# Contents

About This Manual .......................................................... v
   Introductory Remarks .................................................... v
   Understanding This Document ........................................... vi
   Notation ............................................................... vii
   Using the API .......................................................... viii
   Revision History ........................................................ ix

1 Basic Kernel Interface .................................................. 1
   1.1 Kernel Interface Page .................................................. 2
   1.2 KERNELINTERFACE .................................................... 7
   1.3 Virtual Registers .................................................... 11

2 Threads ................................................................. 13
   2.1 ThreadId ............................................................. 14
   2.2 Thread Control Registers (TCRs) .................................... 16
   2.3 EXCHANGEREGISTERS ................................................ 18
   2.4 THREADCONTROL .................................................... 22

3 Scheduling .............................................................. 25
   3.1 Clock ............................................................... 26
   3.2 SYSTEMLOCK ........................................................ 27
   3.3 Time ............................................................... 28
   3.4 THREADSWITCH ..................................................... 30
   3.5 SCHEDULE ........................................................ 31
   3.6 Preempt Flags ....................................................... 34

4 Address Spaces and Mapping .......................................... 35
   4.1 Fpage .............................................................. 36
   4.2 UNMAP ............................................................. 38
   4.3 SPACECONTROL ..................................................... 41

5 IPC ................................................................. 43
   5.1 Messages And Message Registers (MRs) .............................. 44
   5.2 MapItem ............................................................ 49
   5.3 GrantItem ........................................................... 51
   5.4 StringItem .......................................................... 52
   5.5 String Buffers And Buffer Registers (BRs) ......................... 55
   5.6 IPC ............................................................... 57

6 Miscellaneous .......................................................... 65
   6.1 ExceptionHandler .................................................... 66
   6.2 Cop Flags ........................................................... 67
   6.3 PROCESSORCONTROL ............................................... 68
   6.4 MEMORYCONTROL .................................................. 70

7 Protocols .............................................................. 73
   7.1 Thread Start Protocol ................................................ 74
   7.2 Interrupt Protocol .................................................... 75
   7.3 Pagefault Protocol .................................................... 76
   7.4 Preemption Protocol .................................................. 77
   7.5 Exception Protocol ................................................... 78
   7.6 Sigma0 RPC protocol .................................................. 79
   7.7 Generic Booting ...................................................... 82
A IA-32 Interface  
A.1 Virtual Registers .......................................................... 86  
A.2 Systemcalls .................................................................. 89  
A.3 Kernel Features ............................................................. 92  
A.4 IO-Ports ...................................................................... 93  
A.5 Space Control ............................................................... 94  
A.6 Cacheability Hints .......................................................... 95  
A.7 Memory Attributes .......................................................... 96  
A.8 Exception Message Format ............................................... 97  
A.9 Processor Mirroring ........................................................ 98  
A.10 Booting ................................................................... 99  

B IA-64 Interface  
B.1 Virtual Registers ............................................................ 102  
B.2 PAL and SAL Access ....................................................... 104  
B.3 Systemcalls .................................................................. 105  
B.4 PCI Configuration Space .................................................. 110  
B.5 Cacheability Hints ........................................................... 111  
B.6 Memory Attributes ........................................................ 112  
B.7 Memory Descriptors ......................................................... 113  
B.8 Exception Message Format ............................................... 114  

C PowerPC Interface  
C.1 Virtual Registers ............................................................ 116  
C.2 Systemcalls .................................................................. 118  
C.3 Memory Attributes .......................................................... 122  
C.4 Exception Message Format ............................................... 123  
C.5 Processor Mirroring ........................................................ 125  
C.6 Booting ................................................................... 126  

D Alpha Interface  
D.1 Virtual Registers ............................................................ 128  
D.2 Systemcalls .................................................................. 130  
D.3 Booting ................................................................... 134  

E MIPS64 Interface  
E.1 Virtual Registers ............................................................ 136  
E.2 Systemcalls .................................................................. 138  
E.3 Memory Attributes .......................................................... 143  
E.4 Booting ................................................................... 144  

F Development Remarks  
F.1 Exception Handling .......................................................... 145  

Table of Procs, Types, and Constants ........................................ 147  

Index ........................................................................... 153
About This Manual

Introductory Remarks

Purpose of This Document
This L4 Reference Manual serves as defining document for all L4 APIs and ABIs. Primarily, it addresses L4 microkernel implementors as API/ABI suppliers and code-generator or library implementors as API/ABI users. The reference manual assumes intimate knowledge of basic L4 concepts and hardware architecture. Its key point is precise definition, not explanation and illustration. The

L4 System Programmer’s Manual

is intended to support programmers using L4. It explains and illustrates fundamental concepts and describes in more detail how (and why) to use which function, etc.

Maintainers
The document is maintained by the following members of the L4Ka Team:

- Uwe Dannowski (ud3@ira.uka.de)
- Joshua LeVasseur (jtl@ira.uka.de)
- Espen Skoglund (esk@ira.uka.de)
- Volkmar Uhlig (volkmar@ira.uka.de)

Credits
This manual is based on a final draft by Jochen Liedtke. It reflects his outstanding work on the L4 microkernel and systems research in general. Only his vision of system design made this work possible. Jochen defined the state of the art of microkernel design for nearly a decade. We thank him for his support and try to continue the work in his spirit.


Document History

draft by Jochen Liedtke ??/?? - 06/01
review by L4Ka Team 06/01 - 09/01
L4 developers review Q4/01
release 01/02
Understanding This Document

This L4 Reference Manual defines the generic API for all 32-bit and 64-bit machines. As such, the generic reference manual is independent of specific processor architectures. It is complemented by processor-specific ABI specifications. Some of them can be found in the appendix of this document.

In this document, we differentiate between Logical Interface, Generic Binary Interface, Generic Programming Interface, Convenience Programming Interface and Processor-specific Binary Interface.

Logical Interface

The logical interface defines all concepts and logical objects such as system-call operations, logical data objects, data types and their semantics. Altogether, they form the logical L4 API.

Generic Binary Interface

Binary representations of most data types and generic data objects are defined independently of specific processors (although there are two different versions, one for 32-bit and a second one for 64-bit processors). Both versions together form the generic binary interface of L4.

From a purist point of view, logical interface plus generic binary interface could be regarded as a complete specification of the hardware-independent L4 microkernel interface. However, for ease-of-use and standardization reasons, the mentioned two fundamental interfaces are complemented by two more interface classes:

Generic Programming Interface

The generic programming interface defines the objects of the logical interface and the generic binary interface as pseudo C++ classes. The language bindings for regular C is for the most part identical to C++. For the cases where the C language causes function naming conflicts, the C version of the function name is given in brackets. For the time being, only the C and C++ versions of the API are specified. The concrete syntax of other language interfaces will be left open. Later on, all language bindings will be included in the generic programming interface.

Convenience Programming Interface

This interface is not part of the L4 microkernel specification in the strict sense. All of its data types and procedures can be implemented using the generic programming interface. Strictly speaking, it is an interface on top of the microkernel that makes the most common operations more easily usable for the programmer.

It is important to understand that convenience and ease-of-use, not completeness, is the criterion for this interface. The convenience programming interface supports programmers by offering operations that together cover about 95% of the required microkernel functionality. For the remaining 5%, the programmer has to use the basic (not so convenient) operations of the generic programming interface.

Obviously, the convenience programming interface is not mandatory. Consequently, from a minimalist point of view, there is no need to include it in the generic L4 specification.

Nevertheless, for reasons of standardization and thus portability of software, every complete L4 language binding has to include the entire convenience programming interface.

Implementation remark: Although the convenience interface can be completely implemented on top of the generic programming interface, i.e., processor independently, the implementor of the convenience interface may implement it hardware-dependently and thus incorporate any optimization that becomes possible through a specific processor-specific binary interface.

The last interface class is not part of the generic L4 API specification.

Processor-specific Binary Interface

Defines the processor-specific binary interface.
Notation

Basic Data Types
This reference manual describes the L4 API and ABI for both 32-bit and 64-bit processors. The data type Word denotes a 32-bit unsigned integer on a 32-bit processor and a 64-bit unsigned integer on a 64-bit processor. Word64, Word32, and Word16 denote 64, 32, and 16-bit words independent of the processor type.

Privileged Threads
Some system calls can only be executed by privileged threads. Any thread belonging to the same address space as one of the initial threads created by the kernel upon boot-time (see page 82) are treated as privileged.

Bit Fields
Bit-field lengths are denoted as subscripts $i/j$ where $i$ relates to a 32-bit processor and $j$ to a 64-bit processor. Bit-field subscripts $(i)$ specify bit fields that have the same size for both 32-bit and 64-bit processors. Byte offsets are given as $\pm i / \pm j$ for 32-bit and 64-bit processors. If all bit-fields of a specified word only adds up to 32 bits, the remaining upper 32 bits on 64-bit processors are undefined or ignored.

Undefined, Ignored, and Unchanged

\[\sim\] Output parameters or bit fields can be undefined. Corresponding parameters or fields are denoted by $\sim$. They have no defined value on output, i.e., they may have any value or may even be unaccessible. Any algorithm relying on the value of undefined parameters or bit fields is defined to be incorrect.

\[-\] Input parameters or bit fields can be specified as ignored, denoted by $\sim$. Such parameters or fields can hold any value without affecting the invoked service. $\sim$ is also used to define bit fields that are available for additional information. For example, fpage denotations contain some ignored bits that are used for access control bits in some system calls.

\[\equiv\] In processor-specific interfaces, registers are sometimes defined to be unchanged. This is denoted by $\equiv$.

Upward Compatibility
The following holds for future API versions and sub-versions that are specified as upward-compatible to the current version.

Output parameters and bit fields.
Fields currently defined as undefined ($\sim$) may be specified as defined. Such newly defined fields will only deliver additional information. They can be ignored if the system call is used exactly like specified in the current API.

Input parameters and bit fields.
Fields currently defined as ignored ($\sim$) may be specified as defined. However, the content of such fields will be only relevant for newly defined features. Such fields will be ignored if a system call is used with the “old” semantics specified in this API.
Using the API

Naming
A programmer can use all function, type, and constant definitions defined in the generic and convenience programming interfaces throughout this manual. All definitions must, however, be prefixed with the string “L4,” and type names must contain the “_t” suffix (e.g., use “L4_Ipc ()” and “L4_MsgTag_t” rather than “Ipc ()” and “MsgTag”). The interfaces are currently only defined for C++ and C. In some cases the naming used for function names causes conflicts in the C language. These conflicts must be resolved using the alternative name specified in brackets after the function definition.

Include Files
The relevant include files containing the required definitions and declarations are specified in the beginning of the generic and convenience interface sections. In general there is one include file for each chapter in the manual. If only the basic L4 data types are needed they can be included using <l4/types.h>.
Revision History

Revision 1
Initial revision.

Revision 2
- Clarified the specification of the kernel-interface page and kernel configuration page magic.
- UntypedWords and StringItems Acceptor constants collided with function UntypedWords(MsgTag) and StringItems(MsgTag) function declaration. Renamed to UntypedWordsAcceptor and StringItemsAcceptor.
- Changed kernel ids for L4Ka kernels.
- Fixed return types for operators on the Time type.
- Changed \texttt{wrx} access rights in fpages to \texttt{rwx}. Also changed \texttt{WRX} reference bits in fpages returned from \texttt{UNMAP} system call to \texttt{RW X}.
- Renamed Put functions operating on MsgBuffer to Append.
- Address space deletion is now performed by deleting the last thread of an AS. This makes creation and deletion symmetrical (via ThreadControl). Before, all threads but the last were deleted by ThreadControl, and the last by SpaceControl.
- Added functions for creating ThreadIDs and for retrieving version and thread numbers from them. Fixed size of MyLocalId and MyGlobalId TCRs.
- Specified that the first three thread version numbers available for user threads are dedicated to \( \sigma_0 \), \( \sigma_1 \), and root task respectively.
- Changed the encoding of \( \mu \) in the magic field of the KIP back to 0xE6 to be compatible with previous versions of the kernel.
- Changed memory descriptors (e.g., dedicated memory) in the kernel-interface page and kernel configuration page to use an array of typed descriptors instead of a static number of predefined ones.
- Added an appendix for the PowerPC interface.
- Added Niltag MsgTag constant.
- Decreased size of MsgBuffer structure to 32.
- Changed single Fpage& argument of Unmap() and Flush() into pass by value.
- Changed the ia32 kernel feature string “small” to “smallspaces”.
- Added appendix for the ia64 interface.
- Changed the ia32 IPC and LIPC ABI to be better suitable for common hardware featuring sysenter/sysexit and gcc.
- Added ProcDesc convenience functions.
- Specified which include files to use for the various parts of the API.
- Allow privileged threads to access ia32 Model-Specific Registers.
- Changed the ia64 ABI for system-call links and the IPIC and LIPC system-calls.
- The UTCB location of a new thread is now explicitly specified by a parameter to the \texttt{THREADCONTROL} system-call.
- Added C versions of conflicting function names.
– Added a number of convenience functions for fpages, map items, grant items, string items and kernel interface page fields.
– Added description of the send base in map and grant items.
– Changed subversion numbering for Version X.2 and Version 4 API.
– Renamed the XferTimeout TCR to XferTimeouts and split into separate send and receive timeouts.
– Added two thread specific words to each the architecture specific TCR sections. These words are free to be used by, e.g., IDL compilers.
– Changed name of L4Ka kernels to the official name. Added L4Ka::Strawberry.
– Added appendices for Alpha and MIPS64.

Revision 3
– Clarified description of the supplier field in the kernel-interface page.
– Added NumMemoryDescriptors() convenience function.
– Clarified the return value of MemoryDescType() function.
– Fixed faulty specification of WaitTimeout() and ReplyWaitTimeout().
– Added a new h-flag to control parameter in the EXCHANGEREGISTERS system-call. The h-flag controls whether the resume/halt flag should be ignored or not.
– Changed parameter type of TimePeriod() from “int” to “Word64”.
– Fixed typo in specification of the MsgTag input/output IPC parameter.
– Added comment to IPC system-call about the read-once semantics of message registers.
– Added member name “raw” to all L4 types declared as structs.
– Renamed start() and stop() functions to Start() and Stop().
– Describe semantics of undefined UTCB memory regions.
– The first 10 message registers on PowerPC are now defined as backed by physical registers.
– The first 9 message registers on Alpha are now defined as backed by physical registers.
– Fixed MR0 register allocation for IA32 syscalls and adapted syscalls accordingly.
Basic Kernel Interface
# 1.1 Kernel Interface Page

The kernel-interface page contains API and kernel version data, system descriptors including memory descriptors, and system-call links. The remainder of the page is undefined.

The page is a microkernel object. It is directly mapped through the microkernel into each address space upon address-space creation. It is not mapped by a pager, cannot be mapped or granted to another address space and cannot be unmapped. The creator of a new address space can specify the address where the kernel interface page has to be mapped. This address will remain constant throughout the lifetime of that address space. Any thread can obtain the address of the kernel interface page through the `KERNELINTERFACE` system call (see page 7).

