
</tr>
<tr>
<td>ip_icmp_err_burst</td>
<td>_icmp_err_burst</td>
</tr>
</tbody>
</table>
TCP/IP Tunable Parameters

TABLE A-1  IP Parameter Name Changes  (Continued)

<table>
<thead>
<tr>
<th>Previous IP Parameter Name</th>
<th>IP Property Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip_icmp_return_data_bytes</td>
<td>_icmp_return_data_bytes (IPv4)</td>
</tr>
<tr>
<td>ip6_icmp_return_data_bytes</td>
<td>_icmp_return_data_bytes (IPv6)</td>
</tr>
<tr>
<td>ip_ire_pathmtu_interval</td>
<td>_pathmtu_interval</td>
</tr>
<tr>
<td>ip_respond_to_echo_broadcast</td>
<td>_respond_to_echo_broadcast (IPv4)</td>
</tr>
<tr>
<td>ip6_respond_to_echo_broadcast</td>
<td>_respond_to_echo_broadcast (IPv6)</td>
</tr>
<tr>
<td>ip_respond_to_echo_multicast</td>
<td>_respond_to_echo_multicast (IPv4)</td>
</tr>
<tr>
<td>ip6_respond_to_echo_multicast</td>
<td>_respond_to_echo_multicast (IPv6)</td>
</tr>
<tr>
<td>ip_send_redirects</td>
<td>_send_redirects (IPv4)</td>
</tr>
<tr>
<td>ip6_send_redirects</td>
<td>_send_redirects (IPv6)</td>
</tr>
<tr>
<td>ip_strict_dst_multihoming</td>
<td>hostmodel</td>
</tr>
</tbody>
</table>

TCP Parameter Name Changes (Oracle Solaris 11)

In the Oracle Solaris 11 release, the following TCP parameters have been renamed to TCP properties.

You can set a TCP property by using syntax similar to the following:

```bash
# ipadm set-prop -p _deferred_ack_interval=100 tcp
```

You can display TCP property information by using syntax similar to the following:

```bash
# ipadm show-prop -p _deferred_ack_interval tcp
```

TABLE A-2  TCP Parameter Name Changes

<table>
<thead>
<tr>
<th>Previous TCP Parameter Name</th>
<th>TCP Property Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcp_deferred_ack_interval</td>
<td>_deferred_ack_interval</td>
</tr>
<tr>
<td>tcp_local_dack_interval</td>
<td>_local_dack_interval</td>
</tr>
<tr>
<td>tcp_deferred_acks_max</td>
<td>_deferred_acks_max</td>
</tr>
<tr>
<td>tcp_local_acks_max</td>
<td>_local_acks_max</td>
</tr>
<tr>
<td>tcp_wscale_always</td>
<td>_wscale_always</td>
</tr>
<tr>
<td>tcp_tstamp_always</td>
<td>_tstamp_always</td>
</tr>
<tr>
<td>Previous TCP Parameter Name</td>
<td>TCP Property Name</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>tcp_xmit_hiwat</td>
<td>send_buf</td>
</tr>
<tr>
<td>tcp_recv_hiwat</td>
<td>recv_buf</td>
</tr>
<tr>
<td>tcp_max_buf</td>
<td>max_buf</td>
</tr>
<tr>
<td>tcp_cwnd_max</td>
<td>_cwnd_max</td>
</tr>
<tr>
<td>tcp_slow_start_initial</td>
<td>_slow_start_initial</td>
</tr>
<tr>
<td>tcp_slow_start_after_idle</td>
<td>_slow_start_after_idle</td>
</tr>
<tr>
<td>tcp_sack_permitted</td>
<td>sack</td>
</tr>
<tr>
<td>tcp_rev_src_routes</td>
<td>_rev_src_routes</td>
</tr>
<tr>
<td>tcp_time_wait_interval</td>
<td>_time_wait_interval</td>
</tr>
<tr>
<td>tcp_ecn_permitted</td>
<td>ecn</td>
</tr>
<tr>
<td>tcp_conn_req_max_q</td>
<td>_conn_req_max_q</td>
</tr>
<tr>
<td>tcp_conn_req_max_q0</td>
<td>_conn_req_max_q0</td>
</tr>
<tr>
<td>tcp_conn_req_min</td>
<td>_conn_req_min</td>
</tr>
<tr>
<td>tcp_rst_sent_rate_enabled</td>
<td>_rst_sent_rate_enabled</td>
</tr>
<tr>
<td>tcp_rst_sent_rate</td>
<td>_rst_sent_rate</td>
</tr>
<tr>
<td>tcp_keepalive_interval</td>
<td>_keepalive_interval</td>
</tr>
<tr>
<td>tcp_ip_abort_interval</td>
<td>_ip_abort_interval</td>
</tr>
<tr>
<td>tcp_rexmit_interval_initial</td>
<td>_rexmit_interval_initial</td>
</tr>
<tr>
<td>tcp_rexmit_interval_max</td>
<td>_rexmit_interval_max</td>
</tr>
<tr>
<td>tcp_rexmit_interval_min</td>
<td>_rexmit_interval_min</td>
</tr>
<tr>
<td>tcp_rexmit_interval_extra</td>
<td>_rexmit_interval_extra</td>
</tr>
<tr>
<td>tcp_tstamp_if_wscale</td>
<td>_tstamp_if_wscale</td>
</tr>
<tr>
<td>tcp_recv_hiwat_minmss</td>
<td>_recv_hiwat_minmss</td>
</tr>
</tbody>
</table>