<table>
<thead>
<tr>
<th>L4 version parts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier</td>
<td>KernDescPtr</td>
</tr>
<tr>
<td>KernelVer</td>
<td></td>
</tr>
<tr>
<td>KernelGenDate</td>
<td></td>
</tr>
<tr>
<td>KernelId</td>
<td></td>
</tr>
<tr>
<td>InternalFreq</td>
<td>ProcDescPtr</td>
</tr>
<tr>
<td>ExternalFreq</td>
<td></td>
</tr>
<tr>
<td>MemoryDesc</td>
<td>MemDescPtr</td>
</tr>
</tbody>
</table>

```
<table>
<thead>
<tr>
<th>EXCHANGE REGISTERS</th>
<th>SCHEDULE</th>
<th>THREADSWITCH</th>
<th>SYSTEM CLOCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProcessorInfo</td>
<td>PageInfo</td>
<td>ThreadInfo</td>
<td>ClockInfo</td>
</tr>
<tr>
<td>ProcDescPtr</td>
<td>BootInfo</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>KipAreaInfo</td>
<td>UtcbInfo</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>MemoryInfo</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>KernDescPtr</td>
<td>API Flags</td>
<td>API Version</td>
<td>~(0/32)</td>
</tr>
<tr>
<td>+C / +18</td>
<td>+8 / +10</td>
<td>+4 / +8</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Note that this kernel interface page is basically upward compatible to the kernel info page of versions 2 and X.0. Only the former clock field is missing and now used differently. The magic byte string “L4µK” at the beginning of the object identifies the kernel interface page.

**Version/id number convention:** Version/subversion/subsubversion numbers and id/subid numbers with the most significant bit 0 denote official versions/ids and are globally unique through all suppliers. Version/id numbers that have the most significant bit set to 1 denote experimental versions/ids and may be unique only in the context of a supplier.

### API Description

**API Version**

<table>
<thead>
<tr>
<th>version (8)</th>
<th>subversion (8)</th>
<th>~ (16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0x83</td>
<td>0x80</td>
<td>Version 2</td>
</tr>
<tr>
<td>0x83</td>
<td>0x81</td>
<td>Experimental Version X.1</td>
</tr>
<tr>
<td>0x84</td>
<td>rev</td>
<td>Experimental Version X.2 (Revision rev)</td>
</tr>
<tr>
<td>0x04</td>
<td>rev</td>
<td>Version 4 (Revision rev)</td>
</tr>
</tbody>
</table>

**API Flags**

<table>
<thead>
<tr>
<th>ee</th>
<th>~ (28/60)</th>
<th>ww</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>= 00 : little endian,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>= 01 : big endian.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ww</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>= 00 : 32-bit API,</td>
<td></td>
</tr>
<tr>
<td>= 01 : 64-bit API.</td>
<td></td>
</tr>
</tbody>
</table>

Note that this field can not be used directly to differentiate between little endian and big endian mode since the ee field resides in different bytes for both modes. Furthermore, the offset address of the API Flags is different for 32-bit and 64-bit modes. In summary, a direct inspection of the kernel interface page is not sufficient to securely differentiate between 32/64-bit modes and little/big endian modes.

Secure mode detection is enabled through the KERNELINTERFACE system call (see page 7). It delivers the API Flags in a register.

### System Description

**ProcessorInfo**

<table>
<thead>
<tr>
<th>s (4)</th>
<th>~ (12/44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>processors − 1 (16)</td>
</tr>
</tbody>
</table>

s

The size of the area occupied by a single processor description is $2^s$. Location of description fields for the first processor is denoted by ProcDescPtr. Description fields for subsequent processors are located directly following the previous one.

processors

Number of available system processors.

**PageInfo**

<table>
<thead>
<tr>
<th>page-size mask (22/54)</th>
<th>~ (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w x</td>
</tr>
</tbody>
</table>

**page-size mask**

If bit $k − 10$ of the page-size mask field (bit $k$ of the entire word) is set to 1 hardware and kernel support pages of size $2^k$. If the bit is 0 hardware and/or kernel do not support pages of size $2^k$. Note that pages of size $2^k$ can be used, even if $2^k$ is no supported hardware page size. Information about supported hardware page sizes is only a performance hint.
Identifies the supported access rights (read, write, execute) that can be set independently of other access rights. A 1-bit signals that the right can be set and reset on a mapped page. For $rwx = 010$, only write permission could be controlled orthogonally. The processor would implicitly permit read and execute access on any mapped page. For $rwx = 111$, all three rights could be set and reset independently.

### ThreadInfo

<table>
<thead>
<tr>
<th>UserBase (12)</th>
<th>SystemBase (12)</th>
<th>t (8)</th>
</tr>
</thead>
</table>

Number of valid thread-number bits. The thread number field may be larger but only bits $0 \ldots t - 1$ are significant for this kernel. Higher bits must all be 0.

### SystemBase

Lowest thread number used for system threads (see page 14). Thread numbers below this value denote hardware interrupts.

### ClockInfo

<table>
<thead>
<tr>
<th>SchedulePrecision (16)</th>
<th>ReadPrecision (16)</th>
</tr>
</thead>
</table>

ReadPrecision

Specifies the minimal time difference $\neq 0$ that can be detected by reading the system clock through the `SYSTEMCLOCK` system call. Basically, this is the precision of the system clock when reading it.

SchedulePrecision

Specifies the maximal jitter ($\pm$) for a scheduled thread activation based on a wakeup time (provided that no thread of higher or equal priority is active and timer interrupts are enabled). Precisions are given as time periods (see page 28).

### UtcbInfo

<table>
<thead>
<tr>
<th>~ (10/42)</th>
<th>s (6)</th>
<th>a (6)</th>
<th>m (10)</th>
</tr>
</thead>
</table>

$s$
The minimal area size for an address space’s UTCB area is $2^s$. The size of the UTCB area limits the total number of threads $k$ to $2^a mk \leq 2^s$.

$m$
UTCB size multiplier.

$a$
The UTCB location must be aligned to $2^a$. The total size required for one UTCB is $2^a m$.

### KipAreaInfo

<table>
<thead>
<tr>
<th>~ (26/58)</th>
<th>s (6)</th>
</tr>
</thead>
</table>

$s$
The size of the kernel interface page area is $2^s$.

### BootInfo

Prior to kernel initialization a boot loader can write an arbitrary value into the BootInfo field of the kernel configuration page (see page 82). Post-initialization code, e.g., a root server can later read the field from the kernel interface page. Its value is neither changed nor interpreted by the kernel. This is a generic method for passing system information across kernel initialization.

### Processor Description

#### ProcDescPtr

Points to an array containing a description for each system processor. The ProcessorInfo field contains the dimension of the array. ProcDescPtr is given as an address relative to the kernel interface page’s base address.

#### ExternalFreq

External Bus frequency in kHz.
InternalFreq
Internal processor frequency in kHz.

**Kernel Description**

**KernDescPtr**
Points to a region that contains 4 kernel-version words (see below) followed by a number of 0-terminated plaintext strings. The first plaintext string identifies the current kernel followed by further optional kernel-specific versioning information. The remaining plaintext strings identify architecture dependent kernel features (see Appendix A.3). A zero length string (i.e., a string containing only a 0-character) terminates the list of feature descriptions. KernDescPtr is given as an address relative to the kernel interface page’s base address.

**KernelId**
Can be used to identify the microkernel.

<table>
<thead>
<tr>
<th>id</th>
<th>subid</th>
<th>kernel</th>
<th>supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>L4/486</td>
<td>GMD</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>L4/Pentium</td>
<td>IBM</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>L4/x86</td>
<td>UKa</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>L4/Mips</td>
<td>UNSW</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>L4/Alpha</td>
<td>TUD, UNSW</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Fiasco</td>
<td>TUD</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>L4Ka::Hazelnut</td>
<td>UKa</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>L4Ka::Pistachio</td>
<td>UKa</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>L4Ka::Strawberry</td>
<td>UKa</td>
</tr>
</tbody>
</table>

**KernelGenDate**
Kernel generation date.

| ~ (16/48) | year-2000 (7) | month (4) | day (5) |

**KernelVer**
Can be used to identify the microkernel version. Note that this kernel version is not necessarily related to the API version.

**Supplier**
The four least significant bytes of the supplier field specify a character string identifying the kernel supplier:

- “GMD”
- “IBM”
- “UNSW”
- “TUD”
- “UKa”

- “GMD” – GMD
- “IBM” – IBM Research
- “UNSW” – University of New South Wales, Sydney
- “TUD” – Technische Universität Dresden
- “UKa” – Universität Karlsruhe (TH)

**System-Call Links**

**SC**
Link for normal system call.

**pSC**
Link for privileged system call, i.e., a system call that can only be performed by a privileged thread.

The system-call links specify how the application can invoke system-calls for the current microkernel. The interpretation of the system-call links is ABI specific, but will typically be addresses relative to the kernel interface page’s base address where kernel provided system-call stubs are located.
**Memory Description**

**MemoryInfo**

| MemDescPtr (16/32) | n (16/32) |

**MemDescPtr**
Location of first memory descriptor (as an offset relative to the kernel-interface page’s base address). Subsequent memory descriptors are located directly following the first one. For memory descriptors that specify overlapping memory regions, later descriptors take precedence over earlier ones.

**n**
Number of memory descriptors.

**MemoryDesc**

<table>
<thead>
<tr>
<th>high/2(^{10}) (22/54)</th>
<th>~ (10)</th>
<th>+4 / +8</th>
</tr>
</thead>
<tbody>
<tr>
<td>low/2(^{10}) (22/54)</td>
<td>v</td>
<td>t (4)</td>
</tr>
<tr>
<td></td>
<td>~</td>
<td>type (4)</td>
</tr>
</tbody>
</table>

**high**
High address of memory region.

**low**
Low address of memory region.

**v**
Indicates whether memory descriptor refers to physical memory \((v = 0)\) or virtual memory \((v = 1)\).

**type**
Identifies the type of the memory descriptor.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0</td>
<td>Undefined</td>
</tr>
<tr>
<td>0x1</td>
<td>Conventional memory</td>
</tr>
<tr>
<td>0x2</td>
<td>Reserved memory (i.e., reserved by kernel)</td>
</tr>
<tr>
<td>0x3</td>
<td>Dedicated memory (i.e., memory not available to user)</td>
</tr>
<tr>
<td>0x4</td>
<td>Shared memory (i.e., available to all users)</td>
</tr>
<tr>
<td>0xE</td>
<td>Defined by boot loader</td>
</tr>
<tr>
<td>0xF</td>
<td>Architecture dependent</td>
</tr>
</tbody>
</table>

\(t, \ type = 0xE\)
The type of the memory descriptor is dependent on the bootloader. The \(t\) field specifies the exact semantics. Refer to bootloader specification for more info.

\(t, \ type = 0xF\)
The type of the memory descriptor is architecture dependent. The \(t\) field specifies the exact semantics. Refer to architecture specific part for more info (see page 113).

\(t, \ type \neq 0xE, \ type \neq 0xF\)
The type of the memory descriptor is solely defined by the \(type\) field. The content of the \(t\) field is undefined.
1.2 **KERNELINTERFACE**  [Slow Systemcall]

Delivers base address of the *kernel interface page*, *API version*, and *API flags*. The latter two values are copies of the corresponding fields in the kernel interface page. The API information is delivered in registers through this system call (a) to enable unrestricted structural changes of the kernel interface page in future versions, and (b) to enable secure detection of the kernel’s endian mode (little/big) and word width (32/64).

The structure of the *kernel interface page* is described on page 2. The page is a microkernel object. It is directly mapped through the microkernel into each address space upon address-space creation. It is not mapped by a pager, can not be mapped or granted to another address space and can not be unmapped. The creator of a new address space can specify the address where the kernel interface page has to be mapped. This address will remain constant through the lifetime of that address space.

Any thread can determine the address of the kernel interface page through this system call. Since the system call may be slow it is highly recommended to store the address in a static variable for further use.

It is also possible to use a unique address for the kernel interface page in all address spaces of a (sub)system. Then, the kernel interface page can be accessed by fixed absolute addresses without using the current system call.

Besides other things, the page describes the current API, ABI, and microkernel version so that a server or an application can find out whether and how it can run on the current microkernel. Since the kernel interface page also contains API- and ABI-specific data for most other system calls the page’s base address is typically required before any other system call can be used.

To enable version detection independently of the API and ABI, the current system call is guaranteed to work in all L4 versions. The syscall code will never change and will be the same on compatible processors. (If a processor is upward compatible to multiple incompatible processors the kernel should offer multiple syscall codes for this function.)

---

### Output Parameters

<table>
<thead>
<tr>
<th><strong>kernel interface page</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ver X.1 and above</strong></td>
</tr>
<tr>
<td>base address ( (32/64) )</td>
</tr>
<tr>
<td>Kernel interface page address, always page aligned. 0 is no valid address.</td>
</tr>
<tr>
<td><strong>Ver X.0 and below</strong></td>
</tr>
<tr>
<td>0 ( (32/64) )</td>
</tr>
<tr>
<td>Older versions (2, X.0, etc.) do not include the kernel interface page as a kernel mapped page. No address is delivered.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>API Version</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>version ( (8) )</td>
</tr>
<tr>
<td>see page 3, “Kernel Interface Page”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>API Flags</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sim (28/60) )</td>
</tr>
<tr>
<td>see page 3, “Kernel Interface Page”</td>
</tr>
</tbody>
</table>
KernelId

| id (8) | subid (8) | ~ (16) |

see page 5, “Kernel Interface Page”

---

Pagefaults

No pagefaults will happen.

---

Generic Programming Interface

System-Call Function:

```c
#include <l4/kip.h>

Void *KernelInterface (Word &ApiVersion, ApiFlags, KernelId)
```

---

Convenience Programming Interface

Derived Functions:

```c
#include <l4/kip.h>

struct MEMORYDESC { Word raw [2] }
struct PROCDESC { Word raw [4] }

Void* KernelInterface () [GetKernelInterface]
    Delivers a pointer to the kernel interface page.

Word ApiVersion ()
Word ApiFlags ()
Word KernelId ()

Void KernelGenDate (Void* KernelInterface, Word &year, month, day)
Word KernelVersion (Void* KernelInterface)
Word KernelSupplier (Void* KernelInterface)
    Delivers the API Version/API Flags/Kernel Id/kernel generation date/kernel version/kernel supplier.

Word NumProcessors (Void* KernelInterface)
Word NumMemoryDescriptors (Void* KernelInterface)
    Delivers number of processors in the system/number of memory descriptors in the kernel-interface page.

Word PageSizeMask (Void* KernelInterface)
Word PageRights (Void* KernelInterface)
    Delivers supported page sizes/page rights for the current kernel/hardware architecture.

Word ThreadIdBits (Void* KernelInterface)
Word ThreadIdSystemBase (Void* KernelInterface)```
**KernelInterface**

- **Word ThreadIdUserBase** (Void* KernelInterface)
  Delivers number of valid bits for thread numbers/lowest thread number for system threads/lowest thread number for user threads.

- **Word ReadPrecision** (Void* KernelInterface)
- **Word SchedulePrecision** (Void* KernelInterface)
  Delivers the SYSTEMCLOCK read precision/maximal jitter for wakeups (both in \(\mu s\)).

- **Word UtcbAreaSizeLog2** (Void* KernelInterface)
- **Word UtcbAlignmentLog2** (Void* KernelInterface)
- **Word UtcbSize** (Void* KernelInterface)
  Delivers required minimum size of UTCB area/alignment requirement for UTCBs/size of a single UTCB.

- **Word KipAreaSizeLog2** (Void* KernelInterface)
  Delivers size of kernel interface page area.

- **Word BootInfo** (Void* KernelInterface)
  Delivers the contents of the boot info field.

- **Char* KernelVersionString** (Void* KernelInterface)
  Delivers the kernel version string.

- **Char* Feature** (Void* KernelInterface, Word num)
  Delivers the \(num\)th kernel feature string, or a null pointer if \(num\) exceeds the number of available feature strings.

- **MemoryDesc* MemoryDesc** (Void* KernelInterface, Word num)
  Delivers the \(num\)th memory descriptor, or a null pointer if \(num\) exceeds the number of available descriptors.

- **ProcDesc* ProcDesc** (Void* KernelInterface, Word num)
  Delivers the \(num\)th processor descriptor, or a null pointer if \(num\) exceeds the number of processors of the system (see ProcessorInfo).

---

**Support Functions:**

- #include <i4/kip.h>
- **Word UndefinedMemoryType**
- **Word ConventionalMemoryType**
- **Word ReservedMemoryType**
- **Word DedicatedMemoryType**
- **Word SharedMemoryType**
- **Word BootLoaderSpecificMemoryType**
- **Word ArchitectureSpecificMemoryType**

- **bool IsVirtual** (MemoryDesc& m) [IsMemoryDescVirtual]
  Delivers true if memory descriptor specifies a virtual memory region.

- **Word Type** (MemoryDesc& m) [MemoryDescType]
- **Word Low** (MemoryDesc& m) [MemoryDescLow]
- **Word High** (MemoryDesc& m) [MemoryDescHigh]
  Delivers type \((t*16 + type)\), low limit, and high limit of memory region.
Word `ExternalFreq` (ProcDesc& p) \[\text{ProcDescExternalFreq}\]

Word `InternalFreq` (ProcDesc& p) \[\text{ProcDescInternalFreq}\]

Delivers external frequency/internal frequency of processor.
1.3 Virtual Registers

Virtual registers are implemented by the microkernel. They offer a fast interface to exchange data between the microkernel and user threads. Virtual registers are registers in the sense that they are static per-thread objects. Depending on the specific processor type, they can be mapped to hardware registers or to memory locations. Mixtures, some virtual registers to hardware registers, some to memory are also possible. The ABI for virtual-register access depends on the specific processor type and on the virtual-register type, see Appendices A.1, B.1 and C.1 for specific hardware details.

There are three classes of virtual registers:

- Thread Control Registers (TCRs), see page 16
- Message Registers (MRs), see page 44
- Buffer Registers (BRs), see page 55

Loading illegal values into virtual registers, overwriting read-only virtual registers, or accessing virtual registers of other threads in the same address space (which may be physically possible if some are mapped to memory locations) is illegal and can have undefined effects on all threads of the current address space. However, since virtual registers can not be accessed across address spaces, they are safe from the kernel’s point of view: Illegal accesses can like any other programming bug only compromise the originator’s address space.

**Remark:** In general, virtual registers can only be addressed directly, not indirectly through pointers. The generic API therefore offers no operations for indirect virtual-register access. However, processor-specific code generators might use indirect access techniques if the ABI permits it.

---

### Generic Programming Interface

```c
#include <l4/message.h>

Void StoreMR (int i, Word& w)

Void LoadMR (int i, Word w)

Delivers/sets MR i.

Void StoreMRs (int i, k, Word& [k] w)

Void LoadMRs (int i, k, Word& [k] w)

Stores/loads MR _i...i+k−1 to/from memory.

Void StoreBR (int i, Word& w)

Void LoadBR (int i, Word w)

Delivers/sets the value of BR i.

Void StoreBRs (int i, k, Word& [k])

Void LoadBRs (int i, k, Word& [k])

Stores/loads BR _i...i+k−1 to/from memory.
```
Chapter 2

Threads
2.1 ThreadId [Data Type]

Thread IDs identify threads and hardware interrupts. A thread ID can be global or local. Global thread IDs are unique through the entire system. They identify threads independently of the address space in which they are used. Local thread IDs exist per address space; the scope of a thread’s local ID is only the thread’s own address space. In different address spaces, the same local thread ID may identify different and unrelated threads.

Note that any thread has a global and a local thread ID. Both global and local thread IDs are encoded in a single word.

**Global Thread ID**

A global thread ID consists of a word, where 18 bits (32-bit processor) or 32 bits (64-bit processor) determine the thread number and 14 bits (32-bit processor) or 32 bits (64-bit processor) are available for a version number. At least one of the lowermost 6 version bits must be 1 to differentiate a global from a local thread ID.

User-thread numbers can be freely allocated within the interval \([\text{UserBase}, 2^t]\), where \(t\) denotes the upper limit of thread IDs. The thread-number interval \([\text{SystemBase}, \text{UserBase})\) is reserved for L4-internal threads. Hardware interrupts are regarded as hardware-implemented threads. Consequently, they are identified by thread IDs. Their corresponding thread numbers are within the interval \([0, \text{SystemBase})\). The values \(\text{SystemBase}, \text{UserBase}, \) and \(t\) are published in the kernel interface page (see page 4).

<table>
<thead>
<tr>
<th>global thread ID</th>
<th>thread no (18/32)</th>
<th>version(14/32) (\neq 0 \pmod{64})</th>
</tr>
</thead>
<tbody>
<tr>
<td>global interrupt ID</td>
<td>intr no (18/32)</td>
<td>1 (14/32)</td>
</tr>
</tbody>
</table>

Global thread IDs have a version field whose content can be freely set by those threads that can create and delete threads. However, the lowermost 6 bits of the version must not all be 0, i.e. \(v \mod 64 \neq 0\) must hold for every version \(v\). For hardware interrupts, the version field is always 1.

The microkernel checks version fields whenever a thread is accessed through its global thread ID. However, the semantics of the version field are not defined by the microkernel. OS personalities are free to use this field for any purpose. For example, they may use it to make thread IDs unique in time.

**Local Thread ID**

Local thread IDs identify threads within the same address space. They are identified by the 6 lowermost bits being 0.

<table>
<thead>
<tr>
<th>local thread ID</th>
<th>local id/64 (26/58)</th>
<th>000000</th>
</tr>
</thead>
</table>

**Special Thread IDs**

Special IDs exist for nilthread and two wild cards. The thread ID anythread matches with any given thread ID, including all interrupt IDs. The ID anylocalthread matches all threads that reside in the same address space.

<table>
<thead>
<tr>
<th>nilthread</th>
<th>0 (32/64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>anythread</td>
<td>(-1) (32/64)</td>
</tr>
<tr>
<td>anylocalthread</td>
<td>(-1) (26/58)</td>
</tr>
</tbody>
</table>
Generic Programming Interface

```c
#include <l4/thread.h>

struct THREADID { Word raw }
```

ThreadId nilthread
ThreadId anythread
ThreadId anylocalthread

ThreadId GlobalId (Word threadno, version)
Delivers a thread ID with indicated thread and version number.

Word Version (ThreadId t)
Word ThreadNo (ThreadId t)
Delivers version/thread number of indicated global thread ID.

Convenience Programming Interface

```c
#include <l4/thread.h>

bool == (ThreadId l, r) [IsThreadEqual]
bool != (ThreadId l, r) [IsThreadNotEqual]
Check if thread IDs match or differ. The result of comparing a local ID with a global ID will always indicate a mismatch, even if the IDs refer to the same thread.

bool SameThreads (ThreadId l, r)
{ GlobalId (l) == GlobalId (r) }
Check if thread IDs refer to the same thread. Also works if one ID is local and the other is global.

bool IsNilThread (ThreadId t)
{ t == nilthread }

bool IsLocalId (ThreadId t)
bool IsGlobalId (ThreadId t)
Check if thread ID is a local/global one.

ThreadId LocalId (ThreadId t) [LocalIdOf]
ThreadId GlobalId (ThreadId t) [GlobalIdOf]
Delivers the local/global ID of the specified local thread. Specifying a non-local thread delivers nilthread (see EXCHANGEREGISTERS, page 18).

ThreadId MyLocalId ()
ThreadId MyGlobalId ()
Delivers the local/global ID of the currently running thread (see TCRs, page 16).

ThreadId Myself ()
{ MyGlobalId () }
## 2.2 Thread Control Registers (TCRs)

TCRs are a fast mechanism to exchange relatively static control information between user thread and microkernel. TCRs are static non-transient per-thread registers.

<table>
<thead>
<tr>
<th>Register Name</th>
<th>Size</th>
<th>Access</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VirtualSender/ActualSender</td>
<td>32/64</td>
<td>R/W</td>
<td>see IPC</td>
</tr>
<tr>
<td>IntendedReceiver</td>
<td>32/64</td>
<td>R-only</td>
<td>see IPC</td>
</tr>
<tr>
<td>XferTimeouts</td>
<td>32/64</td>
<td>R/W</td>
<td>see IPC</td>
</tr>
<tr>
<td>ErrorCode</td>
<td>32/64</td>
<td>R-only</td>
<td>see IPC</td>
</tr>
<tr>
<td>Preempt Flags</td>
<td>8</td>
<td>R/W</td>
<td>see Scheduling</td>
</tr>
<tr>
<td>Cop Flags</td>
<td>8</td>
<td>W-only</td>
<td>see Miscellaneous</td>
</tr>
<tr>
<td>ExceptionHandler</td>
<td>32/64</td>
<td>R/W</td>
<td>see Miscellaneous</td>
</tr>
<tr>
<td>Pager</td>
<td>32/64</td>
<td>R/W</td>
<td>see Protocols</td>
</tr>
<tr>
<td>UserDefinedHandle</td>
<td>32/64</td>
<td>R/W</td>
<td>see Threads</td>
</tr>
<tr>
<td>ProcessorNo</td>
<td>32/64</td>
<td>R-only</td>
<td>see Miscellaneous</td>
</tr>
<tr>
<td>MyLocalId</td>
<td>32/64</td>
<td>R-only</td>
<td>see Threads, IPC</td>
</tr>
<tr>
<td>MyGlobalId</td>
<td>32/64</td>
<td>R-only</td>
<td>see Threads, IPC</td>
</tr>
</tbody>
</table>

---

**MyGlobalId**  
Global ID of the thread.

**MyLocalId**  
Local ID of the thread.

**ProcessorNo**  
The processor number on which the thread currently executes.

**UserDefinedHandle**  
This field can be freely set and read by user threads. It can, e.g., be used for storing a thread number, a pointer to an additional user thread control block, etc.
**Generic Programming Interface**

The listed generic functions permit user code to access TCRs independently of the processor-specific TCR model. All functions are user-level functions; the microkernel is not involved.

```c
#include <4/thread.h>

ThreadId MyLocalId ()
ThreadId MyGlobalId ()
    Delivers the local/global ID of the currently running thread (see TCRs, page 16).

ThreadId Myself ()
    { MyGlobalId () }

int ProcessorNo ()
    Delivers the processor number the current thread is running on. Delivered value is a valid index into the processor description array (see Kernel Interface Page, page 4).

Word UserDefinedHandle ()
Void Set_UserDefinedHandle (Word NewValue)
    Delivers/sets the user defined handle of the currently running thread.

ThreadId Pager ()
Void Set_Pager (ThreadId NewPager)
    Delivers/sets the pager for the currently running thread.

ThreadId ExceptionHandler ()
Void Set_ExceptionHandler (ThreadId NewHandler)
    Delivers/sets the exception handler for the currently running thread.

Void Set_CopFlag (Word n)
Void Clr_CopFlag (Word n)
    Sets/clears coprocessor flag \( c_n \).

Word ErrorCode ()
    Delivers the error code of the last IPC (see IPC, page 60).

Word XferTimeouts ()
Void Set_XferTimeouts (Word NewValue)
    Delivers/sets the transfer timeouts for the currently running thread (see IPC, page 59).

ThreadId IntendedReceiver ()
    Delivers the intended receiver of last received IPC (see IPC, page 60).

ThreadId ActualSender ()
    Delivers the actual sender of the last propagated IPC (see IPC, page 59).

Void Set_VirtualSender (ThreadId t)
    Sets the virtual sender for the next deceiving IPC (see IPC, page 59).
```

Code generators of IDL and other compilers are not restricted to the generic interface. They can use any processor-specific methods and optimizations to access TCRs.
2.3 EXCHANGEREGISTERS [Systemcall]

<table>
<thead>
<tr>
<th>Input</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dest</td>
<td>Thread ID of the addressed thread. This may be a local or a global ID. However, the addressed thread must reside in the current address space. Using a local thread ID might be substantially faster in some implementations.</td>
</tr>
<tr>
<td>control</td>
<td>The s-flag refers to the SP register, i to the IP, f to the FLAGS, a to the UserDefinedHandle TCR, p to the pager TCR, and h to the H-flag. If a flag is set to 1, the register/state is overwritten by the corresponding input parameter. Otherwise, the corresponding input parameter is ignored and the register/state is not modified.</td>
</tr>
<tr>
<td>h p u f i s</td>
<td>Controls whether the addressed thread’s ongoing IPC operation should be canceled/aborted through the system call or not.</td>
</tr>
<tr>
<td>S R</td>
<td>An IPC operation of the addressed thread that is currently waiting to send a message will be canceled. An IPC operation that is currently sending a message will be aborted.</td>
</tr>
<tr>
<td>S = 0</td>
<td>An IPC operation of the addressed thread that is currently waiting to send a message will be canceled. An IPC operation that is currently sending a message will be aborted.</td>
</tr>
<tr>
<td>S = 1</td>
<td>An IPC operation of the addressed thread that is currently waiting to receive a message will be canceled. An IPC operation that is currently receiving a message will be aborted.</td>
</tr>
<tr>
<td>R = 0</td>
<td>An IPC operation of the addressed thread that is currently receiving a message will be canceled. An IPC operation that is currently sending a message will be aborted.</td>
</tr>
<tr>
<td>R = 1</td>
<td>No effect if the thread was not halted. Otherwise, thread execution is resumed.</td>
</tr>
</tbody>
</table>

Exchanges or reads a thread’s FLAGS, SP, and IP hardware registers as well as pager and UserDefinedHandle TCRs. Furthermore, thread execution can be suspended or resumed. The destination thread must exist and must reside in the invoker’s address space.

Any IP, SP, or FLAGS modification changes the corresponding user-level registers of the addressed thread. In general, ongoing kernel activities are not influenced. However, a currently active IPC operation can be canceled or aborted. For details see the SR-bit specification below.

Modifications of the pager TCR and the UserDefinedHandle TCR become immediately effective, whether the destination thread executes in user mode or in kernel mode.
User-level thread execution is halted. Note that ongoing IPCs and other kernel operations are not affected by $H$. (See $SR$ for also aborting active IPC.)

### $SP$

The current user-level stack pointer is set to $SP$ if $s = 1$. Ignored for $s = 0$. 

### $IP$

The current user-level instruction pointer is set to $IP$ if $i = 1$. If the thread was stopped/inactive before and resumed by $h = 1 \land H = 0$, it is implicitly started. Ignored for $i = 0$. 

### $FLAGS$

Sets the user-level processor flags of the thread if $f = 1$. Ignored for $f = 0$. The semantics of the $FLAGS$ word depends on the processor type.

### $UserDefinedHandle$

Sets the thread’s $UserDefinedHandle$ TCR if $u = 1$. Ignored for $u = 0$. 

### $pager$

Sets the thread’s $pager$ TCR if $p = 1$. Ignored for $p = 0$. 

### Output Parameters

**result** ≠ nilthread, input parameter dest was a local thread ID

global thread ID of the addressed thread. EXCHANGEREGISTERS succeeded.

**result** ≠ nilthread, input parameter dest was a global thread ID

local thread ID of the addressed thread. EXCHANGEREGISTERS succeeded.

**result** = nilthread

Operation failed. $dest$ might denote a thread that resides in a different address space or might denote no existing thread.

### $control$

<table>
<thead>
<tr>
<th>control</th>
<th>$0$.($29/61$)</th>
<th>$S$</th>
<th>$R$</th>
<th>$H$</th>
</tr>
</thead>
</table>

Reports whether the addressed thread was halted ($H = 1$) or not ($H = 0$) when EXCHANGEREGISTERS was invoked. Note that this output $control$ bit is independent of the input parameter $control$.

### $SR$

Reports whether the addressed thread was within an IPC operation when EXCHANGEREGISTERS was invoked. A value of 0 reports that the addressed thread was not within a send phase ($S = 0$) or not within a receive phase ($R = 0$), respectively. Note that these output $control$ bits are independent of the input parameter $control$.

### $R = 1$

Operation was executed while the addressed thread was within the receive phase of an IPC operation. If the input control word had $R = 1$ the IPC operation was canceled or aborted.

### $S = 1$

Operation was executed while the addressed thread was within the send phase of an IPC operation. If the input control word had $S = 1$ the IPC operation was canceled or aborted.

### $SP$

Old user-level stack pointer of the thread.

### $IP$

Old user-level instruction pointer of the thread.
**FLAGS**

Old user-level flags of the thread. The semantics of this word is processor specific.

**UserDefinedHandle**

Old content of thread’s `UserDefinedHandle` TCR.

**pager**

Old content of thread’s `pager` TCR.

**Pagefaults**

No pagefaults will happen.

---

**Generic Programming Interface**

System-Call Function:

```c
#include <4/thread.h>

ThreadId ExchangeRegisters (ThreadId dest, Word control, sp, ip, flags, UserDefinedHandle, ThreadId pager, Word& old control, old sp, old ip, old flags, old UserDefinedHandle, ThreadId& old pager)
```

**Convenience Programming Interface**

Derived Functions:

```c
#include <4/thread.h>

ThreadId GlobalId (ThreadId t) [GlobalIdOf]
{
    if (IsLocalId (t)) ExchangeRegisters (t,0,–,. . . ) else t
}

Delivers global ID of specified local thread. Specifying a non-local thread delivers `nilthread`.

ThreadId LocalId (ThreadId t) [LocalIdOf]
{
    if (IsGlobalId (t)) ExchangeRegisters (t,0,–,. . . ) else t
}

Delivers local ID of specified local thread. Specifying a non-local thread delivers `nilthread`.

Word UserDefinedHandle (ThreadId t) [UserDefinedHandleOf]

Void Set_UserDefinedHandle (ThreadId t, Word handle) [Set_UserDefinedHandleOf]

Delivers/sets the user defined handle of specified local thread. Result of specifying a non-local thread is undefined.

ThreadId Pager (ThreadId t) [PagerOf]

Void Set_Pager (ThreadId t, p) [Set_PagerOf]

Delivers/sets the pager for specified local thread. Result of specifying a non-local thread is undefined.

Void Start (ThreadId t) [Start_SpIp]

Void Start (ThreadId t, Word sp, ip) [Start_SpIp]

Void Start (ThreadId t, Word sp, ip, flags) [Start_SpIpFlags]

Resume execution of specified local thread (if halted). Optionally modify stack pointer, instruction pointer, and processor flags according to function parameters. Result of specifying a non-local thread is undefined.
ThreadState Stop (ThreadId t)  
ThreadState Stop (ThreadId t, Word& sp, ip, flags)  [Stop_SpIpFlags]

Halt execution of specified local thread and return its current thread state. Do not abort any ongoing IPC operation. Optionally return thread’s stack pointer, instruction pointer, and processor flags in output parameters. Result of specifying a non-local thread is undefined.

ThreadState AbortReceive_and_stop (ThreadId t)  
ThreadState AbortReceive_and_stop (ThreadId t, Word& sp, ip, flags)  [AbortReceive_and_stop_SpIpFlags]

As stop (), except any ongoing IPC receive operation is immediately aborted.

ThreadState AbortSend_and_stop (ThreadId t)  
ThreadState AbortSend_and_stop (ThreadId t, Word& sp, ip, flags)  [AbortSend_and_stop_SpIpFlags]

As stop (), except any ongoing IPC send operation is immediately aborted.

ThreadState AbortIpc_and_stop (ThreadId t)  
ThreadState AbortIpc_and_stop (ThreadId t, Word& sp, ip, flags)  [AbortIpc_and_stop_SpIpFlags]

As stop (), except any ongoing IPC send or receive operations are immediately aborted.

Support Functions:

#include <l4/thread.h>

struct THREADSTATE { Word raw }

bool ThreadWasHalted (ThreadState s)  
bool ThreadWasSending (ThreadState s)  
bool ThreadWasReceiving (ThreadState s)  
bool ThreadWasIpcing (ThreadState s)

Query the thread state returned from one of the stop () functions.
## 2.4 ThreadControl

**[Privileged Systemcall]**

<table>
<thead>
<tr>
<th>Type</th>
<th>dest</th>
<th>Word</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreadId</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ThreadId</td>
<td>SpaceSpecifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ThreadId</td>
<td>scheduler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ThreadId</td>
<td>pager</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Void*</td>
<td>UtcbLocation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A privileged thread, e.g., the root server, can delete and create threads through this function. It can also modify the global thread ID (version field only) of an existing thread.

Threads can be created as **active** or **inactive** threads. Inactive threads do not execute but can be activated by active threads that execute in the same address space.

An actively created thread starts immediately by executing a short receive operation from its pager. (An active thread must have a pager.) The actively started thread expects a start message (MsgTag and two untyped words) from its pager. Once it receives the start message, it takes the value of MR$_1$ as its new IP, the value of MR$_2$ as its new SP, and then starts execution at user level with the received IP and SP.

Interrupt threads are treated as normal threads. They are active at system startup and can **not** be deleted or migrated into a different address space (i.e., SpaceSpecifier must be equal to the interrupt thread ID). When an interrupt occurs the interrupt thread sends an IPC to its pager and waits for an empty end-of-interrupt acknowledgment message (MR$_0=0$).

Interrupt threads never raise pagefaults. To deactivate interrupt message delivery the pager is set to the interrupt thread’s own ID.

### Input Parameters

**dest**

Addressed thread. **Must be a global thread ID.** Only the thread number is effectively used to address the thread. If a thread with the specified thread number exists, its version bits are overwritten by the version bits of dest id and any ongoing IPC operations are aborted. Otherwise, the specified version bits are used for thread creations, i.e., a thread creation generates a thread with ID dest.

**SpaceSpecifier ≠ nilthread, dest not existing**

**Creation.** The space specifier specifies in which address space the thread will reside. Since address space do not have own IDs, a thread ID is used as SpaceSpecifier. Its meaning is: the new thread should execute in the same address space as the thread SpaceSpecifier.

The first thread in a new address space is created with SpaceSpecifier = dest. This operation implicitly creates a new empty address space. Note that the space creation must be completed by a SPACECONTROL operation before the thread(s) can execute.

**SpaceSpecifier ≠ nilthread, dest exists**

**Modification Only.** The addressed thread dest is neither deleted nor created. Modifications can change the version bits of the thread ID, the associated scheduler, the pager, or the associated address space, i.e., migrate the thread to a new address space.

**SpaceSpecifier = nilthread, dest exists**

**Deletion.** The addressed thread dest is deleted. Deleting the last thread of an address space implicitly also deletes the address space.

**scheduler ≠ nilthread**

Defines the scheduler thread that is permitted to schedule the addressed thread. Note that the scheduler thread must exist when the addressed thread starts executing.
The current scheduler association is not modified. This variant is illegal for a creating THREADCONTROL operation.

The pager of dest is set to the specified thread. If dest was inactive before, it is activated.

The current pager association is not modified. If used with a creating THREADCONTROL operation, dest is created as an inactive thread.

The start address of the UTCB of the thread is set to UtcbLocation. The UTCB must fit entirely into the UTCB area of the configured address space, and must be properly aligned according to the UtcbInfo field of the kernel interface page. It is the application’s responsibility to ensure that UTCBs of multiple threads do not overlap. Changing the UtcbLocation of an already active thread is an illegal operation.

The UTCB location is not modified.

Permits to calculate the appropriate page size of the UTCB area fpage and specifies the size and alignment of UTCBs. Note that the size restricts the total number of threads that can reside in an address space.

\[
\begin{array}{c|c|c|c}
\sim (10/42) & s (6) & a (6) & m (10) \\
\end{array}
\]

- \( s \) The minimal area size for an address space’s UTCB area is \( 2^s \). The size of the UTCB area limits the total number of threads \( k \) to \( 2^s mk \leq 2^a \).
- \( m \) UTCB size multiplier.
- \( a \) The UTCB location must be aligned to \( 2^a \). The total size required for one UTCB is \( 2^s m \).

The result is 1 if the operation succeeded completely, 0 otherwise. (The operation may fail because the addressed thread does not exist, UtcbLocation is not valid, and/or the invoker is not sufficiently privileged for the requested operation.)

No pagefaults will happen.

#include <l4/thread.h>

Word ThreadControl (ThreadId dest, SpaceSpecifier, Scheduler, Pager, Void* UtcbLocation)
Convenience Programming Interface

Derived Functions:

```c
#include <l4/thread.h>

Word AssociateInterrupt (ThreadId InterruptThread, InterruptHandler)
    { ThreadControl (InterruptThread, InterruptThread, nilthread, InterruptHandler, -1) }

Associate a handler thread with the specified interrupt source.

Word DeassociateInterrupt (ThreadId InterruptThread)
    { ThreadControl (InterruptThread, InterruptThread, nilthread, InterruptThread, -1) }

Remove association between the specified interrupt source and any potential handler thread.
```
Chapter 3

Scheduling
3.1 Clock [Data Type]

On both 32-bit and 64-bit processors, the system clock is represented as a 64-bit unsigned counter. The clock measures time in 1 µs units, independent of the processor frequency. Although the clock base is undefined, it is guaranteed that the counter will not overflow for at least 1,000 years.

---

**Generic Programming Interface**

```c
#include <l4/schedule.h>

struct Clock { Word64 raw }
```

---

**Convenience Programming Interface**

```c
#include <l4/schedule.h>

Clock + (Clock l, int r) [ClockAddUsec]
Clock + (Clock l, Word64 r)
Clock − (Clock l, int r) [ClockSubUsec]
Clock − (Clock l, Word64 r)

Adds/subtracts a number of µs to/from a clock value. Delivers new clock value. Does not modify the old clock value.

bool < (Clock l, r) [IsClockEarlier]
bool > (Clock l, r) [IsClockLater]
bool <= (Clock l, r)
bool >= (Clock l, r)
bool == (Clock l, r) [IsClockEqual]
bool != (Clock l, r) [IsClockNotEqual]

Compares two clock values.
```
3.2 SYSTEMCLOCK [Systemcall]

\[\text{→ Clock clock}\]

Delivers the current system clock. Typically, the operation does not enter kernel mode.

---

Pagefaults

No pagefaults will happen.

---

Generic Programming Interface

System-Call Function:

\[
\text{#include <l4/schedule.h>}
\]

\[
\text{Clock SystemClock ()}
\]
## 3.3 Time [Data Type]

Time values are used to specify send/receive timeouts for IPC operations (see page 58) and time quanta for scheduling (see page 31). The unit for time periods as well as for time points is 1 µs. Clock ticks thus happen every µs.

Relative time values specify a time period. Time periods are encoded as un-normalized 16-bit floating-point numbers. (Note that for easier handling the mantissa can have leading 0-bits.) The shortest non-zero time period that can be specified is 1 µs, the longest finite period slightly exceeds 610 hours. Two special periods frequently used for timeouts are 0 and ∞, a never ending period. The values 0 and ∞ have special encodings.

<table>
<thead>
<tr>
<th>time period:</th>
<th>c (e)</th>
<th>m (10)</th>
<th>(2^e m , \mu s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 (16)</td>
<td>(0_{(16)})</td>
<td>∞</td>
</tr>
<tr>
<td>0</td>
<td>1 (5)</td>
<td>0 (10)</td>
<td>0</td>
</tr>
</tbody>
</table>

Absolute time values specify a point in time. They are only valid for a limited period, at maximum 67 seconds.

<table>
<thead>
<tr>
<th>time point:</th>
<th>c (e)</th>
<th>m (10)</th>
<th>(2^e \cdot \left( m + {\text{Clock}}_{e+10} \cdot 2^{10} \right))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (4)</td>
<td>(0_{(10)})</td>
<td>if ( {\text{Clock}}_{e+10} = c )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2^e \cdot \left( m + {\text{Clock}}<em>{e+10} \cdot 2^{10} + 2^{10} \right)) if ( {\text{Clock}}</em>{e+10} \neq c )</td>
</tr>
</tbody>
</table>

Absolute time values are thus the more precise the nearer in the future they are. Absolute time values with maximal precision become invalid just after the clock has reached the specified point in time. The validity interval can be expanded, but only by reducing the precision. In general, a time-point value \((c, m, e)\) that is constructed when the current clock value is \(C_0\) is valid from \(C_0\) up to

\[C_0 + \left( 2^{10} - 1 \right) \cdot 2^e\]

Therefore, a time-point value that should remain valid for 10 ms can have a precision of 10 µs whereas a value that should remain valid for an entire second can only have a precision of 1 ms. In general, a precision of 0.1% of the required validity interval can be achieved.

### Generic Programming Interface

```c
#include "l4/schedule.h"

struct TIME { Word16 raw }
```

**Time** *Never*

**Time** *ZeroTime*

**Time** *TimePeriod* *(Word64 microseconds)*
Time timePoint (Clock at)

Convenience Programming Interface

#include <l4/schedule.h>

Time + (Time l, Word r) [TimeAddUsec]
Time += (Time l, Word r) [TimeAddUsecTo]
Time − (Time l, Word r) [TimeSubUsec]
Time −= (Time l, Word r) [TimeSubUsecFrom]

Adds/subtracts a number of microseconds to/from a time value.

Time + (Time l, r) [TimeAdd]
Time += (Time l, r) [TimeAddTo]
Time − (Time l, r) [TimeSub]
Time −= (Time l, r) [TimeSubFrom]

Adds/subtracts a time period to/from a time value. The result of adding/subtracting a time point is undefined.

bool > (Time l, r) [IsTimeLonger]
bool >= (Time l, r) [IsTimeLonger]
bool < (Time l, r) [IsTimeShorter]
bool <= (Time l, r) [IsTimeShorter]
bool == (Time l, r) [IsTimeEqual]
bool != (Time l, r) [IsTimeNotEqual]

Compares two time values. The result of comparing a time period with a time point, or vice versa, is undefined.
The invoking thread releases the processor (non-preemptively) so that another ready thread can be processed.

**Input Parameter**

- **dest = nilthread**: Processing switches to an undefined ready thread which is selected by the scheduler. (It might be the invoking thread.) Since this is “ordinary” scheduling, the thread gets a new timeslice.

- **dest ≠ nilthread**: If `dest` is ready, processing switches to this thread. In this “extraordinary” scheduling, the invoking thread donates its remaining timeslice to the destination thread. (This one gets the donation in addition to its ordinarily scheduled timeslices, if any.) If the destination thread is not ready or resides on a different processor, the system call operates as described for `dest = nilthread`.

**Pagefaults**

No pagefaults will happen.