**UDP Parameter Name Changes (Oracle Solaris 11)**

In the Oracle Solaris 11 release, the following UDP parameters have been renamed to UDP properties.

You can set a UDP property by using syntax similar to the following:
# ipadm set-prop -p send_buf=57344 udp

You can display UDP property information by using syntax similar to the following:

# ipadm show-prop -p send_buf udp

<table>
<thead>
<tr>
<th>PROTO</th>
<th>PROPERTY</th>
<th>PERM</th>
<th>CURRENT</th>
<th>PERSISTENT</th>
<th>DEFAULT</th>
<th>POSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>udp</td>
<td>send_buf</td>
<td>rw</td>
<td>57344</td>
<td>57344</td>
<td>57344</td>
<td>1024-2097152</td>
</tr>
</tbody>
</table>

**TABLE A–3**  UDP Parameter Name Changes

<table>
<thead>
<tr>
<th>Previous UDP Parameter Name</th>
<th>UDP Property Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>udp_max_buf</td>
<td>max_buf</td>
</tr>
<tr>
<td>udp_xmit_hiwat</td>
<td>send_buf</td>
</tr>
<tr>
<td>udp_recv_hiwat</td>
<td>recv_buf</td>
</tr>
</tbody>
</table>

**SCTP Parameter Name Changes (Oracle Solaris 11)**

In the Oracle Solaris 11 release, the following SCTP parameters have been renamed to SCTP properties.

You can set an SCTP property by using syntax similar to the following:

# ipadm set-prop -p _max_init_retr=8 sctp

You can display SCTP property information by using syntax similar to the following:

# ipadm show-prop -p _max_init_retr sctp

<table>
<thead>
<tr>
<th>PROTO</th>
<th>PROPERTY</th>
<th>PERM</th>
<th>CURRENT</th>
<th>PERSISTENT</th>
<th>DEFAULT</th>
<th>POSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>sctp</td>
<td>_max_init_retr</td>
<td>rw</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0-128</td>
</tr>
</tbody>
</table>

**TABLE A–4**  SCTP Parameter Name Changes

<table>
<thead>
<tr>
<th>Previous SCTP Parameter Name</th>
<th>SCTP Property Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>sctp_max_init_retr</td>
<td>_max_init_retr</td>
</tr>
<tr>
<td>sctp_pa_max_retr</td>
<td>_pa_max_retr</td>
</tr>
<tr>
<td>sctp_pp_max_retr</td>
<td>_pp_max_retr</td>
</tr>
<tr>
<td>sctp_cwnd_max</td>
<td>_cwnd_max</td>
</tr>
<tr>
<td>sctp_ipv4_ttl</td>
<td>_ipv4_ttl</td>
</tr>
<tr>
<td>sctp_heartbeat_interval</td>
<td>_heartbeat_interval</td>
</tr>
<tr>
<td>sctp_new_secret_interval</td>
<td>_new_secret_interval</td>
</tr>
<tr>
<td>sctp_initial_mtu</td>
<td>_initial_mtu</td>
</tr>
</tbody>
</table>
Parameters That Are Obsolete or Have Been Removed

The following section describes parameters that are obsolete or have been removed from more recent Solaris releases.

rstchown

This parameter is obsolete starting in the Oracle Solaris 11 release.

Description Indicates whether the POSIX semantics for the chown system call are in effect. POSIX semantics are as follows:

- A process cannot change the owner of a file, unless it is running with UID 0.
A process cannot change the group ownership of a file to a group in which it is not currently a member, unless it is running as UID 0. For more information, see `chown(2)`.