**Generic Programming Interface**

**System-Call Function:**

```c
#include <l4/schedule.h>

Void ThreadSwitch (ThreadId dest)
```

**Convenience Programming Interface**

**Derived Functions:**

```c
#include <l4/schedule.h>

Void Yield ()

{ ThreadSwitch (nilthread) }

Switch processing to a thread selected by the scheduler.
```
3.5 **Schedule** [Systemcall]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dest</code></td>
<td>Destination thread ID. The destination thread must be existent (but can be inactive) and the current thread must be defined as the destination thread’s scheduler (see thread control). Otherwise, the destination thread is not affected. All further input parameters have no effect if the supplied value is $-1$, ensuring that the corresponding internal thread variable is not modified. The following description always refers to values $\neq -1$.</td>
</tr>
<tr>
<td><code>time control</code></td>
<td>Defines the total quantum for the thread. Exhaustion of the total quantum results in an RPC to the thread’s scheduler (i.e., the current thread). (Re)writing the total quantum re-initializes the quantum, independent of the already consumed total quantum. The total quantum is specified as a time period (see page 28). Absolute time values are illegal. A total quantum of $\infty$ can be specified.</td>
</tr>
<tr>
<td><code>ts len</code></td>
<td>New timeslice length for the destination thread. The timeslice length is specified as a time period (see page 28). Absolute time values and the value 0 are illegal. A timeslice length of $\infty$, however, can be specified. In that case, the thread never experiences a preemption due to exhausted time slice. The specified value is always rounded up to the nearest possible timeslice length. In particular, a time period of 1 $\mu$s results in the shortest possible timeslice. Writing the timeslice length initializes the current quantum with the new length. After the quantum is exhausted, the thread is preempted while the quantum is reloaded with <code>ts len</code> for the next timeslice.</td>
</tr>
<tr>
<td><code>total quantum</code></td>
<td>Defines the total quantum for the thread. Exhaustion of the total quantum results in an RPC to the thread’s scheduler (i.e., the current thread). (Re)writing the total quantum re-initializes the quantum, independent of the already consumed total quantum. The total quantum is specified as a time period (see page 28). Absolute time values are illegal. A total quantum of $\infty$ can be specified.</td>
</tr>
<tr>
<td><code>prio</code></td>
<td>New priority for destination thread. Must be less than or equal to current thread’s priority.</td>
</tr>
<tr>
<td><code>preemption control</code></td>
<td>Preemptions by threads that run on a priority lower or equal to this <code>sensitive prio</code> will, (a) if the <code>delay-preemption</code> flag is set, be delayed until the thread executes a <code>thread switch</code> (<code>nithread</code>) system call; and (b) if the <code>signal-preemption</code> flag is set, raise a preemption fault to the exception handler. No preemption delays or signaling will occur if preempted by a thread having a higher priority than <code>sensitive prio</code>, regardless of the state of the <code>delay-preemption</code> and <code>signal-preemption</code> flags.</td>
</tr>
</tbody>
</table>

The system call can be used by schedulers to define the priority, timeslice length, and other scheduling parameters of threads. Furthermore, it delivers thread states. The system call is only effective if the calling thread is defined as the destination thread’s scheduler (see thread control, page 22).
**maximum delay**
The maximum time in µs a pending preemption can be delayed in the destination thread. The value 0 effectively disables preemption delay.

**processor control**

| 0 (16/48) | processor number (16) |

**processor number**
Specifies the processor number to which the thread should be migrated. The processor number must be valid, i.e., smaller than the total number of processors (see kernel interface page at page 3). Otherwise, the parameter is ignored. The first processor number is denoted as 0.

**Output Parameters**

| result | ~ (24/56) | tstate (8) |

**tstate =**
Thread state:

| 0 | Error. The operation failed completely. The invoker has specified invalid parameters or is not sufficiently privileged for the requested operation. |
| 1 | Dead. The thread is unable to execute or does not exist. |
| 2 | Inactive. The thread is inactive/stopped. |
| 3 | Running. The thread is ready to execute at user-level. |
| 4 | Pending send. A user-invoked IPC send operation currently waits for the destination (recipient) to become ready to receive. |
| 5 | Sending. A user-invoked IPC send operation currently transfers an outgoing message. |
| 6 | Waiting to receive. A user-invoked IPC receive operation currently waits for an incoming message. |
| 7 | Receiving. A user-invoked IPC receive operation currently receives an incoming message. |

**time control**

| rem ts (16) | rem total (16) |

**rem ts**
Remainder of the current timeslice.

**rem total**
Remaining total quantum of the thread.

**Pagefaults**

No pagefaults will happen.
Generic Programming Interface

System-Call Function:

```
#include <l4/schedule.h>

Word Schedule (ThreadId dest, Word TimeControl, ProcessorControl, prio, PreemptionControl, Word& old_TimeControl)
```

Convenience Programming Interface

Derived Functions:

```
#include <l4/schedule.h>

Word Set_Priority (ThreadId dest, Word prio)
    { Schedule (dest, -1, -1, prio, -1) }

Word Set_ProcessorNo (ThreadId dest, Word ProcessorNo)
    { Schedule (dest, -1, ProcessorNo, -1, -1) }

Word Timeslice (ThreadId dest, Time & ts, Time & tq)
    Delivers the remaining timeslice and total quantum of the given thread.
    { Schedule (dest, ts * 2^16 + tq, -1, -1, -1) }

Word Set_Timeslice (ThreadId dest, Time ts, Time tq)
    { Schedule (dest, ts * 2^16 + tq, -1, -1, -1) }

Word Set_PreemptionDelay (ThreadId dest, Word sensitivePrio, Word maxDelay)
    { Schedule (dest, -1, -1, -1, SensitivePrio * 2^16 + MaxDelay) }
```
### 3.6 Preempt Flags \([TCR]\)

The *preemption flags* TCR controls asynchronous preemptions (timeslice exhausted or activation of a higher-priority thread including device interrupts).

![Preempt Flags](image)

The *ds*-flags are used to control the microkernel. User threads can set/reset them. The *I*-flag signals an event to the user. It is set by the microkernel and typically read/reset by the user.

- **s = 0**  
  Asynchronous preemptions are not signaled to the exception handler.

- **s = 1**  
  Asynchronous preemptions are signaled as preemption faults to the exception handler. If \(d = 0\) this happens immediately. Otherwise, it is delayed until the thread continues execution after the preemption.

- **d = 0**  
  All asynchronous preemptions happen immediately. If they are signaled as preemption faults \((s = 1)\), this happens after the preemption took place, i.e., when the thread gets reactivated.

- **d = 1**  
  Asynchronous preemptions are delayed if the priority of the preemptor is lower or equal than the *sensitive priority* for the current thread. (The sensitive priority is set by the scheduler, see page 31.) A delayed preemption does not interrupt the current thread immediately but is postponed until the current thread invokes a systemcall *thread switch* (*nilthread*). However, a pending preemption must not be delayed for longer than the *maximum delay* that was set by the thread’s scheduler. Such a preemption-delay overflow resets the *d*-bit and is signaled to the exception handler.

- **I = 0**  
  No asynchronous preemption is pending.

- **I = 1**  
  An asynchronous preemption is currently pending, i.e., the thread should as soon as possible reset the *d*-flag and invoke *thread switch*. Invoking *thread switch* re-enables the *maximum delay* for the next delayed asynchronous preemption. Invoking *thread switch* is not required if no asynchronous preemption is pending \((I = 0)\) after the user thread has reset the *d*-flag.

---

#### Generic Programming Interface

```c
#include "l4/schedule.h"

bool EnablePreemptionFaultException ()
    Sets/resets the *s*-flag and delivers the old *s*-flag value (true = set).

bool DisablePreemptionFaultException ()

bool DisablePreemption ()
    Sets/resets the *d*-flag and delivers the old *d*-flag value (true = set).

bool EnablePreemption ()

bool PreemptionPending ()
    Resets the *I*-flag and delivers the old *I*-flag value (true = set).
```
Chapter 4

Address Spaces and Mapping
4.1 Fpage  [Data Type]

Fpages (Flexpages) are regions of the virtual address space. An fpage consists of all pages mapped actually in this region sans kernel mapped objects, i.e., kernel interface page and UTCBs. Fpages have a size of at least 1 K. For specific processors, the minimal fpage size may be larger; e.g., a Pentium processor offers a minimal page size of 4 K while the Alpha processor offers smallest pages of 8 K. Fpages smaller than the minimal page size are treated as nilpages. The kernel interface page (see page 3) specifies which page sizes are supported by the hardware/kernel. An fpage of size $2^s$ has a $2^s$-aligned base address $b$, i.e., $b \equiv 0 \pmod{2^s}$, where $s \geq 10$ for all architectures.

Mapped fpages are considered inseparable objects. That is, if an fpage is mapped, the mapper cannot later partially unmap the mapped page; the whole fpage must be unmapped in a single operation. The mappee can, however, separate the fpage and map fpages (objects) of smaller size. Partially unmapping an fpage might or might not work on some systems. The kernel will give no indication as to whether such an operation succeeded or not.

$$\text{fpage } (b, 2^s)$$

<table>
<thead>
<tr>
<th>$b/2^{10}$</th>
<th>$s$</th>
<th>$0$</th>
<th>$r$</th>
<th>$w$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(22/54)</td>
<td>(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Special fpage denoters describe the complete user address space and the nilpage, an fpage which has no base address and a size of 0:

- **complete**
  $$\begin{array}{c|c|c|c|c|c|}
  0 & s = 1 & 0 & r & w & x \\
  (22/54) & (6) & & & & \\
  \end{array}$$

- **nilpage**
  $$\begin{array}{c|c|c|c|c|c|}
  0 & s = 0 & 0 & r & w & x \\
  (32/64) & (6) & & & & \\
  \end{array}$$

**Access Rights**

- **$rwx$**

  The $rwx$ bits define the accessibility of the fpage:

  $r$: readable  
  $w$: writable  
  $x$: executable

  A bit set to one permits the corresponding access to the newly-mapped/granted page provided that the mapper itself possesses that access right. If the mapper does not have the access right itself or if the bit is set to zero the mapped/granted page will not get the corresponding access right.

  Note that processor architectures may impose restrictions on the access-right combinations. However, read-only (including execute), $rwx = 101$, and read/write/execute, $rwx = 111$, should be valid for any processor architecture. The kernel interface page (see page 3) specifies which access rights are supported in the processor architecture.

---

**Generic Programming Interface**

```c
#include <l4/space.h>

struct FPAGE { Word raw }
```

- **Word Readable**
- **Word Writable**
Word *Executable*
Word *FullyAccessible*
Word *ReadExecOnly*
Word *NoAccess*

**Fpage Nilpage**

**Fpage CompleteAddressSpace**

```c
bool IsNilFpage (Fpage f)
{
    f == Nilpage
}
```

**Fpage Fpage** (Word BaseAddress, int FpageSize ≥ 1K)

**Fpage FpageLog2** (Word BaseAddress, int Log2FpageSize < 64)

Delivers an fpage with the specified location and size.

**Word Address** (Fpage f)

**Word Size** (Fpage f)

**Word SizeLog2** (Fpage f)

Delivers address/size of specified fpage.

**Word Rights** (Fpage f)

**Void SetRights** (Fpage f, Word AccessRights)

Delivers/sets the access rights for the specified fpage.

**Fpage +** (Fpage f, Word AccessRights)  \[FpageAddRights\]

**Fpage +=** (Fpage f, Word AccessRights)  \[FpageAddRightsTo\]

**Fpage −** (Fpage f, Word AccessRights)  \[FpageRemoveRights\]

**Fpage −=** (Fpage f, Word AccessRights)  \[FpageRemoveRightsFrom\]

Adds/removes specified access rights from fpage. Delivers new fpage value.
The specified fpages (located in MR 0...) are unmapped. Fpages are mapped as part of the IPC operation (see page 57).

### Input Parameters

**control**

| 0 (25/57) | k (6) |

- **k** Specifies the highest MRₖ that holds an fpage to be unmapped. The number of fpages is thus \( k + 1 \).
- **f = 0** The fpages are unmapped recursively in all address spaces in which threads of the current address space have mapped them before. However, the fpages remain unchanged in the current address space.
- **f = 1** The fpages are unmapped like in the \( f = 0 \) case and, in addition, also in the current address space.

**FpageList MR₀...k** Fpages to be processed.

**Fpage MRᵢ**

Fpage to be unmapped. (The term *unmapped* is used even if effectively no access right is removed.) A nilpage specifies a no-op.

**0rwx** Any access bit set to 1 revokes the corresponding access right. A 0-bit specifies that the corresponding access right should not be affected. Typical examples:

- **=0111** Complete unmap of the fpage.
- **=0010** Partial unmap, revoke writability only. As a result, the fpage is set to read-only.
- **=0000** No unmap. This case is particularly useful if only *dirty* and *accessed* bits should be read and reset without changing the mapping.

### Output Parameters

**FpageList MR₀...k** The accessed status bits in the fpages are updated.
The status bits *Read*, *Written*, and *eXecuted* of all pages processed by the unmap operation are reset and the bitwise OR-ed old values of all the processed pages are delivered in MR \( 0...k \).

For processors that do not differentiate between read access and execute access, the \( R \) and \( X \) bits are unified: either both are set or both are reset. Resetting status bits is not a recursive operation. However, the status bit values for pages within the current space will also reflect accesses performed on recursive mappings.

### \( R = 0 \)

No part of the fpage has been *Referenced* after the last unmap operation (or after the initial map operation). This includes all recursively mapped pages.

*Remark:* The meaning of *referenced* slightly differs from *read*. Not being referenced means that not only no read access but that also no write and execute access occurred.

### \( R = 1 \)

At least one page of the specified fpage (including all recursive mappings) has been referenced after the last unmap operation (or after the initial map operation). All in-kernel \( R \) bits are reset.

*Remark:* The meaning of *referenced* slightly differs from *read*. Write accesses and execute accesses also set the \( R \) bit.

### \( W = 0 \)

No part of the fpage has been written after the last unmap operation (or after the initial map operation), i.e., the fpage is *clean*. This includes all recursively mapped pages.

### \( W = 1 \)

At least one page of the specified fpage (including all recursive mappings) has been written after the last unmap operation (or after the initial map operation), i.e., the fpage is *dirty*. All in-kernel dirty bits are reset.

### \( X = 0 \)

No part of the fpage has been *eXecuted* after the last unmap operation (or after the initial map operation). This includes all recursively mapped pages.

### \( X = 1 \)

At least one page of the specified fpage (including all recursive mappings) has been executed after the last unmap operation (or after the initial map operation). All in-kernel \( X \) bits are reset.

*Remark:* For processors that do not differentiate between read and execute accesses, the \( X \) bit is set to 1 iff \( R = 1 \).

---

### Pagefaults

No pagefaults will happen.

---

### Generic Programming Interface

**System-Call Function:**

```c
#include <l4/space.h>

Void Unmap (Word control)
```

---

### Convenience Programming Interface

**Derived Functions:**

```c
#include <l4/space.h>

Void Unmap (Fpage f) \[UnmapFpage]\n    { LoadMR (0, f); Unmap (0); StoreMR (0, f); }

Void Unmap (Word n, Fpage & [n] fpages) \[UnmapFpages]\n    { LoadMRs (0, n, fpages); Unmap (n - 1); StoreMRs (0, n, fpages); }
```

Recursively unmaps the specified fpage(s) from all address spaces except the current one.
Void Flush (Fpage f)
    { LoadMR (0, f); Unmap (64); StoreMR (0, f); }

Void Flush (Word n, Fpage \& [n] fpages)
    { LoadMRs (0, n, fpages); Unmap (64 + n - 1); StoreMRs (0, n, fpages); }

Recursively unmaps the specified fpage(s) from all address spaces, including the current one.

Fpage GetStatus (Fpage f)
    { LoadMR (0, f - FullyAccessible); Unmap (0); StoreMR (0, f); f }

Resets and delivers the status bits of the specified fpage.

bool WasReferenced (Fpage f)
bool WasWritten (Fpage f)
bool WasExecuted (Fpage f)

Checks the status bits of specified fpage. The specified fpage must be the output of an Unmap (), Flush (), or GetStatus () function.
4.3 **SPACECONTROL**  

[Privileged Systemcall]

<table>
<thead>
<tr>
<th>ThreadId</th>
<th>SpaceSpecifier</th>
<th>control</th>
<th>Fpage</th>
<th>KernelInterfacePageArea</th>
<th>Fpage</th>
<th>UtcbArea</th>
<th>Redirector</th>
</tr>
</thead>
</table>

A privileged thread, e.g., the root server, can configure an address spaces through this function.

### Input Parameters

**SpaceSpecifier**  
Since address spaces do not have ids, a thread ID is used as SpaceSpecifier. It specifies the address space in which the thread resides. The SpaceSpecifier thread must exist although it may be inactive or not yet started. In particular, the thread may reside in an empty address space that is not yet completely created.

**KernelInterfacePageArea**  
Specifies the fpage where the kernel should map the kernel interface page. The supplied fpage must have a size specified in the KipAreaInfo field of the kernel interface page and must fit entirely into the user-accessible part of the address space. Address 0 of the kernel interface page is mapped to the fpage’s base address. The value is ignored if there is at least one active thread in the address space.

**KipAreaInfo** [KernelInterfacePage Field]  
Permits calculation of the appropriate page size of the KernelInterface area fpage.

\[
\sim \left( \frac{26}{58} \right) \quad s \quad (6)
\]

\( s \)  
The size of the kernel interface page area is \( 2^s \).

**UtcbArea**  
Specifies the fpage where the kernel should map the UTBs of all threads executing in the address space. The fpage must fit entirely into the user-accessible part of an address space. The fpage size has to be at least the smallest supported hardware-page size. In fact, the size of the UTCB area restricts the maximum number of threads that can be created in the address space. See the kernel interface page for the space and alignment that is required for UTBs. The value is ignored if there is at least one active thread in the address space.

**UtcbInfo** [KernelInterfacePage Field]  
Permits to calculate the appropriate page size of the UTCB area fpage and specifies the size and alignment of UTBs. Note that the size restricts the total number of threads that can reside in an address space.

\[
\sim \left( \frac{10}{42} \right) \quad s \quad (6) \quad a \quad (6) \quad m \quad (10)
\]

\( s \)  
The minimal area size for an address space’s UTCB area is \( 2^s \). The size of the UTCB area limits the total number of threads \( k \) to \( 2^m mk \leq 2^s \).

\( m \)  
UTCB size multiplier.
The UTCB location must be aligned to $2^n$. The total size required for one UTCB is $2^n m$.

**Redirector** = nthread

The current redirector setting for the specified space is not modified.

**Redirector** = anythread

All threads within the specified space are allowed to communicate with any thread in the system.

**Redirector** ≠ anythread, ≠ nthread

All threads within the specified address space are only allowed to send an IPC to a local thread or to a thread in the same address space as the specified redirector. All other send operations will be deflected to the redirector, the redirected bit (see page 60) in the received message will be set, and the IntendedReceiver TCR will indicate the intended receiver of the message.

**control**

The control field is architecture specific (see Appendix A.5). It is undefined for some architectures, but should for reasons of upward compatibility be set to zero.

### Output Parameters

**result**

The result is 1 if the operation succeeded completely, 0 otherwise. (The operation may fail because the addressed thread does not exist, and/or the invoker is not sufficiently privileged for the requested operation, and/or the address space contained more than one thread).

**control**

Delivers the space control value that was effective for the thread when the operation was invoked. The value is architecture specific.

### Pagefaults

No pagefaults will happen.

### Generic Programming Interface

**System-Call Function:**

```c
#include <l4/space.h>

Word SpaceControl (ThreadId SpaceSpecifier, Word control, Fpage KernelInterfacePageArea, UtcbArea, ThreadId Redirector, Word& old_Control)
```
5.1 Messages And Message Registers (MRs)  [Virtual Registers]

Messages can be sent and received through the IPC system call (see page 57). Basically, the sender writes a message into the sender’s message registers (MRs) and the receiver reads it from the receiver’s MRs. Each thread has 64 MRs, MR0...63. A message can use some or all MRs to transfer untyped words; it can include memory strings and pages which are also specified using MRs.

MRs are virtual registers (see page 11), but they are more transient than TCRs. MRs are read-once registers: once an MR has been read, its value is undefined until the MR is written again. The send phase of an IPC implicitly reads all MRs; the receive phase writes the received message into MRs.

The read-once property permits to implement MRs not only by special registers or memory locations, but also by general registers. Writing to such an MR has to block the corresponding general register for code-generator use; reading the MR can release it. Typically, code generated by an IDL compiler will load MRs just before an IPC system call and store them to user variables just afterwards.

Messages

A message consists of up to 3 sections: the mandatory message tag, followed by an optional untyped-words section, followed by an optional typed-items section. The message tag is always held in MR0. It contains message control information and the message label which can be freely set by the user. The kernel associates no semantics with it. Often, the message label is used to encode a request key or to define the method that should be invoked by the message.

```
MsgTag [MR0]  

| label (16/48) | flags (4) | t (6) | u (6) |
```

- **u** Number of untyped words following word 0. MR1...u hold the untyped words. u = 0 denotes a message without untyped words.

- **t** Number of typed-item words following the untyped words or the message tag if no untyped words are present. The typed items use MRu+1...u+t. A message without typed items has t = 0.

- **flags** Message flags, see IPC system call, page 57.

- **label** Freely available, often used to specify the request type or invoked method.

**untyped words [MR1...u]**

The optional untyped-words section holds arbitrary data that is untyped from the kernel’s point of view. The data is simply copied to the receiver. The kernel associates no semantics with it.

**typed items [MRu+1...u+t]**
The optional typed-items section is a sequence of items such as string items (page 52), map items (page 49), and grant items (page 51). Typed message items have their type encoded in the lowermost 4 bits of their first word:

- $0hhC$: StringItem see page 52
- $100C$: MapItem see page 49
- $101C$: GrantItem see page 51
- $110C$: Reserved
- $111C$: Reserved

The $C$ bit signals whether the typed item is followed by another typed item ($C = 1$) or is the last one of the typed-item section ($C = 0$). The typed items must exactly fit into $M_{r+1...r+t}$.

Note that $C$ and $t$ redundantly describe the message. This is by intention. The $C$ bit allows efficient message parsing, whereas $t + u$ can be used to store all MRs of a message to memory without parsing the complete message. Upon message sending, the $C$ bits are completely ignored. The kernel will, however, ensure that the MRs on the receiver side will have the $C$ bits set properly.

### Example Messages

**struct (label, Word [2] $w$)**

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>MR 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word $w_2$ (32/64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Word $w_1$ (32/64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>label (16/48) flags $t = 0$ $u = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**struct (label, MapItem $m$)**

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>MR 1,2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MapItem $m$</td>
<td>1 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>label (16/48) flags $t = 2$ $u = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**struct (label, Word $w$, StringItem $s_1$, $s_2$)**

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>MR 4,5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>StringItem $s_2$</td>
<td>0 h h 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>StringItem $s_1$</td>
<td>0 h h 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Word $w$ (32/64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>label (16/48) flags $t = 4$ $u = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**struct (label, Word [3] $w$, MapItem $m$, GrantItem $g$, StringItem $s$)**
Generic Programming Interface

The listed generic functions permit user code to access message registers independently of the processor-specific MR model. All functions are user-level functions; the microkernel is not involved.

**MsgTag**

```c
#include <l4/ipc.h>

struct MsgTag { Word raw }

MsgTag **NilTag**

A message tag with no untyped or typed words, no label, and no flags.

```
bool == (MsgTag l, r) [IsMsgTagEqual]
bool ! = (MsgTag l, r) [IsMsgTagNotEqual]
```

Compares all field values of two message tags.

**Word Label** (Msg Tag t)

**Word UntypedWords** (Msg Tag t)

**Word TypedWords** (Msg Tag t)

Delivers the message label, number of untyped words, and number of typed words, respectively.

```
MsgTag + (MsgTag t, Word label) [MsgTagAddLabel]
MsgTag += (MsgTag t, Word label) [MsgTagAddLabelTo]
```

Adds a label to a message tag. Old label information is overwritten by the new label.

**MsgTag MsgTag ()**

**Void Set_MsgTag** (MsgTag t)

Delivers/sets MR₀.
Convenience Programming Interface

IDL-code generators are not restricted to the generic interface for accessing MRs. Instead, they can use processor-specific methods and thus generate heavily optimized code for MR access.

However, such processor-specific MR operations are not generally defined and should be used exclusively by processor-specific IDL code generators. All other programs must use the operations defined in this generic interface.

```c
# include <l4/ ipc.h>

struct Msg { Word raw[64] }

Void Put (Msg& msg, Word l, int u, Word & ut, int t, {MapItem, GrantItem, StringItem} & Items) [MsgPut]
Loads the specified parameters into the memory object msg. The parameters u and t respectively indicate number of untyped words and number of typed words (i.e., the total size of all typed items). It is assumed that the msg object is large enough to contain all items.

Void Get (Msg & msg, Word & ut, {MapItem, GrantItem, StringItem} & Items) [MsgGet]
Stores the msg object into the specified parameters. Type consistency between the message in the memory object and the specified parameter list is not checked.

MsgTag MsgTag (Msg& msg) [MsgMsgTag]
Void Set_MsgTag (Msg & msg, MsgTag t) [Set_MsgTag]
Delivers/sets the message tag of the msg object.

Word Label (Msg& msg) [MsgLabel]
Void SetLabel (Msg & msg, Word label) [Set_MsgLabel]
Delivers/sets the label of the msg object.

Void Load (Msg& msg) [MsgLoad]
Loads message registers MR_0... from the msg object.

Void Store (MsgTag t, Msg& msg) [MsgStore]
Stores the message tag t and the current message beginning with MR_1 to the memory object msg. The number of message registers to be stored is derived from t.

Void Clear (Msg & msg) [MsgClear]
Empties the msg object (i.e., clears the message tag).

Void Append (Msg & msg, Word w) [MsgAppendWord]
Void Append (Msg & msg, MapItem m) [MsgAppendMapItem]
Void Append (Msg & msg, GrantItem g) [MsgAppendGrantItem]
Void Append (Msg & msg, StringItem s) [MsgAppendSimpleStringItem]
Void Append (Msg & msg, StringItem & s) [MsgAppendStringItem]
Appends an untyped or a typed item to the msg object. Compound strings must always be passed in by reference. A compound string passed by value will be treated as a simple string (see page 52). It is assumed that there is enough memory in the msg object to contain the new item.

Void Put (Msg& msg, Word u, Word w) [PutWord]
Puts an untyped word at untyped word position u (first untyped word has position 0) in the msg object. It is assumed that the object contains at least u + 1 untyped words.

Void Put (Msg & msg, Word t, MapItem m) [MsgPutMapItem]
Void **Put** (Msg& msg, Word t, GrantItem g)  
[MsgPutGrantItem]

Void **Put** (Msg& msg, Word t, StringItem s)  
[MsgPutSimpleStringItem]

Void **Put** (Msg& msg, Word t, StringItem& s)  
[MsgPutStringItem]

Puts a typed item into the msg object, starting at typed word position t (first typed word has position 0). Compound strings must always be passed in by reference. A compound string passed by value will be treated as a simple string (see page 52). It is assumed that the object has enough typed words to contain the new item.

Word **Get** (Msg& msg, Word u)  
[MsgWord]

Void **Get** (Msg& msg, Word u, Word& w)  
[MsgGetWord]

Delivers the untyped words at position u. It is assumed that the object contains at least u + 1 untyped words.

Word **Get** (Msg& msg, Word t, MapItem& m)  
[MsgGetMapItem]

Word **Get** (Msg& msg, Word t, GrantItem& g)  
[MsgGetGrantItem]

Word **Get** (Msg& msg, Word t, StringItem& s)  
[MsgGetStringItem]

Delivers the typed item starting at typed word position t. It is assumed that the requested item is of the right size and type. Returns the size (in words) of the delivered item.

---

**Low-Level MR Access**

#include <l4/ipc.h>

Void **StoreMR** (int i, Word& w)

Void **LoadMR** (int i, Word w)

Delivers/sets MR i.

Void **StoreMRs** (int i, k, Word& [k] w)

Void **LoadMRs** (int i, k, Word& [k] w)

Stores/loads MR i...i+k−1 to/from memory.
5.2 MapItem [Data Type]

An fpage (see page 36) or IO fpage that should be mapped is sent to the mappee as part of a message. A map operation is a no-op within the same address space. The fpage is specified by a two-word descriptor:

<table>
<thead>
<tr>
<th>snd fpage (28/60)</th>
<th>0 r w x</th>
<th>MR \textsubscript{j+1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>snd base / 1024 \textsubscript{22/54}</td>
<td>0 (6)</td>
<td>1 00 C \textsubscript{T}</td>
</tr>
</tbody>
</table>

**access rights rwx** The effective access rights for the newly mapped page are calculated by bitwise AND-ing the access rights specified in the snd fpage and the access rights that the mapper itself has on that fpage. As such, the mapper can restrict the effective access rights but not widen them.

**snd base**

The send base specifies the semantics of the map operation if the size of the snd fpage is larger or smaller than the window in which the receiver is willing to accept a mapping (see page 55). If the size of the snd fpage, \(2^s\), is larger than the receive window, \(2^r\), the send base indicates which region of the snd fpage that is transmitted. More precisely:

\[
\text{send region} = fpage (addr_r + 2^r k, 2^r), \text{ for some } k \geq 0 :
\]

\[
addr_r + 2^r k \leq addr_r + (snd\ base \ mod\ 2^r) < addr_r + 2^r k + 2^r
\]

and where \(addr_r\) is the base address of the snd fpage. If the size of the snd fpage, \(2^s\), is smaller than the receive window, \(2^r\), the send base indicates where in the receive window the snd fpage is mapped. More precisely:

\[
\text{receive region} = fpage (addr_r + 2^s k, 2^s), \text{ for some } k \geq 0 :
\]

\[
addr_r + 2^s k \leq addr_r + (snd\ base \ mod\ 2^s) < addr_r + 2^s k + 2^s
\]

and where \(addr_r\) is the base address of the receive window.

Pages already mapped in the mappee’s address space that would conflict with new mappings are implicitly unmapped before new pages are mapped. For performance reasons extension of access rights is possible without prior unmapping, iff the very same mapping already exists. This is the case, when

- the mapper maps from the same address space as the existing mapping; and
- the mapper maps from the same virtual source address as the existing mapping; and
- the mapper maps to the same virtual destination address as the existing mapping; and
- the object (physical address) is the same as the existing mapping.

Access rights can not be revoked by mapping. The access rights of the resulting mapping are a bitwise OR of the existing and the new mapping’s access rights. Access rights are not extended recursively.

---

**Generic Programming Interface**

```
#include <l4/ipc.h>

struct MAPITEM { Word raw[2] }
```

\textit{MapItem MapItem} (Fpage f, Word SndBase)

Delivers a map item with the specified fpage and send base.
bool MapItem (MapItem m)  [IsMapItem]
    Delivers true if map item is valid. Otherwise delivers false.

Fpage SndFpage (MapItem m)  [MapItemSndFpage]

Word SndBase (MapItem m)  [MapItemSndBase]
    Delivers fpage/send base of map item.
5.3 GrantItem [Data Type]

An fpage (see page 36) or IO fpage that should be granted is sent to the mappee as part of a message. It is specified by a two-word descriptor:

<table>
<thead>
<tr>
<th>snd fpage (28/60)</th>
<th>0 r w x</th>
<th>MR_{i+1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>snd base / 1024  (22/54)</td>
<td>0 (6)</td>
<td>1 0 1 C’</td>
</tr>
</tbody>
</table>

**access rights rwx** The effective access rights for the granted page are calculated by bitwise anding the access rights specified in the snd fpage and the access rights that the mapper itself has on that fpage. As such, the granter can restrict the effective access rights but not widen them.

**snd base**

The send base specifies the semantics of the map operation if the size of the snd fpage is larger or smaller than the window in which the receiver is willing to accept a mapping (see page 55). If the size of the snd fpage, 2^s, is larger than the receive window, 2^r, the send base indicates which region of the snd fpage that is transmitted. More precisely:

\[
send \text{ region} = fpage (addr_s + 2^r k, 2^r), \text{ for some } k \geq 0:
\]

\[
addr_s + 2^r k \leq addr_s + (snd \ base \ mod \ 2^r) < addr_s + 2^r k + 2^r
\]

and where \(addr_s\) is the base address of the snd fpage. If the size of the snd fpage, 2^s, is smaller than the receive window, 2^r, the send base indicates where in the receive window the snd fpage is mapped. More precisely:

\[
receive \text{ region} = fpage (addr_r + 2^s k, 2^s), \text{ for some } k \geq 0:
\]

\[
addr_r + 2^s k \leq addr_r + (snd \ base \ mod \ 2^s) < addr_r + 2^s k + 2^s
\]

and where \(addr_r\) is the base address of the receive window.

Pages already mapped in the grantee’s address space that would conflict with new mappings are implicitly unmapped before new pages are mapped.

---

**Generic Programming Interface**

```c
#include <l4/ipc.h>

struct GrantItem { Word raw [2] }

GrantItem GrantItem (Fpage f, Word SndBase)
Delivers a grant item with the specified fpage and send base.

bool GrantItem (GrantItem g) [IsGrantItem]
Delivers true if grant item is valid. Otherwise delivers false.

Fpage SndFpage (GrantItem g) [GrantItemSndFpage]
Delivers fpage/send base of grant item.

Word SndBase (GrantItem g) [GrantItemSndBase]
```

5.4 StringItem [Data Type]

A string item specifies a sequence of bytes in user space. No alignment is required, the maximal string size is 4 MB. In send messages, such a string is copied to the receiver buffer when transferring the message. String items are also used to specify receive buffers in buffer registers on the receiver’s side.

**Simple String**

A simple string is a contiguous sequence of bytes.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>string ptr (32/64)</td>
<td>The start address of the string to be sent or the start address of the buffer for receiving a string (no alignment restrictions). However, the string/buffer must fit entirely into the legally addressable user space.</td>
</tr>
<tr>
<td>string length (22/54)</td>
<td>The length of the string to be sent or the size of the receive buffer. In the second case, strings up to (including) this length can be received. Maximum string length is 4 M bytes, even if the according field is 54 bits wide on 64-bit processors.</td>
</tr>
<tr>
<td>hh</td>
<td>Cacheability hint. Except for hh = 00, the semantics of this parameter depends on the processor type (see Appendices A.6 and B.5).</td>
</tr>
</tbody>
</table>

hh = 00 Use the processor’s default cacheability strategy. Typically, cache lines are allocated for data read and written (assuming that the processor’s default strategy is write-back and write-allocate).

**Compound String**

A compound string is a noncontiguous string that consists of multiple contiguous substrings which can be scattered around the entire user address space. The substrings must not overlap. For send and receive IPC operations, a compound string is handled as a single logical string. When sending such a string through IPC, the substrings are transferred as if they were one contiguous string (gather). On the receiver side, a compound string buffer is treated as one logical buffer. The corresponding received string is scattered among the compound buffer’s substrings.

A compound string can be specified as a sequence of substrings where each substring has the form of a simple string except that the continuation flag c is set for all but the last substring. If j subsequent substrings have the same size, e.g., for equally sized buffers, a single length word can be used for all j substrings so that only j + 1 words instead of 2j words are required.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>length word</td>
<td>The type information 0hhC is only required for the first word of a string descriptor. The field is ignored for further length words in a compound-string descriptor.</td>
</tr>
<tr>
<td>substring length (22/54)</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>Number of subsequent string-ptr words. These string ptrs specify j substrings that have all the same substring length.</td>
</tr>
<tr>
<td>c = 0</td>
<td>Continuation flag reset. The compound string descriptor ends with the jth string ptr word following the current length word.</td>
</tr>
<tr>
<td>c = 1</td>
<td>Continuation flag set. The current length word and j string-ptr words are followed by (at least) one substring descriptor, i.e., another length word, etc.</td>
</tr>
</tbody>
</table>
Generic Programming Interface

```c
#include <l4/ipc.h>

struct STRINGITEM { Word raw [*] }

bool StringItem (StringItem& s) [IsStringItem]
Delivers true if string item is valid. Otherwise delivers false.

bool CompoundString (StringItem& s)
Delivers the c-flag value (true = set).

Word Substrings (StringItem& s)
Void* Substring (StringItem& s, Word n)
Delivers number of substrings/address of nth substring.

StringItem StringItem (int size, Void* address)
Delivers a simple string item with the specified size and location.

StringItem & += (StringItem& dest, StringItem AdditionalSubstring) [AddSubstringTo]
Append substring to the string item. It is assumed that there is enough memory in the string item
to contain the new substring.

StringItem & += (StringItem& dest, Void* AdditionalSubstringAddress) [AddSubstringAddressTo]
Append a new substring pointer to the string item. It is assumed that there is enough memory in
the string item to contain the new substring pointer.
```

Convenience Programming Interface

Support Functions:

```c
#include <l4/ipc.h>

struct CACHEALLOCATIONHINT { Word raw }

CacheAllocationHint UseDefaultCacheLineAllocation
```
bool == (CacheAllocationHint l, r)  [IsCacheAllocationHintEqual]
bool != (CacheAllocationHint l, r)  [IsCacheAllocationHintNotEqual]

Compares two cache allocation hints.

CacheAllocationHint CacheAllocationHint (StringItem s)
    Delivers the cache allocation hint of the string item.

StringItem + (StringItem s, CacheAllocationHint h)  [AddCacheAllocationHint]
StringItem += (StringItem s, CacheAllocationHint h)  [AddCacheAllocationHintTo]
    Adds a cache allocation hint to a string item. An already existing hint is overwritten.
5.5 String Buffers And Buffer Registers (BRs)  [Pseudo Registers]

For receiving messages that contain string items, the receiver has to specify appropriate string buffers. Such buffers are described by string items (see page 52). A buffer can be contiguous (simple string) or non-contiguous (compound string).

Such buffer descriptors are held in 34 per-thread Buffer Registers BR_0...33. The number of buffer registers is sufficient to specify, for example, one compound buffer of 32 equally-sized sub-buffers. Up to 16 buffers can be specified provided that not more than 34 BRs are required.

When a message is received, the first message string item is copied into the first buffer string item which starts at BR_1; the next message string item is copied to the next buffer string item, etc. The list of buffer strings is terminated by having the C bit in the item type specifier of the last string zeroed.

BRs are registers in the sense that they are per-thread objects and can only be addressed directly, not indirectly through pointers. BRs are static objects like TCRs, i.e., they keep their values until explicitly modified. BRs can be mapped to either special registers or to memory locations.

<table>
<thead>
<tr>
<th>Acceptor [BR_0]</th>
<th>RcvWindow (28/60)</th>
<th>000 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR_0 specifies which typed items are accepted when a message is received.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RcvWindow Fpage (without access bits) that specifies the address-space window in which mappings and grants are accepted. Nilpage denies any mapping or granting; CompleteAddressSpace accepts any mapping or granting.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s String Items are accepted iff s = 1.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

buffer string items [BR_1...]

contain the valid buffer string items. Ignored if s = 0 in BR_0.

**Generic Programming Interface**

The listed generic functions permit user code to access buffer registers independently of the processor-specific BR model. All functions are user-level functions; the microkernel is not involved.

**Acceptor**

```c
#include <l4/ipc.h>

struct Acceptor { Word raw }

Acceptor UntypedWordsAcceptor
Acceptor StringItemsAcceptor
Acceptor MapGrantItems (Fpage RcvWindow)

Delivers an acceptor which allows untyped words, string items, or mappings and grants.
```

```c
Acceptor + (Acceptor l, r) [AddAccept]  
Acceptor += (Acceptor l, r) [AddAcceptTo]  

Adds mappings/grants or string items to an acceptor. Adding a non-nil receive window will replace an existing window.
```

```c
Acceptor − (Acceptor l, r) [RemoveAccept]  
Acceptor −= (Acceptor l, r) [RemoveAcceptFrom]  

Removes mappings/grants or string items from an acceptor. Removing a non-nil receive window will deny all mappings or grants, regardless of the size of the receive window.
```
bool StringItems (Acceptor a) [HasStringItems]  
Checks whether string items/mappings are allowed.

Fpage RcvWindow (Acceptor a)  
Delivers the address space window where mappings and grants are accepted. Delivers nilpage if mappings or grants are not allowed.

Void Accept (Acceptor a)  
Sets BR0.

Void Accept (Acceptor a, MsgBuffer& b) [AcceptStrings]  
Sets BR0 and loads the buffer description b into BR1...

Accepte Accepted ()  
Delivers BR0.

---

Convenience Programming Interface

MsgBuffer

#include <l4/ipc.h>

struct MSGBUFFER { Word raw[32] }

Void Clear (MsgBuffer& b) [MsgBufferClear]  
Clears the message buffer (i.e., inserts a single empty string into it).

Void Append (MsgBuffer& b, StringItem s) [MsgBufferAppendSimpleRcvString]  
Appends a string buffer to the message buffer. Compound strings must always be passed in by reference. A compound string passed by value will be treated as a simple string. It is assumed that there is enough memory in the message buffer object to contain the new string buffer.

Low-Level BR Access

#include <l4/ipc.h>

Void StoreBR (int i, Word& w)  
Void LoadBR (int i, Word w)  
Delivers/sets the value of BRi.

Void StoreBRs (int i, k, Word& [k])  
Void LoadBRs (int i, k, Word& [k])  
Stores/loads BRi...i+k−1 to/from memory.

Code generators of IDL and other compilers are not restricted to the generic interface. They can use any processor-specific methods and optimizations to access BRs.
5.6 IPC [Systemcall]

IPC is the fundamental operation for inter-process communication and synchronization. It can be used for intra- and inter-address-space communication. All communication is synchronous and unbuffered: a message is transferred from the sender to the recipient if and only if the recipient has invoked a corresponding IPC operation. The sender blocks until this happens or until a period specified by the sender has elapsed without the destination becoming ready to receive.

IPC can be used to copy data as well as to map or grant fpages from the sender to the recipient. For the description of messages see page 44. A single IPC call combines an optional send phase and an optional receive phase. Which phases are included is determined by the parameters to and FromSpecifier. Transitions between send phase and receive phase are atomic.

Ipc operations are also controlled by MRs, BRs and some TCRs. RcvTimeout and SndTimeout are directly specified as system-call parameters. Each timeout can be 0, ∞ (i.e., never expire), relative or absolute. For details on timeouts see page 28.

Variants

To enable implementation-specific optimizations, there exist two variants of the IPC system call. Functionally, both variants are identical. Transparently to the user, a kernel implementation can unify both variants or implement differently optimized functions.