**DataType** Signed integer  
**Default** 1, indicating that POSIX semantics are used  
**Range** 0 = POSIX semantics not in force or 1 = POSIX semantics used  
**Units** Toggle (on/off)  
**Dynamic?** Yes  
**Validation** None  
**When to Change** When POSIX semantics are not wanted. Note that turning off POSIX semantics opens the potential for various security holes. Doing so also opens the possibility of a user changing ownership of a file to another user and being unable to retrieve the file without intervention from the user or the system administrator.

**Commitment Level** Obsolete

### Obsolete TCP/IP Module Parameters

**ip_multidata_outbound (Oracle Solaris 11)**  
This parameter is obsolete in the Oracle Solaris 11 release.

**tcp_mdt_max_pbufs (Oracle Solaris 11)**  
This parameter is obsolete in the Oracle Solaris 11 release.
Revision History for This Manual

This section describes the revision history for this manual.

- “Current Version: Oracle Solaris 11 Release” on page 177
- “New or Changed Parameters in the Oracle Solaris Release” on page 177

Current Version: Oracle Solaris 11 Release

The current version of this manual applies to the Oracle Solaris 11 release.

New or Changed Parameters in the Oracle Solaris Release

The following sections describe new, changed, or obsolete kernel tunables.

- Oracle Solaris 11: The `rsc_hown` parameter is obsolete. For more information, see “What’s New in Oracle Solaris System Tuning?” on page 17.
- Oracle Solaris 11: The `ipadm` command replaces the `ndd` command for setting TCP, IP, UDP, and SCTP properties. In addition, the names of the network parameters have changed to better correlate to the `ipadm` format. For more information, see “Overview of Tuning IP Suite Parameters” on page 117.
- Oracle Solaris 11: This release includes the `disp_rechoose_interval` parameter. For more information, see “disp_rechoose_interval” on page 73.
- Oracle Solaris 11: This release includes the `ngroups_max` parameter description. For more information, see “ngroups_max” on page 39.
- Oracle Solaris 11: This release includes the `zfs_arc_min` and `zfs_arc_max` parameter descriptions. For more information, see “zfs_arc_min” on page 26 and “zfs_arc_max” on page 27.
- Oracle Solaris 11: This release includes several `igb` and `ixgbe` network driver parameters. For more information, see “igb Parameters” on page 57 and “ixgbe Parameters” on page 58.
Oracle Solaris 11: This release includes the `ddi_msix_alloc_limit` parameter that can be used to increase the number of MSI-X interrupts that a device instance can allocate. For more information, see "ddi_msix_alloc_limit" on page 56.

Oracle Solaris 11: This release includes the `kmem_stackinfo` parameter, which can be enabled to monitor kernel thread stack usage. For more information, see "kmem_stackinfo" on page 54.

Oracle Solaris 11: Memory locality group parameters are provided in this release. For more information about these parameters, see "Locality Group Parameters" on page 79.

Oracle Solaris 11: Parameter information was updated to include sun4v systems. For more information, see the following references:
- "maxphys" on page 62
- "tmpfs:tmpfs_maxkmem" on page 66
- "SPARC System Specific Parameters" on page 75
Index

A
_addip_enabled, 154
_addrs_per_if, 121
autofs, 164
autoup, 33
dopageflush, 34

E
ecn, 132
enable_tsb_rss_sizing, 78

F
fastscan, 47
forwarding, 120
fs, 165
fsflush, 32
ftp, 165

G
gc_intr_mode, 161

H
handspreadpages, 49
_heartbeat_interval, 147
hires_tick, 74
hoplimit (ipv6), 120
hostmodel, 121

D
ddi_msix_alloc_limit parameter, 56
default_stksize, 28
default_tsb_size, 77
_deferred_ack_interval, 124, 148
_deferred_acks_max, 125
desfree, 42
dhcpagent, 164
disp_rechoose_interval, 73, 170
dnlc_dir_enable, 65
dnlc_dir_max_size, 66
dnlc_dir_min_size, 65
doiflush, 35
I
  _icmp_err_burst, 118
  _icmp_err_interval, 118
  _icmp_return_data_bytes, 123
  _ignore_path_mtu, 149
  inetinit, 165
  init, 166
  _initial_mtu, 148
  _initial_out_streams, 153
  _initial_ssthresh, 149
  intr_force, 57
  intr_throttling, 59
  _ip_abort_interval, 138
  ip_squeue_fanout, 122
  ip_squeue_worker_wait, 137
  ipcl_conn_hash_size, 136
  ipsec, 166
  _ipv4_ttl, 146
  _ipv6_hoplimit, 147

K
  kbd, 166
  _keealive_interval, 138
  keyserv, 166
  kmem_flags, 53
  kmem_stackinfo, 54

L
  largest_anon_port, 136, 143, 155
  lgrp_mem_pset_aware, 81
  _local_dack_interval, 124
  _local_dacks_max, 126
  logevent_max_q_sz, 30
  login, 167
  lotsfree, 41
  lpg_alloc_prefer, 79
  lpg_mem_default_policy, 80
  lwp_default_stksize, 29

M
  max_buf (SCTP), 151
  max_buf (TCP), 128
  max_buf (UDP), 142
  _max_in_streams, 152
  _max_init_retr, 145
  max_nprocs, 38
  maxpgio, 50
  maxphys, 62
  maxpid, 37
  maxuprc, 39
  maxusers, 36
  min_percent_cpu, 48
  minfree, 43
  moddebug, 55
  mpathd, 167
  mr_enable, 57

N
  nca_conn_hash_size, 158
  nca_conn_req_max_q, 158
  nca_conn_req_max_q0, 158
  nca_ppmax, 159
  nca_vpmx, 159
  ncsize, 64
  nedd, 118
  _new_secret_interval, 148
  nfs_max_threads, 93
  nfs:nacache, 106
  nfs:nfs_allow_preeepoch_time, 85
  nfs:nfs_async_clusters, 103
  nfs:nfs_async_timeout, 106
  nfs:nfs_cots_timeo, 86
  nfs:nfs_disable_rdir_cache, 101
  nfs:nfs_do_symlink_cache, 88
  nfs:nfs_dynamic, 89
  nfs:nfs_lookup_neg_cache, 90
  nfs:nfs_nra, 95
  nfs:nfs_shrinkreaddir, 98
  nfs:nfs_write_error_interval, 100
  nfs:nfs_write_error_to_cons_only, 100
  nfs:nfs3_async_clusters, 104
  nfs:nfs3_bsize, 102
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nfs:nfs3_cots_timeo</td>
<td>86</td>
</tr>
<tr>
<td>nfs:nfs3_do_symlink_cache</td>
<td>88</td>
</tr>
<tr>
<td>nfs:nfs3_dynamic</td>
<td>90</td>
</tr>
<tr>
<td>nfs:nfs3_jukebox_delay</td>
<td>107</td>
</tr>
<tr>
<td>nfs:nfs3_lookup_neg_cache</td>
<td>91</td>
</tr>
<tr>
<td>nfs:nfs3_max_threads</td>
<td>94</td>
</tr>
<tr>
<td>nfs:nfs3_max_transfer_size</td>
<td>108</td>
</tr>
<tr>
<td>nfs:nfs3_max_transfer_size_clts</td>
<td>110</td>
</tr>
<tr>
<td>nfs:nfs3_max_transfer_size_cots</td>
<td>110</td>
</tr>
<tr>
<td>nfs:nfs3_nra</td>
<td>96</td>
</tr>
<tr>
<td>nfs:nfs3_pathconf_disable_cache</td>
<td>84</td>
</tr>
<tr>
<td>nfs:nfs3_shrinkreaddir</td>
<td>99</td>
</tr>
<tr>
<td>nfs:nfs4_async_clusters</td>
<td>105</td>
</tr>
<tr>
<td>nfs:nfs4_bsize</td>
<td>102</td>
</tr>
<tr>
<td>nfs:nfs4_cots_timeo</td>
<td>87</td>
</tr>
<tr>
<td>nfs:nfs4_do_symlink_cache</td>
<td>89</td>
</tr>
<tr>
<td>nfs:nfs4_lookup_neg_cache</td>
<td>92</td>
</tr>
<tr>
<td>nfs:nfs4_max_threads</td>
<td>95</td>
</tr>
<tr>
<td>nfs:nfs4_max_transfer_size</td>
<td>109</td>
</tr>
<tr>
<td>nfs:nfs4_nra</td>
<td>97</td>
</tr>
<tr>
<td>nfs:nfs4_pathconf_disable_cache</td>
<td>84</td>
</tr>
<tr>
<td>nfs:nrnode</td>
<td>97</td>
</tr>
<tr>
<td>nfslogd</td>
<td>167</td>
</tr>
<tr>
<td>ngroups_max</td>
<td>39, 169</td>
</tr>
<tr>
<td>noexec_user_stack</td>
<td>31</td>
</tr>
<tr>
<td>nss</td>
<td>167</td>