<table>
<thead>
<tr>
<th>IPC</th>
<th>Default IPC function. Must always be used except if all criteria for using LIPC are fulfilled.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIPC</td>
<td>IPC function that may be optimized for sending messages to local threads. Should be used whenever it is absolutely clear that in the overwhelming majority of all invocations</td>
</tr>
<tr>
<td></td>
<td>• a send phase is included; and</td>
</tr>
<tr>
<td></td>
<td>• the destination thread is specified as a local thread ID; and</td>
</tr>
<tr>
<td></td>
<td>• a receive phase is included; and</td>
</tr>
<tr>
<td></td>
<td>• the destination thread runs on the same processor; and</td>
</tr>
<tr>
<td></td>
<td>• the RcvTimeout is ∞, and</td>
</tr>
<tr>
<td></td>
<td>• the IPC includes no map/grant operations.</td>
</tr>
</tbody>
</table>

Input Parameters

<table>
<thead>
<tr>
<th>to = nilthread</th>
<th>IPC includes no send phase.</th>
</tr>
</thead>
<tbody>
<tr>
<td>to ≠ nilthread</td>
<td>Destination thread; IPC includes a send phase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FromSpecifier = nilthread</th>
<th>IPC includes no receive phase.</th>
</tr>
</thead>
</table>
**FromSpecifier = anythread**
PC includes a receive phase. Incoming messages are accepted from any thread (including hardware interrupts).

**FromSpecifier = anylocalthread**
PC includes a receive phase. Incoming messages are accepted from any thread that resides in the current address space.

**FromSpecifier ≠ nilthread, ≠ anythread, ≠ anylocalthread**
PC includes a receive phase. Incoming messages are accepted only from the specified thread. (Note that hardware interrupts can be specified.)

<table>
<thead>
<tr>
<th>Timeouts</th>
<th>SnrTimeout (16)</th>
<th>RcvTimeout (16)</th>
</tr>
</thead>
</table>

**RcvTimeout**
The receive phase waits until either a message transfer starts or the **RcvTimeout** expires. Ignored for send-only IPC operations.
For relative receive timeout values, the receive timeout starts to run after the send phase has successfully completed. If the receive timeout expires before the message transfer has been started IPC fails with “receive timeout”. A pending incoming message is received if the timeout period is 0.

**SnrTimeout**
If the send timeout expires before the message transfer could start the IPC operation fails with “send timeout”. A send timeout of 0 ensures that IPC happens only if the addressed receiver is ready to receive when the send IPC operation is invoked. Otherwise, IPC fails immediately, i.e., without blocking.

<table>
<thead>
<tr>
<th>MsgTag [MR0]</th>
</tr>
</thead>
<tbody>
<tr>
<td>label (16/48)</td>
</tr>
</tbody>
</table>

Message head of the message to be sent. Only the upper 16/48 bits are freely available. The lower 16 bits hold the **SndControl** parameter. It describes the message to be sent and contains some control bits; ignored if no send phase.

**u**
Number of untyped words following word 0. MR 1...u hold the untyped words. **u = 0** denotes a message with no untyped words.

**t**
Number of words holding typed items that follow the untyped words (or the message tag if no untyped words are present). The typed items use MR u+1 and following MRs, potentially up to MR 63. **t = 0** denotes a message without typed items.

**p=0**
Normal (unpropagated) send operation. The recipient gets the original sender’s id.

**p=1**
Propagating send operation. The **VirtualSender** TCR specifies the id of the originator thread. (i.e., the thread to send the message on behalf of). If originator thread and current sender, or current sender and receiver reside in the same address space, propagation is always permitted. Otherwise, IPC occurs unpropagated. Propagation is also allowed if one of the participating threads is a system thread, or if the current sender is a redirector for the originator thread (or there exists a chain of redirectors from the originator to the current sender).
If propagation is permitted, the receiver receives the originator’s id instead of the current sender’s id, the p bit in the receiver’s MsgTag is set, and the current sender’s id is stored in the receiver’s **ActualSender** TCR. If the originator thread is waiting (closed) for a reply from the current sender, the originator’s state is additionally modified so that it now waits for the new receiver instead of the current sender.

**label**
Freely available, often used to specify the request type or invoked method, respectively.

[MR1...u]
Untyped words to be sent. Ignored if no send phase.

[MRu+1...u+t]
Typed items to be sent. Ignored if no send phase.
Once a message transfer has been started, the time for transferring the message is roughly bounded by the minimum of sender’s and receiver’s XferTimeout. “Roughly” means that xfer timeouts are only checked when message copy raises a pagefault in the sender’s or in the receiver’s address space. Copying data and mapping/granting is assumed to take no time. A relative transfer timeout always refers to the beginning of the message transfer (actually when the first page fault is raised). Logically, at that point it is transferred into an absolute timeout which then is used as send and receive timeout for the first and all subsequent page-fault RPCs in the message transfer.

If the effective transfer timeout expires during the message transfer, IPC fails with “xfer timeout” (on both sides). Additional information specifies whether the page fault was in the receiver’s or in the sender’s address space and which part of the message was already transferred. Each thread has two transfer timeouts. One for the send phase and one for the receive phase.

BR₀ specifies which typed items are accepted when a message is received.

Fpage (without access bits) that specifies the address-space window in which mappings and grants are accepted. Nilpage denies any mapping or granting; CompleteAddressSpace accepts any mapping or granting.

StringItems are accepted iff \( s = 1 \).

contain the valid buffer string items. Ignored if \( s = 0 \) in BR₀.

Thread ID of the sender from which the IPC was received. Thread IDs are delivered as local thread IDs iff they identify a thread executing in the same address space as the current thread. It does not matter whether the sender specified the destination as local or global id. Only defined for IPC operations that include a receive phase.

If the IPC operation included a receive phase, MR₀ contains the message tag of the received message. The upper 16/48 bits contain the user-specified label. The lower bits describe the received message, contain the error indicator, and the cross-processor IPC indicator.

MR₀ is defined even if the IPC operation did not include a receive phase. In the send-only case, MR₀ returns the error indicator.

Number of untyped words following word 0. \( u = 0 \) means no untyped words. For IPC operations without receive phase, \( u = 0 \) is delivered.

Number of received words that hold typed items. \( t = 0 \) means no typed items. For IPC operations without receive phase, \( t = 0 \) is delivered.

Propagated IPC. If reset \( p = 0 \) the IPC was not propagated. If set \( p = 1 \) the IPC was propagated and the FromSpecifier indicates the originator thread’s id. The ActualSender specifies the id of the thread which performed the propagation.
Redirected IPC. If reset \((r = 0)\) the IPC was not a redirected one. If set \((r = 1)\) the IPC was redirected to the current thread, and the \textit{IntendedReceiver} TCR specifies the id of the thread supposed to receive the message.

Cross-processor IPC. If reset \((X = 0)\) the received IPC came from a thread running on the same processor as the receiver. If set \((X = 1)\) the received IPC was cross-processor. For IPC operations without receive phase, \(X = 0\) is delivered.

Error indicator. If reset \((E = 0)\) the IPC operation terminated successfully. If set \((E = 1)\) IPC failed. If the send phase was successful but a receive timeout occurred afterwards, or if a message could only be partially transferred, the entire IPC fails. The error code and additional information can be retrieved from the ErrorCode TCR. The fields \textit{label}, \textit{t}, and \textit{u} are valid if the error code signals a partially received message.

Label of the received message. For IPC operations without receive phase, the label is 0.

Untyped words that have been received. Undefined if no receive phase.

Typed items that have been received. Undefined if no receive phase.

Only defined if the error indicator \(E\) in MR0 is set. IPC failed, i.e., was not correctly completed. The \(x\) field depends on the error code, see below. The \(p\) field specifies whether the error occurred during send or receive phase. If the error occurred during the receive phase the send phase (if any) was completed successfully before. If the error occurred during the send phase, the receive phase (if any) was skipped.

\(p\) Specifies whether the error occurred during the send phase \((p = 0)\) or the receive phase \((p = 1)\).

Error happened before a partner thread was involved in the message transfer. Therefore, the error is signaled only to the thread that invoked the failing IPC operation.

\(e = 1\) Timeout. \(\textit{From}\) is undefined in this case.

\(e = 2\) Non-existing partner. If the error occurred in the send phase, \textit{to} does not exist. (\textit{Anythread} as a destination is illegal and will also raise this error.) If the error occurred in the receive phase, \textit{FromSpecifier} does not exist. (\textit{FromSpecifier} = \textit{anythread} is legal, and thus will never raise this error.)

\(e = 3\) Canceled by another thread (system call \textit{exchange registers}).

A partner thread is already involved in the IPC operation, and the error is therefore signaled to both threads.

The message transfer has been started and could not be completed. The \(\textit{offset}\) identifies exactly the number of bytes that have been been transferred successfully so far through string items.

\(e = 4\) Message Overflow. A message overflow can occur (1) if a receiving buffer string is too short, (2) if not enough buffer string items are present, and (4) if a map/grant of an fpage fails because the system has not enough page-table space available. The \(\textit{offset}\) in conjunction with the received MRs permits sender and receiver to exactly determine the reason.

\(e = 5\) Xfer timeout during page fault in the invoker’s address space.
Pagefaults

Three different types of pagefault can occur during ipc: pre-send, post-receive, and xfer pagefaults. Only xfer pagefaults are critical from a security point of view. Fortunately, messages without strings will never raise xfer pagefaults and need thus no special pagefault provisions:

**Pre-send pagefaults**

happen in the sender’s context before the message transfer has really started. The destination thread is not involved; in particular, it is not locked. Therefore, the destination thread might receive another message or time out while the sender’s pre-send pagefault is handled. Send and transfer timeouts do not control pre-send pagefaults. Pre-send pagefaults are uncritical from a security point of view, since only the sender’s own pager is involved and only the sender could suffer from its potential misbehavior.

**Post-receive pagefaults**

happen in the receiver’s context after the message has been transferred. The sender thread is no longer involved, especially, it is no longer locked. Consequently, post-receive pagefault are not subject to send and transfer timeouts. Like pre-send pagefaults, post-receive pagefaults are also uncritical from a security perspective since only the receiver and its pager are involved.

**Xfer pagefaults**

happen while the message is being transferred and both sender and receiver are involved. Therefore, xfer pagefaults are critical from a security perspective: If such a pagefault occurs in the receiver’s space, the sender may be starved by a malicious receiver pager. An xfer pagefault in the sender’s space and a malicious sender pager may starve the receiver. As such, xfer pagefaults are controlled by the minimum of sender’s and receiver’s xfer timeouts.

However, xfer pagefaults can only happen when transferring strings. Send messages without strings or receive buffers without receive string buffers are guaranteed not to raise xfer pagefaults.

Generic Programming Interface

System-Call Function:

```c
#include <l4/ipc.h>

MsgTag Ipc (ThreadId to, FromSpecifier, Word Timeouts, ThreadId& from)
MsgTag Lipc (ThreadId to, FromSpecifier, Word Timeouts, ThreadId& from)
```

Note that message registers have read-once semantics and that returning the message tag implies reading MR0. The contents of the message tag is therefore lost if the application does not implicitly store the return value of IPC or LIPC.

Convenience Programming Interface

Derived Functions:

```c
#include <l4/ipc.h>

MsgTag Call (ThreadId to)
    { Call (to, never, never) }
```
**supported functions**

```c
#include <l4/ipc.h>

bool IpcSucceeded (MsgTag t)
bool IpcFailed (MsgTag t)

Delivers the state of the error indicator (the E bit of MR0).

bool IpcPropagated (MsgTag t)
bool IpcRedirected (MsgTag t)
bool IpcXcpu (MsgTag t)

Checks if the IPC was propagated/redirected/cross cpu.

Word ErrorCode ()
ThreadId IntendedReceiver ()
```
ThreadId ActualSender ()
      Delivers the error code/intended receiver TCR/actual sender.

Void SetPropagation (Msg& Tag t)
      Sets the propagation bit.

Void SetVirtualSender (ThreadId t)
      Sets the virtual sender TCR.

Word Timeouts (Time SndTimeout, RcvTimeout)
      Delivers a word containing both timeout values.
Chapter 6

Miscellaneous
6.1 ExceptionHandler  [TCR]

An exception handler thread can be installed to receive exception IPCs.

ExceptionHandler

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#include <l4/thread.h>

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.

\#ifndef

ThreadId ExceptionHandler ()

\#endif

Void SetExceptionHandler (ThreadId new)

Delivers/sets the exception handler TCR.
6.2 Cop Flags \([\text{TCR}]\)

The coprocessor flags TCR helps the kernel to optimize thread switching for some hardware architectures.

**Cop Flags**

\[
\begin{array}{c}
1 & \ldots & 0 \\
\end{array}
\]

By resetting a \(c_i\)-bit to 0, a thread tells the system that it no longer needs coprocessor \(i\). If the kernel finds \(c_i = 0\), it concludes that registers and state of coprocessor \(i\) do not have to be saved. However, the kernel ensures that the coprocessor can not be used as a covert channel between different address spaces.

Once a thread has reset bit \(c_i\), it must set \(c_i\) to 1 before it issues the next operation on coprocessor \(i\). Otherwise, coprocessor registers and state might be arbitrarily modified while using it.

Note that the \(c_i\)-bits are *write-only*. Reading them results in an undefined value. Upon thread creation, all \(c_i\)-bits are set to 1.

---

**Generic Programming Interface**

```c
#include <l4/thread.h>

Void Set_CopFlag (Word n)
Void Clr_CopFlag (Word n)

Sets/clears coprocessor flag \(c_n\).
```
6.3 ProcessorControl

Control the internal frequency, external frequency, or voltage for a system processor.

### Input Parameters

- **ProcessorNo**
  - Specifies the processor to control. Number must be a valid index into the processor descriptor array (see Kernel Interface Page, page 4).

  All further input parameters have no effect if the supplied value is \(-1\), ensuring that the corresponding value is not modified. The following description always refers to values \(\neq -1\).

- **InternalFrequency**
  - Sets internal frequency for processor to the given value (in kHz).

- **ExternalFrequency**
  - Sets external frequency for processor to the given value (in kHz).

- **voltage**
  - Sets voltage for processor to the given value (in mV). A value of 0 shuts down the processor.

### Output Parameters

- **result**
  - The result is 1 if the operation succeeded completely, 0 otherwise. (The operation may fail because the specified processor number is invalid or the invoker is not sufficiently privileged for the requested operation.)

Note that the active internal and external frequency of all processors are available to all threads via the kernel interface page.

### Pagefaults

No pagefaults will happen.

### Generic Programming Interface

**System-Call Function:**

```
#include <l4/misc.h>
```
Word ProcessorControl (Word ProcessorNo, control, InternalFrequency, ExternalFrequency, voltage)
6.4 MEMORY CONTROL [Privileged Systemcall]

Word control → Void
Word attribute0
Word attribute1
Word attribute2
Word attribute3

Set the page attributes of the fpages (MR $0...k$) to the attribute specified with the fpage.

Input Parameters

- **control**
  - 0 (26/58)
  - \(k\) (6)

\(k\) Specifies the highest MR $k$ that holds an fpage to set the attributes. The number of fpages is thus $k + 1$.

- **attribute\(_i\)**
  - Specifies the attribute to associate with an fpage. The semantics of the attribute\(_i\) values are hardware specific, except for the value 0 which specifies default semantics.

- **FpageList MR \(_0...k\)**
  - Fpages to be processed.

- **Fpage MR \(_i\)**
  - Fpage to change the attributes. A nilpage specifies a no-op.

\(a\) selects attribute\(_a\) to be set as the fpages memory attributes.

Pagefaults

No pagefaults will happen.

Generic Programming Interface

**System-Call Function:**

```c
#include <l4/misc.h>

Void MemoryControl (Word control, Word& attributes[4])
```

Word DefaultMemory
Convenience Programming Interface

Derived Functions:

```c
#include <l4/misc.h>

Void Set_PageAttribute (Fpage f, Word attribute)
    { LoadMR (0, f); MemoryControl (0, &attribute); }

    { LoadMRs (0, n, fpages); MemoryControl (n - 1, attributes); }
```
Chapter 7

Protocols
7.1 Thread Start Protocol

Newly created active threads start immediately by receiving a message from its pager. The received message contains the initial instruction-pointer and stack-pointer for the thread.

<table>
<thead>
<tr>
<th>From Pager</th>
<th>Initial SP (32/64)</th>
<th>Initial IP (32/64)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (16/48)</td>
<td>0 (4)</td>
</tr>
<tr>
<td>MR 0</td>
<td></td>
<td>t = 0 (6)</td>
</tr>
<tr>
<td>MR 1</td>
<td></td>
<td>u = 2 (6)</td>
</tr>
<tr>
<td>MR 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.2 Interrupt Protocol

Interruption Protocol:

Interruptions are delivered as an IPC call to the interrupt handler thread (i.e., the pager of the interrupt thread). The interrupt is disabled until the interrupt handler sends a re-enable message.

**From Interrupt Thread**

| -1 (12/44) | 0 (4) | 0 (4) | t = 0 (6) | u = 0 (6) |

**To Interrupt Thread**

| 0 (16/48) | 0 (4) | t = 0 (6) | u = 0 (6) |
### 7.3 Pagefault Protocol

A thread generating a pagefault will cause the kernel to transparently generate a pagefault IPC to the faulting thread’s pager. The behavior of the faulting thread is undefined if the pager does not exactly follow this protocol.

#### To Pager

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>faulting user-level IP</td>
<td>MR₂</td>
</tr>
<tr>
<td>fault address</td>
<td>MR₁</td>
</tr>
<tr>
<td></td>
<td>MR₀</td>
</tr>
</tbody>
</table>

The `rwx` bits specify the fault reason:

- `r` read fault
- `w` write fault
- `x` execute fault

A bit set to one reports the type of the attempted access. On processors that do not differentiate between read and execute accesses, `x` is never set. Read and execute accesses will both be reported by the `r` bit.

#### Acceptor [BR₀]

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MR₁, MR₂</td>
</tr>
<tr>
<td></td>
<td>MR₀</td>
</tr>
</tbody>
</table>

The acceptor covers the complete user address space. The kernel accepts mappings or grants into this region on behalf of the faulting thread. The received message is discarded.

#### From Pager

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MapItem / GrantItem</td>
<td>MR₁, MR₂</td>
</tr>
<tr>
<td></td>
<td>MR₀</td>
</tr>
</tbody>
</table>

The `rw` bits specify the fault reason:
7.4 Preemption Protocol

From Preempted Thread

<table>
<thead>
<tr>
<th></th>
<th>Clock ( /2^{32/64} ) ( (32/64) )</th>
<th>MR (_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clock ( \text{mod}2^{32/64} ) ( (32/64) )</td>
<td>MR (_1)</td>
</tr>
<tr>
<td>(-3) ( (12/44) )</td>
<td>0 ( (4) )</td>
<td>0 ( (4) )</td>
</tr>
</tbody>
</table>

The preemption message contains the system clock when the thread was preempted. The preemption message is sent with relative timeout 0. If the message cannot be delivered (e.g., due to timeouts) the message is dropped.
7.5 Exception Protocol

The exception IPC contains a label, the faulting instruction pointer, and additional architecture specific exception words. The reply from the exception handler contains a label, an instruction pointer where the faulting thread is resumed, and an optional number of additional architecture specific words.

Note that the stack pointer is not explicitly specified to allow architecture specific optimizations.

### To Exception Handler

<table>
<thead>
<tr>
<th>exception word (k-1) (32/64)</th>
<th>MR (k+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>exception word 0 (32/64)</td>
<td>MR 2</td>
</tr>
<tr>
<td>IP (32/64)</td>
<td>MR 1</td>
</tr>
<tr>
<td>label (12/44)</td>
<td>0 (4)</td>
</tr>
<tr>
<td></td>
<td>0 (4)</td>
</tr>
<tr>
<td></td>
<td>(t = 0)</td>
</tr>
<tr>
<td></td>
<td>(u = k)</td>
</tr>
</tbody>
</table>

\(k\)  
Number of exception words.

\(label\)  
Specifies the exception type.

\(-4\)  
System exceptions are defined for all architectures.

\(-5\)  
Architecture specific exceptions.

### From Exception Handler

<table>
<thead>
<tr>
<th>exception reply word (k-1) (32/64)</th>
<th>MR (k+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>exception reply word 0 (32/64)</td>
<td>MR 2</td>
</tr>
<tr>
<td>IP (32/64)</td>
<td>MR 1</td>
</tr>
<tr>
<td>0 (16/48)</td>
<td>0 (4)</td>
</tr>
<tr>
<td>(t = 0)</td>
<td>(u = k)</td>
</tr>
</tbody>
</table>

\(k\)  
Number of exception reply words.

\(IP\)  
Location where execution is resumed in the faulting thread.
7.6 Sigma0 RPC protocol

$\sigma_0$ is the initial address space. Although it is not part of the kernel, its basic protocol is defined with the kernel. Specific $\sigma_0$ implementations may extend this protocol.

The address space $\sigma_0$ is idempotent, i.e., all virtual addresses in this address space are identical to the corresponding physical address. Note that pages requested from $\sigma_0$ continue to be mapped idempotently if the receiver specifies its complete address space as receive fpage.

$\sigma_0$ gives pages to the kernel and to arbitrary tasks, but only once. The idea is that all pagers request the memory they need in the startup phase of the system so that afterwards $\sigma_0$ has exhausted all its memory. Further requests will then automatically be denied.

Kernel Protocol

<table>
<thead>
<tr>
<th>To $\sigma_0$</th>
<th>$\sim (32/64)$</th>
<th>MR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>requested fpage $(32/64)$</td>
<td></td>
<td>MR 1</td>
</tr>
<tr>
<td>$-6 (12/44)$</td>
<td>$0 (4)$</td>
<td>$0 (4)$</td>
</tr>
</tbody>
</table>

requested fpage

$-1 (22/54)$ 8 (6) 0 rwx

$s = 0$ Kernel requests the amount of memory recommended by $\sigma_0$ for kernel use (pagetable and other kernel-internal data).

$s \neq 0$ Kernel requests an fpage of size $2^s$. The fpage can be located at an arbitrary position but must contain ordinary memory. If a free fpage of size $2^s$ is available, it is granted to the kernel.

$rwx$ The rwx bits are ignored. $\sigma_0$ always grants fpages with maximum access rights to the kernel.

From $\sigma_0$

Kernel memory recommendation

<table>
<thead>
<tr>
<th>$0 (32/64)$</th>
<th>MR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>amount $(32/64)$</td>
<td></td>
</tr>
<tr>
<td>$0 (16/48)$</td>
<td>$0 (4)$</td>
</tr>
</tbody>
</table>

amount Amount of memory recommended for kernel use (in bytes).

Grant Response

GrantItem

<table>
<thead>
<tr>
<th>$0 (16/48)$</th>
<th>$0 (4)$</th>
<th>$t = 2 (6)$</th>
<th>$u = 0 (6)$</th>
<th>MR 1,2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MR 0</td>
</tr>
</tbody>
</table>
### Grant Reject

<table>
<thead>
<tr>
<th>nilpage (32/64)</th>
<th>MR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (28/60)</td>
<td>1 0 1 0</td>
</tr>
<tr>
<td>0 (16/48)</td>
<td>0 (4)</td>
</tr>
<tr>
<td></td>
<td>t = 2 (6)</td>
</tr>
<tr>
<td></td>
<td>u = 0 (6)</td>
</tr>
</tbody>
</table>

### User Protocol

**To σ₀**

<table>
<thead>
<tr>
<th>requested attributes (32/64)</th>
<th>MR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>requested fpage (32/64)</td>
<td>MR 1</td>
</tr>
<tr>
<td>−6 (12/44)</td>
<td>0 (4)</td>
</tr>
<tr>
<td></td>
<td>0 (4)</td>
</tr>
<tr>
<td></td>
<td>t = 0 (6)</td>
</tr>
<tr>
<td></td>
<td>u = 2 (6)</td>
</tr>
</tbody>
</table>

**requested fpage**

| b/2^10 (22/54) | s (6) | 0 r w x |

σ₀ deals with fpages of arbitrary size. A successful response from σ₀ contains an fpage of physically contiguous memory.

- **b ≠ −1** Requests the specific fpage with base address b and size $2^s$. If the fpage is neither owned by the kernel nor by a user thread (not even partially), the requested fpage is mapped to the requestor’s address space and the fpage is marked as owned by the requesting thread (i.e., fpage is not marked as being owned by the address space in which thread resides). Fpages belonging to dedicated-memory (see page 83) can be requested. If the requested fpage is already owned by the requestor only the page attributes are modified. No new mapping operations happens.

- **b = −1** Requests an fpage of size $2^s$ but with arbitrary address. If a free fpage of size $2^s$ is available, it is mapped to the requestor’s address space and marked as owned by the requesting thread (i.e., fpage is not marked as being owned by the address space in which thread resides). σ₀ is free to use the requested-attribute for choosing a best fitting page. Fpages belonging to dedicated-memory (see page 83) are not considered to be free and will not be delivered upon such anonymous requests. No new mapping operations happens.

**rwx** The rwx bits are ignored. σ₀ always maps fpages with maximum access rights to the requestor.

**requested attributes**

- $= 0$ The page is requested with default attributes.

- $\neq 0$ The page is requested with some architecture dependent attributes.

**From σ₀**

**Map Response**

<table>
<thead>
<tr>
<th>MapItem</th>
<th>MR 1,2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (16/48)</td>
<td>0 (4)</td>
</tr>
<tr>
<td>t = 2 (6)</td>
<td>u = 0 (6)</td>
</tr>
</tbody>
</table>
\[ \sigma_0 \text{ responds with a map reject message if the page is reserved (i.e., kernel space) or already mapped to a different thread, or if memory is exhausted.} \]
7.7 Generic Booting  [Protocol]

Machine-specific boot procedures are described on pages 99 ff.

After booting, L4 initializes itself. It generates the basic address space-servers $\sigma_0$, $\sigma_1$ and a root server which is intended to boot the higher-level system.

$\sigma_0$, $\sigma_1$ and the root server are user-level servers and not part of the pure kernel. The predefined ones can be replaced by modifying the following table in the L4 image before starting L4. An empty area specifies that the corresponding server should not be started. Note, that $\sigma_0$ is a mandatory service. The kernel debugger $kdebug$ is also not part of the kernel and can accordingly be replaced by modifying the table.

<table>
<thead>
<tr>
<th>MemoryDesc</th>
<th>MemDescPtr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sim$</td>
<td>BootInfo</td>
</tr>
<tr>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>$\sim$</td>
<td>$\sim$</td>
</tr>
<tr>
<td>Kdebug.config1</td>
<td>Kdebug.config0</td>
</tr>
<tr>
<td>root server.high</td>
<td>root server.low</td>
</tr>
<tr>
<td>$\sigma_1$.high</td>
<td>$\sigma_1$.low</td>
</tr>
<tr>
<td>$\sigma_0$.high</td>
<td>$\sigma_0$.low</td>
</tr>
<tr>
<td>Kdebug.high</td>
<td>Kdebug.low</td>
</tr>
<tr>
<td>$\sim$</td>
<td>API Version</td>
</tr>
</tbody>
</table>

The addresses are offsets relative to the configuration page’s base address. The configuration page is located at a page boundary and can be found by searching for the magic “L4 μK” starting at the load address. The IP and SP values however, are absolute addresses. The appropriate code must be loaded at these addresses before L4 is started.

**IP**

Physical address of a server’s initial instruction pointer (start).

**SP**

Physical address of a server’s initial stack pointer (stack bottom).

**Kdebug.init**

Physical address of $kdebug$’s initialization routine.
**Kdebug.entry**  
Physical address of kdebug’s exception handler entry point.

**Kdebug.low**  
Physical address of first byte of kernel debugger. Must be page aligned.

**Kdebug.high**  
Physical address of last byte of kernel debugger. Must be the last byte in page.

**Kdebug.config**  
Configuration fields which can be freely interpreted by the kernel debugger. The specific semantics of these fields are provided with the specific kernel debuggers.

**BootInfo**  
Prior to kernel initialization a boot loader can write an arbitrary value into this field. Post-initialization code, e.g., a root server can later read the field. Its value is neither changed nor interpreted by the kernel. This is the generic method for passing system information across kernel initialization.

**MemoryInfo**

<table>
<thead>
<tr>
<th>MemDescPtr</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(16/32)</td>
<td>(16/32)</td>
</tr>
</tbody>
</table>

**MemDescPtr**  
Location of first memory descriptor (as an offset relative to the configuration page’s base address). Subsequent memory descriptors are located directly following the first one. For memory descriptors that specify overlapping memory regions, later descriptors take precedence over earlier ones.

**\( n \)**  
Initially equals the number of available memory descriptors in the configuration page. Before starting L4 this number must be initialized to the number of inserted memory descriptors.

**MemoryDesc**

<table>
<thead>
<tr>
<th>( high/2^{10} ) (22/54)</th>
<th>( \sim (10) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( low/2^{10} ) (22/54)</td>
<td>( v )</td>
</tr>
</tbody>
</table>

Memory descriptors should be initialized before starting L4. The kernel may after startup insert additional memory descriptors or modify existing ones (e.g., for reserved kernel memory).

**\( high \)**  
High address of memory region.

**\( low \)**  
Low address of memory region.

**\( v \)**  
Indicates whether memory descriptor refers to physical memory (\( v = 0 \)) or virtual memory (\( v = 1 \)).

**\( type \)**  
Identifies the type of the memory descriptor.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x0</td>
<td>Undefined</td>
</tr>
<tr>
<td>0x1</td>
<td>Conventional memory</td>
</tr>
<tr>
<td>0x2</td>
<td>Reserved memory (i.e., reserved by kernel)</td>
</tr>
<tr>
<td>0x3</td>
<td>Dedicated memory (i.e., memory not available to user)</td>
</tr>
<tr>
<td>0x4</td>
<td>Shared memory (i.e., available to all users)</td>
</tr>
<tr>
<td>0xE</td>
<td>Defined by boot loader</td>
</tr>
<tr>
<td>0xF</td>
<td>Architecture dependent</td>
</tr>
</tbody>
</table>

**\( t \)**  
Identifies the precise type for boot loader specific or architecture dependent memory descriptors.

\( type = 0xE \)

The type of the memory descriptor is dependent on the bootloader. The \( t \) field specifies the exact semantics. Refer to boot loader specification for more info.
type = $0xF$

The type of the memory descriptor is architecture dependent. The $t$ field specifies the exact semantics. Refer to architecture specific part for more info (see page 113).

\[
\text{type} \neq 0xE, \text{ type} \neq 0xF
\]

The type of the memory descriptor is solely defined by the $\text{type}$ field. The content of the $t$ field is undefined.
A.1 Virtual Registers [ia32]

Thread Control Registers (TCRs)

TCRs are implemented as part of the ia32-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. Setting the UTCB address of an active thread via THREADCONTROL is similar to deletion and re-creation. There is a fixed correlation between the UtcbLocation parameter when invoking THREADCONTROL and the UTCB address. The UTCB address of the current thread can be loaded through a machine instruction

\[
\text{mov} \quad %gs[0], %r
\]

UTCB objects of the current thread can then be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible. ThreadWord0 and ThreadWord1 are free to be used by systems software (e.g., IDL compilers). The kernel associates no semantics with these words.

\[
\begin{array}{|c|}
\hline
\sim (32) \\
\hline
\vdots \\
\hline
\text{ThreadWord0 (32)} \\
\hline
\text{ThreadWord1 (32)} \\
\hline
\text{VirtualSender/ActualSender (32)} \\
\hline
\text{IntendedReceiver (32)} \\
\hline
\text{XferTimeouts (32)} \\
\hline
\text{ErrorCode (32)} \\
\hline
\sim (16) \quad \text{cop flags (8)} \quad \text{preempt flags (8)} \\
\hline
\text{ExceptionHandler (32)} \\
\hline
\text{Pager (32)} \\
\hline
\text{UserDefinedHandle (32)} \\
\hline
\text{ProcessorNo (32)} \\
\hline
\text{MyGlobalId (32)} \\
\hline
\end{array}
\]

\[
\text{MyLocalId = UTCB address (32)} \quad \text{gs:[0]}
\]

The TCR MyLocalId is not part of the UTCB. On ia32 it is identical with the UTCB address and can be loaded from memory location gs:[0].
**Message Registers (MRs)**

Memory-mapped MRs are implemented as part of the ia32-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. Setting the UTCB address of an active thread via `THREADCONTROL` is similar to deletion and re-creation. There is a fixed correlation between the `UtcpLocation` parameter when invoking `THREADCONTROL` and the UTCB address. The UTCB address of the current thread can be loaded through a machine instruction

```
mov %gs:[0], %r
```

UTCB objects of the current thread can then be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible.

MR 0 is always mapped to a general register. MR 1 and MR 2 are mapped to general registers when reading a received message; in all other cases, MR 1 and MR 2 are mapped to memory locations. MR 3...63 are always mapped to memory.

- **MR 0**
  - ESI

- **MR 1 (only for msg receive)**
  - EBX

- **MR 2 (only for msg receive)**
  - EBP

- **MR 1...63 [UTCB fields]**
  - MR 63 (32) +252
  - ...
  - MR 4 (32) +16
  - MR 3 (32) +12
  - MR 2 (except for msg receive) (32) +8
  - MR 1 (except for msg receive) (32) ← UTCB address + 4

**Buffer Registers (BRs)**

BRs are implemented as part of the ia32-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. Setting the UTCB address of an active thread via `THREADCONTROL` is similar to deletion and re-creation. There is a fixed correlation between the `UtcpLocation` parameter when invoking `THREADCONTROL` and the UTCB address. The UTCB address of the current thread can be loaded through a machine instruction

```
mov %gs:[0], %r
```

UTCB objects of the current thread can then be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible.
**VIRTUAL REGISTERS**

<table>
<thead>
<tr>
<th>BR 0…32 [UTCB fields]</th>
</tr>
</thead>
<tbody>
<tr>
<td>~ (32)</td>
</tr>
<tr>
<td>BR 0 (32)</td>
</tr>
<tr>
<td>BR 1 (32)</td>
</tr>
<tr>
<td>BR 32 (32)</td>
</tr>
</tbody>
</table>

--- UTCB address

--- UTCB address

--- UTCB address

--- UTCB address

--- UTCB address

--- UTCB address

--- UTCB address

**UTCB Memory With Undefined Semantics**

The kernel will associate no semantics with memory located at `UTCB address...UTCB address + 3`. The application can use this memory as thread local storage, e.g., for implementing the L4 API. Note, however, that the memory contents within this region may be overwritten during a system-call operating on message registers.

All undefined UTCB memory which is not covered by the above mentioned region may have kernel defined semantics.
A.2 Systemcalls [ia32]

The system-calls which are invoked by the call instruction take the target of the calls the from system-call link fields in the kernel interface page (see page 2). Each system-call link specifies an address relative to the kernel interface page’s base address. An application may use instructions other than call to invoke the system-calls, but must ensure that a valid return address resides on the stack.

---

**KERNELINTERFACE** [Slow Systemcall]

- EAX
- ECX
- EDX
- ESI
- EDI
- EBX
- EBP
- ESP

**KernelInterface → EAX** base address

**ECX** API Version

**EDX** API Flags

**ESI** Kernel ID

**EDI**

**EBX**

**EBP**

**ESP**

**lock: nop**

---

**EXCHANGEREGISTERS** [Systemcall]

- **dest**
- **EAX**
- **ECX**
- **SP**
- **EDX**
- **IP**
- **EDX**
- **ESI**
- **EBX**
- **EBP**
- **ESP**

**UserDefinedHandle**

**FLAGS**

**Pager**

- **call ExchangeRegisters**

**EAX** result

**ECX** control

**EDX**

**ESI**

**EDI**

**EBX**

**EBP**

**ESP**

**SP**

**IP**

**FLAGS**

**UserDefinedHandle**

**Pager**

“FLAGS” refers to the user-modifiable ia32 processor flags that are held in the EFLAGS register.

---

**THREADCONTROL** [Privileged Systemcall]

- **dest**
- **EAX**
- **ECX**
- **EDX**
- **ESI**
- **EDI**
- **EBX**
- **EBP**
- **ESP**

**Pager**

**Scheduler**

**SpaceSpecifier**

**UtcbLocation**

- **call ThreadControl**

**EAX** result

**ECX**

**EDX**

**ESI**

**EDI**

**EBX**

**EBP**

**ESP**

---

**SYSTEMCLOCK** [Systemcall]

- **EAX**
- **ECX**
- **EDX**
- **ESI**
- **EDI**
- **EBX**
- **EBP**
- **ESP**

- **call SystemClock**

**EAX** clock 0...31

**ECX**

**EDX** clock 32...63

**ESI**

**EDI**

**EBX**

**EBP**

**ESP**
### Thread Switch

**Thread Switch** [Systemcall]

<table>
<thead>
<tr>
<th>dest</th>
<th>EAX</th>
<th>EAX</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ECX</td>
<td></td>
<td>ECX</td>
<td></td>
</tr>
<tr>
<td>- EDX</td>
<td></td>
<td>EDX</td>
<td></td>
</tr>
<tr>
<td>- ESI</td>
<td></td>
<td>ESI</td>
<td></td>
</tr>
<tr>
<td>- EDI</td>
<td></td>
<td>EDI</td>
<td></td>
</tr>
<tr>
<td>- EBX</td>
<td></td>
<td>EBX</td>
<td></td>
</tr>
<tr>
<td>- EBP</td>
<td></td>
<td>EBP</td>
<td></td>
</tr>
<tr>
<td>- ESP</td>
<td></td>
<td>ESP</td>
<td></td>
</tr>
</tbody>
</table>

- **Thread Switch** → call **Thread Switch**

### Schedule

**Schedule** [Systemcall]

<table>
<thead>
<tr>
<th>dest</th>
<th>EAX</th>
<th>EAX</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ECX</td>
<td></td>
<td>ECX</td>
<td></td>
</tr>
<tr>
<td>- EDX</td>
<td></td>
<td>EDX</td>
<td></td>
</tr>
<tr>
<td>- ESI</td>
<td></td>
<td>ESI</td>
<td></td>
</tr>
<tr>
<td>- EDI</td>
<td></td>
<td>EDI</td>
<td></td>
</tr>
<tr>
<td>- EBX</td>
<td></td>
<td>EBX</td>
<td></td>
</tr>
<tr>
<td>- EBP</td>
<td></td>
<td>EBP</td>
<td></td>
</tr>
<tr>
<td>- ESP</td>
<td></td>
<td>ESP</td>
<td></td>
</tr>
</tbody>
</table>

- **Schedule** → call **Schedule**

### IPC

**IPC** [Systemcall]

<table>
<thead>
<tr>
<th>to</th>
<th>EAX</th>
<th>EAX</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ECX</td>
<td></td>
<td>ECX</td>
<td></td>
</tr>
<tr>
<td>- EDX</td>
<td></td>
<td>EDX</td>
<td></td>
</tr>
<tr>
<td>- ESI</td>
<td></td>
<td>ESI</td>
<td></td>
</tr>
<tr>
<td>- EDI</td>
<td></td>
<td>EDI</td>
<td></td>
</tr>
<tr>
<td>- EBX</td>
<td></td>
<td>EBX</td>
<td></td>
</tr>
<tr>
<td>- EBP</td>
<td></td>
<td>EBP</td>
<td></td>
</tr>
<tr>
<td>- ESP</td>
<td></td>
<td>ESP</td>
<td></td>
</tr>
</tbody>
</table>

- **IPC** → call **IPC**

### LIPC

**LIPC** [Systemcall]

<table>
<thead>
<tr>
<th>to</th>
<th>EAX</th>
<th>EAX</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ECX</td>
<td></td>
<td>ECX</td>
<td></td>
</tr>
<tr>
<td>- EDX</td>
<td></td>
<td>EDX</td>
<td></td>
</tr>
<tr>
<td>- ESI</td>
<td></td>
<td>ESI</td>
<td></td>
</tr>
<tr>
<td>- EDI</td>
<td></td>
<td>EDI</td>
<td></td>
</tr>
<tr>
<td>- EBX</td>
<td></td>
<td>EBX</td>
<td></td>
</tr>
<tr>
<td>- EBP</td>
<td></td>
<td>EBP</td>
<td></td>
</tr>
<tr>
<td>- ESP</td>
<td></td>
<td>ESP</td>
<td></td>
</tr>
</tbody>
</table>

- **LIPC** → call **LIPC**

### Unmap

**Unmap** [Systemcall]

<table>
<thead>
<tr>
<th>control</th>
<th>EAX</th>
<th>EAX</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ECX</td>
<td></td>
<td>ECX</td>
<td></td>
</tr>
<tr>
<td>- EDX</td>
<td></td>
<td>EDX</td>
<td></td>
</tr>
<tr>
<td>- ESI</td>
<td></td>
<td>ESI</td>
<td></td>
</tr>
<tr>
<td>- EDI</td>
<td></td>
<td>EDI</td>
<td></td>
</tr>
<tr>
<td>- EBX</td>
<td></td>
<td>EBX</td>
<td></td>
</tr>
<tr>
<td>- EBP</td>
<td></td>
<td>EBP</td>
<td></td>
</tr>
<tr>
<td>- ESP</td>
<td></td>
<td>ESP</td>
<td></td>
</tr>
</tbody>
</table>

- **Unmap** → call **Unmap**
### SPACECONTROL  [Privileged Systemcall]

<table>
<thead>
<tr>
<th>SpaceSpecifier control</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>KernelInterfacePageArea</td>
<td>ECX</td>
</tr>
<tr>
<td>UtcbArea</td>
<td>EDX</td>
</tr>
<tr>
<td>Redirector</td>
<td>EDI</td>
</tr>
<tr>
<td>−</td>
<td>EBX</td>
</tr>
<tr>
<td>−</td>
<td>EBP</td>
</tr>
<tr>
<td>−</td>
<td>ESP</td>
</tr>
</tbody>
</table>

**SPACE Control** → call SpaceControl

<table>
<thead>
<tr>
<th>EAX</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECX</td>
<td>control</td>
</tr>
<tr>
<td>EDX</td>
<td>~</td>
</tr>
<tr>
<td>ESI</td>
<td>~</td>
</tr>
<tr>
<td>EDI</td>
<td>~</td>
</tr>
<tr>
<td>EBX</td>
<td>~</td>
</tr>
<tr>
<td>EBP</td>
<td>~</td>
</tr>
<tr>
<td>ESP</td>
<td>~</td>
</tr>
</tbody>
</table>

### PROCESSORCONTROL  [Privileged Systemcall]

<table>
<thead>
<tr>
<th>ProcessorNo</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>InternalFrequency</td>
<td>ECX</td>
</tr>
<tr>
<td>ExternalFrequency</td>
<td>EDX</td>
</tr>
<tr>
<td>voltage</td>
<td>EDI</td>
</tr>
<tr>
<td>−</td>
<td>EBX</td>
</tr>
<tr>
<td>−</td>
<td>EBP</td>
</tr>
<tr>
<td>−</td>
<td>ESP</td>
</tr>
</tbody>
</table>

**Processor Control** → call ProcessorControl

<table>
<thead>
<tr>
<th>EAX</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECX</td>
<td>~</td>
</tr>
<tr>
<td>EDX</td>
<td>~</td>
</tr>
<tr>
<td>ESI</td>
<td>~</td>
</tr>
<tr>
<td>EDI</td>
<td>~</td>
</tr>
<tr>
<td>EBX</td>
<td>~</td>
</tr>
<tr>
<td>EBP</td>
<td>~</td>
</tr>
<tr>
<td>ESP</td>
<td>~</td>
</tr>
</tbody>
</table>

### MEMORYCONTROL  [Privileged Systemcall]

<table>
<thead>
<tr>
<th>control</th>
<th>EAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>attribute₀</td>
<td>ECX</td>
</tr>
<tr>
<td>attribute₁</td>
<td>EDX</td>
</tr>
<tr>
<td>MR₀</td>
<td>ESI</td>
</tr>
<tr>
<td>Utcb</td>
<td>EDI</td>
</tr>
<tr>
<td>attribute₂</td>
<td>EBX</td>
</tr>
<tr>
<td>attribute₃</td>
<td>EBP</td>
</tr>
<tr>
<td>−</td>
<td>ESP</td>
</tr>
</tbody>
</table>

**Memory Control** → call MemoryControl

| EAX | ~ |
|-----|~|
| ECX | ~ |
| EDX | ~ |
| ESI | ~ |
| EDI | ~ |
| EBX | ~ |
| EBP | ~ |
| ESP | ~ |
### A.3 Kernel Features  [ia32]

The ia32 architecture supports the following kernel feature descriptors in the kernel interface page (see page 5).