</tr>
<tr>
<td>nstrpush</td>
<td>70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>recv_buf (SCTP)</td>
<td>150</td>
</tr>
<tr>
<td>recv_buf (TCP)</td>
<td>128</td>
</tr>
<tr>
<td>recv_buf (UDP)</td>
<td>142</td>
</tr>
<tr>
<td>_recv_hiwat_minmss</td>
<td>141</td>
</tr>
<tr>
<td>reserved_procs</td>
<td>37</td>
</tr>
<tr>
<td>_respond_to_echo_broadcast</td>
<td>119</td>
</tr>
<tr>
<td>_respond_to_echo_multicast</td>
<td>119</td>
</tr>
<tr>
<td>_rev_src_routes</td>
<td>131</td>
</tr>
<tr>
<td>_rexmit_interval_extra</td>
<td>140</td>
</tr>
<tr>
<td>_rexmit_interval_initial</td>
<td>139</td>
</tr>
<tr>
<td>_rexmit_interval_max</td>
<td>139</td>
</tr>
<tr>
<td>_rexmit_interval_min</td>
<td>140</td>
</tr>
<tr>
<td>rlim_fd_cur</td>
<td>63</td>
</tr>
<tr>
<td>rlim_fd_max</td>
<td>62</td>
</tr>
<tr>
<td>rpcmod:clnt_idle_timeout</td>
<td>112</td>
</tr>
<tr>
<td>rpcmod:clnt_max_conns</td>
<td>111</td>
</tr>
<tr>
<td>rpcmod:cotsmaxdupreqs</td>
<td>114</td>
</tr>
<tr>
<td>rpcmod:maxdupreqs</td>
<td>113</td>
</tr>
<tr>
<td>rpcmod:svc_default_stksize</td>
<td>112</td>
</tr>
<tr>
<td>rpcmod:svc_idle_timeout</td>
<td>112</td>
</tr>
<tr>
<td>_rst_sent_rate</td>
<td>135</td>
</tr>
<tr>
<td>_rst_sent_rate_enabled</td>
<td>134</td>
</tr>
<tr>
<td>rstchown</td>
<td>175</td>
</tr>
<tr>
<td>_rto_max</td>
<td>151, 152</td>
</tr>
<tr>
<td>_rto_min</td>
<td>151</td>
</tr>
<tr>
<td>rx_copy_threshold</td>
<td>61</td>
</tr>
<tr>
<td>rx_limit_per_intr</td>
<td>59</td>
</tr>
<tr>
<td>rx_queue_number</td>
<td>58</td>
</tr>
<tr>
<td>rx_ring_size</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sack</td>
<td>130</td>
</tr>
<tr>
<td>scctp_maxburst</td>
<td>153</td>
</tr>
<tr>
<td>segspt_minfree</td>
<td>72</td>
</tr>
<tr>
<td>send_buf (SCTP)</td>
<td>149</td>
</tr>
<tr>
<td>send_buf (TCP)</td>
<td>127</td>
</tr>
<tr>
<td>send_buf (UDP)</td>
<td>141</td>
</tr>
<tr>
<td>_send_redirects</td>
<td>119</td>
</tr>
<tr>
<td>_shutack_wait_bound</td>
<td>153</td>
</tr>
<tr>
<td>_slow_start_after_idle</td>
<td>129</td>
</tr>
<tr>
<td>_slow_start_initial</td>
<td>129</td>
</tr>
<tr>
<td>slowscan</td>
<td>48</td>
</tr>
</tbody>
</table>
smallest_anon_port, 135, 143, 155
sq_max_size, 160
strmsgs, 70, 71
su, 168
sun4v, 75
swapfs_minfree, 52
swapfs_reserve, 51
syslog, 168

T

tar, 168
throttlefree, 44
_time_wait_interval, 131
timer_max, 75
tmpfs_maxkmem, 66
tmpfs_minfree, 67
tsb_alloc_hiwater, 76
tsb_rss_size, 78
_tstamp_always, 127
_tstamp_if_wscale, 141
ttl (ipv4), 120
tune_t_fshelfr, 33
tune_t_minarmem, 47
tx_copy_threshold, 61
tx_queue_number, 58
tx_ring_size, 60

U

utmpd, 168

W

_wscale_always, 126

X

_xmit_lowat, 150

Z

zfs_arc_max, 27, 170
zfs_arc_min, 26, 170