<table>
<thead>
<tr>
<th>String</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>“smallspaces”</td>
<td>Kernel has small address spaces enabled.</td>
</tr>
</tbody>
</table>
A.4 IO-Ports [ia32]

On ia32 processors, IO-ports are handled as fpages. IO fpages can be mapped, granted, and unmapped like memory fpages. Their minimal granularity is 1. An IO-fpage of size $2^s$ has a $2^s$-aligned base address $p$, i.e. $p \mod 2^s = 0$. An fpage with base port address $p$ and size $2^s$ is denoted as described below.

$$IO\ fpage \ (p, 2^s)$$

| $p$ (16/48) | $s'$ (6) | $s = 2^{(6)}$ | 0 r w x |

IO-ports can only be mapped idempotently, i.e., physical port $x$ is either mapped at IO address $x$ in the task’s IO address space, or it is not mapped at all.

---

**Generic Programming Interface**

```c
#include <l4/space.h>

Fpage IOFpage (Word BaseAddress, int FpageSize)
Fpage IOFpageLog2 (Word BaseAddress, int Log2FpageSize < 64)
```

Delivers an IO fpage with the specified location and size.
A.5 Space Control  [ia32]

The `SPACECONTROL` system call has an architecture dependent `control` parameter to specify various address space characteristics. For ia32, the `control` parameter has the following semantics.

### Input Parameter

<table>
<thead>
<tr>
<th>control</th>
<th>0 (23)</th>
<th>small (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>s</code></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A value of 1 indicates the intention to change the `small address space number` for the specified address space. The small space number will remain unchanged if `s = 0`.

- If `s = 1`, sets the small address space number for the specified address space. Small address space numbers from 1 to 255 are available. A value of 0 indicates a regular large address space.
- An assigned small space number is effective on all CPUs in an SMP system.

The position (`pos`) of the least significant bit of `small` indicates the size of the small space by the following formula: `size = 2^{pos} * 4 MB`. After removing the least significant bit, the remaining bits of `small` indicate the location of the space within a 512 MB region using the following formula: `location = small * 2 MB`. Setting the small space number fails if the specified region overlaps with an already existing one.

The `small` field is ignored if `s = 0`, or if the kernel does not support small spaces (see Kernel Features, page 92).

### Output Parameter

<table>
<thead>
<tr>
<th>control</th>
<th>0 (23)</th>
<th>small (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>e</code></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Indicates if the change of small space number was effective (`e = 1`). Undefined if `s = 0` in the input parameter.

- `small` The old value for the small space number. A value of 0 is possible even if the space has previously been put into a small address space. An implicit change to small space number 0 can happen if a thread within the space accesses memory beyond the specified small space size.

### Generic Programming Interface

```c
#include <l4/space.h>

Word LargeSpace

Word SmallSpace  (Word location, size)

Delivers a small space number with the specified `location` and `size` (both in MB). It is assumed that `size = 2^{p} * 4` for some value `p < 8`.```
A.6 Cacheability Hints  [ia32]

String items can specify cacheability hints to the kernel (see page 52). For ia32, the cacheability hints have the following semantics.

\[
\begin{align*}
hh &= 00 & \text{Use the processor’s default cacheability strategy. Typically, cache lines are allocated for data read and written (assuming that the processor’s default strategy is write-back and write-allocate).} \\
hh &= 01 & \text{Allocate cache lines in the entire cache hierarchy for data read or written.} \\
hh &= 10 & \text{Do not allocate new cache lines (entire cache hierarchy) for data read or written.} \\
hh &= 11 & \text{Allocate only new L1 cache line for data read or written. Do not allocate cache lines in lower cache hierarchies.}
\end{align*}
\]

Convenience Programming Interface

```
#include <l4/ipc.h>

CacheAllocationHint UseDefaultCacheLineAllocation
CacheAllocationHint AllocateNewCacheLines
CacheAllocationHint DoNotAllocateNewCacheLines
CacheAllocationHint AllocateOnlyNewL1CacheLines
```
A.7 Memory Attributes [ia32]

The ia32 architecture in general supports the following memory attributes values.

<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0</td>
</tr>
<tr>
<td>Uncacheable</td>
<td>1</td>
</tr>
<tr>
<td>Write Combining</td>
<td>2</td>
</tr>
<tr>
<td>Write Through</td>
<td>5</td>
</tr>
<tr>
<td>Write Protected</td>
<td>6</td>
</tr>
<tr>
<td>Write Back</td>
<td>7</td>
</tr>
</tbody>
</table>

Note that some attributes are only supported on certain processors. See the “IA-32 Intel Architecture Software Developer’s Manual, Volume 3: System Programming Guide” for the semantics of the memory attributes and which processors they are supported on.

---

**Generic Programming Interface**

```c
#include <l4/misc.h>

Word DefaultMemory
Word UncacheableMemory
Word WriteCombiningMemory
Word WriteThroughMemory
Word WriteProtectedMemory
Word WriteBackMemory
```
## A.8 Exception Message Format

**To Exception Handler**

<table>
<thead>
<tr>
<th>Register</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAX (32)</td>
<td>MR 12</td>
</tr>
<tr>
<td>ECX (32)</td>
<td>MR 11</td>
</tr>
<tr>
<td>EDX (32)</td>
<td>MR 10</td>
</tr>
<tr>
<td>EBX (32)</td>
<td>MR  9</td>
</tr>
<tr>
<td>ESP (32)</td>
<td>MR  8</td>
</tr>
<tr>
<td>EBP (32)</td>
<td>MR  7</td>
</tr>
<tr>
<td>ESI (32)</td>
<td>MR  6</td>
</tr>
<tr>
<td>EDI (32)</td>
<td>MR  5</td>
</tr>
<tr>
<td>ErrorCode (32)</td>
<td>MR  4</td>
</tr>
<tr>
<td>ExceptionNo (32)</td>
<td>MR  3</td>
</tr>
<tr>
<td>EFLAGS (32)</td>
<td>MR  2</td>
</tr>
<tr>
<td>EIP (32)</td>
<td>MR  1</td>
</tr>
<tr>
<td>MR 0</td>
<td></td>
</tr>
</tbody>
</table>

¢4/−5 (12/44) 0 (4) 0 (4) t = 0 (6) u = 12 (6)

#PF (page fault), #MC (machine check exception), and some #GP (general protection), #SS (stack segment fault), and #NM (no math coprocessor) exceptions are handled by the kernel and therefore do not generate exception messages.

Note that executing an INT n instructions in 32-bit mode will always raise a #GP (general protection). The exception handler may interpret the error code (8n + 2, see processor manual) and emulate the INT n accordingly.
A.9 Processor Mirroring

Segments

L4 uses a flat (unsegmented) memory model. There are only three segments available: \textit{user\_space}, a read/write segment, \textit{user\_space\_exec}, an executable segment, and \textit{utcb\_address}, a read-only segment. Both \textit{user\_space} and \textit{user\_space\_exec} cover (at least) the complete user-level address space. \textit{Utcb\_address} covers only enough memory to hold the UTCB address.

The values of segment selectors are undefined. When a thread is created, its segment registers SS, DS, ES and FS are initialized with \textit{user\_space}, GS with \textit{utcb\_address}, and CS with \textit{user\_space\_exec}. Whenever the kernel detects a general protection exception and the segment registers are not loaded properly, it reloads them with the above mentioned selectors. From the user’s point of view, the segment registers cannot be modified.

However, the binary representation of \textit{user\_space} and \textit{user\_space\_exec} may change at any point during program execution. Never rely on any particular value.

Furthermore, the LSL (load segment limit) machine instruction may deliver wrong segment limits, even floating ones. The result of this instruction is always undefined.

Debug Registers

User-level debug registers exist per thread. DR0...3, DR6 and DR7 can be accessed by the machine instructions \texttt{mov n,DRx} and \texttt{mov DRx,r}. However, only task-local breakpoints can be activated, i.e., bits G0...3 in DR7 cannot be set. Breakpoints operate per thread. Breakpoints are signaled as #DB exception (INT 1).

Note that user-level breakpoints are suspended when kernel breakpoints are set by the kernel debugger.

Model-Specific Registers

All privileged threads in the system have read and write access to all the Model-Specific Registers (MSRs) of the CPU. Modification of some MSRs may lead to undefined system behavior. Any access to an MSR by an unprivileged thread will raise an exception.
### A.10 Booting

**PC-compatible Machines**

L4 can be loaded at any 16-byte-aligned location beyond 0x1000 in physical memory. It can be started in real mode or in 32-bit protected mode at address 0x100 or 0x1000 relative to its load address. The protected-mode conditions are compliant to the Multiboot Standard Version 0.6.

<table>
<thead>
<tr>
<th>Start Preconditions</th>
<th>Real Mode</th>
<th>32-bit Protected Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>load base (L)</td>
<td>( L \geq 0x1000 ), 16-byte aligned</td>
<td>( L \geq 0x1000 )</td>
</tr>
<tr>
<td>load offset (X)</td>
<td>( X = 0x100 ) or ( X = 0x1000 )</td>
<td>( X = 0x100 ) or ( X = 0x1000 )</td>
</tr>
<tr>
<td>Interrupts</td>
<td>disabled</td>
<td>disabled</td>
</tr>
<tr>
<td>Gate A20</td>
<td>( \sim )</td>
<td>open</td>
</tr>
<tr>
<td>EFLAGS</td>
<td>I=0</td>
<td>I=0, VM=0</td>
</tr>
<tr>
<td>CR0</td>
<td>PE=0</td>
<td>PE=1, PG=0</td>
</tr>
<tr>
<td>(E)IP</td>
<td>( X )</td>
<td>( L + X )</td>
</tr>
<tr>
<td>CS</td>
<td>( L/16 )</td>
<td>0, 4GB, 32-bit exec</td>
</tr>
<tr>
<td>SS,DS,ES</td>
<td>( \sim )</td>
<td>0, 4GB, read/write</td>
</tr>
<tr>
<td>EAX</td>
<td>( \sim )</td>
<td>0x2BADB002</td>
</tr>
<tr>
<td>EBX</td>
<td>( \sim )</td>
<td>( \ast P )</td>
</tr>
<tr>
<td>((P + 0))</td>
<td>( \sim )</td>
<td>( \sim ) OR 1</td>
</tr>
<tr>
<td>((P + 4))</td>
<td>n/a</td>
<td>below 640 K mem in K</td>
</tr>
<tr>
<td>((P + 8))</td>
<td></td>
<td>beyond 1M mem in K</td>
</tr>
<tr>
<td>all remaining registers &amp; flags</td>
<td>( \sim )</td>
<td>( \sim )</td>
</tr>
</tbody>
</table>

L4 relocates itself to 0x1000, enters protected mode if started in real mode, enables paging and initializes itself.
Appendix B

IA-64 Interface
B.1 Virtual Registers

Thread Control Registers (TCRs)

TCRs are mapped to memory locations. They are implemented as part of the ia64-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. (In fact, the ia64 UTCB address is identical to the thread’s local ID.) Register ar.k6 always contains the UTCB address of the current thread. UTCBs of other threads must not be accessed, even if they are physically accessible. ThreadWord0 and ThreadWord1 are free to be used by systems software (e.g., IDL compilers). The kernel associates no semantics with these words.

<table>
<thead>
<tr>
<th>Field</th>
<th>Offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreadWord 1</td>
<td>+352</td>
</tr>
<tr>
<td>ThreadWord 0</td>
<td>+344</td>
</tr>
<tr>
<td>ErrorCode</td>
<td>+72</td>
</tr>
<tr>
<td>VirtualSender/ActualSender</td>
<td>+64</td>
</tr>
<tr>
<td>IntendedReceiver</td>
<td>+56</td>
</tr>
<tr>
<td>XferTimeouts</td>
<td>+48</td>
</tr>
<tr>
<td>∼ (48)</td>
<td></td>
</tr>
<tr>
<td>cop flags (8)</td>
<td>+40</td>
</tr>
<tr>
<td>preempt flags (8)</td>
<td></td>
</tr>
<tr>
<td>ExceptionHandler</td>
<td>+32</td>
</tr>
<tr>
<td>Pager</td>
<td>+24</td>
</tr>
<tr>
<td>UserDefinedHandle</td>
<td>+16</td>
</tr>
<tr>
<td>ProcessorNo</td>
<td></td>
</tr>
<tr>
<td>←− UTCB address + 8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyLocalId = UTCB address</td>
<td>ar.k6</td>
</tr>
<tr>
<td>MyGlobalId</td>
<td>ar.k5</td>
</tr>
</tbody>
</table>

Message Registers (MRs)

Memory-mapped MRs are implemented as part of the ia64-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. (In fact, the ia64 UTCB address is identical to the thread’s local ID.) Register ar.k6 always contains the UTCB address of the current thread. UTCBs of other threads must not be accessed, even if they are physically accessible.

MR 0...7 are mapped to the eight first output registers on the register stack. The exact location of the first eight message registers therefore depends on the configuration of the current frame marker (CFM). MR 8...63 are mapped to memory. It is valid to configure less than eight output registers in the current register frame if a message to be transferred spans less than eight message registers. The number of message registers must not exceed the number of output registers, however.
Virtual Registers

MR_{0...7}

<table>
<thead>
<tr>
<th>MR 0</th>
<th>MR 1</th>
<th>MR 2</th>
<th>MR 3</th>
<th>MR 4</th>
<th>MR 5</th>
<th>MR 6</th>
<th>MR 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>out 0</td>
</tr>
</tbody>
</table>

MR_{8...63} [UTCB fields]

<table>
<thead>
<tr>
<th>MR 0</th>
<th>MR 1</th>
<th>MR 2</th>
<th>MR 3</th>
<th>MR 4</th>
<th>MR 5</th>
<th>MR 6</th>
<th>MR 7</th>
<th>MR 8</th>
<th>MR 9</th>
<th>MR 10</th>
<th>...</th>
<th>MR 63</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<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>
<td></td>
<td></td>
<td></td>
<td>+888</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+456</td>
</tr>
</tbody>
</table>

Buffer Registers (BRs)

BRs are implemented as part of the ia64-specific user-level thread control block (UTCB). The address of the current thread’s UTCB will not change over the lifetime of the thread. (In fact, the ia64 UTCB address is identical to the thread’s local ID.) Register ar.k6 always contains the UTCB address of the current thread. UTCBs of other threads must not be accessed, even if they are physically accessible.

BR_{0...32} [UTCB fields]

<table>
<thead>
<tr>
<th>BR 0</th>
<th>BR 1</th>
<th>BR 2</th>
<th>BR 3</th>
<th>BR 4</th>
<th>BR 5</th>
<th>BR 6</th>
<th>BR 7</th>
<th>BR 8</th>
<th>BR 9</th>
<th>BR 10</th>
<th>...</th>
<th>BR 32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<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>
<td></td>
<td></td>
<td></td>
<td>+336</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+88</td>
</tr>
</tbody>
</table>

UTCB Memory With Undefined Semantics

The kernel will associate no semantics with memory located at UTCB address + 384...UTCB address + 447. The application can use this memory as thread local storage, e.g., for implementing the L4 API. Note, however, that the memory contents within this region may be overwritten during a system-call operating on message registers.

All undefined UTCB memory which is not covered by the above mentioned region may have kernel defined semantics.
B.2 PAL and SAL Access  [ia64]

The microkernel provides special system-calls for accessing Processor Abstraction Level (PAL) and System Abstraction Layer (SAL) procedures. The location of the additional system-call links in the kernel interface page are as follows:

<table>
<thead>
<tr>
<th>Location</th>
<th>System-call</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel Interface Page + 0x220</td>
<td>PAL_CALL</td>
</tr>
<tr>
<td>Kernel Interface Page + 0x228</td>
<td>SAL_CALL</td>
</tr>
</tbody>
</table>

---

### Generic Programming Interface

**System-Call Function:**

```c
#include <l4/arch.h>

Word PAL_Call (Word idx, a1, a2, a3, Word& r1, r2, r3)

Invoke the PAL procedure specified by idx. a1...a3 are the arguments to the PAL procedure. r1...r3 are the return values. The system-call returns the status of the procedure invocation. See the “Intel Itanium Architecture Software Developer’s Manual, Volume 2: System Architecture” for the possible values of idx, and the contents of arguments and return values.

As of now, no invocation of PAL procedures is allowed by any user-level thread.
```

```c
#include <l4/arch.h>

Word SAL_Call (Word idx, a1, a2, a3, a4, a5, a6, Word& r1, r2, r3)

Invoke the SAL procedure specified by idx. a1...a6 are the arguments to the SAL procedure. r1...r3 are the return values. The system-call returns the status of the procedure invocation. See the “Itanium Processor Family System Abstraction Layer Specification” for possible values of idx, and the contents of arguments and return values.

As of now, only the PCI_CONFIG_READ and PCI_CONFIG_WRITE procedure calls can be invoked from a user-level thread.
```

---

### Convenience Programming Interface

**Derived Functions:**

```c
#include <l4/arch.h>

Word SAL_PCI_ConfigRead (Word address, size, Word& value)

Read from the PCI configuration space at address with the indicated word size (1, 2 or 4 bytes). The read value is returned in value. Return the status of the operation (0 if success). The operation will only succeed if the address in the PCI configuration space is mapped readable (see page 110)
```

```c
#include <l4/arch.h>

Word SAL_PCI_ConfigWrite (Word address, size, value)

Write value to the PCI configuration space at address with the indicated word size (1, 2 or 4 bytes). Return the status of the operation (0 if success). The operation will only succeed if the address in the PCI configuration space is mapped writeable (see page 110).
```
B.3 Systemcalls  [ia64]

The system-calls which are invoked by the br.call instruction take the target of the calls from system-call link fields in the kernel interface page (see page 2). Each system-call link value, v, specifies either an absolute address (if \( v \geq 1\text{MB} \)) or an address relative to the kernel interface page's base address (if \( v < 1\text{MB} \)). An application may use instructions other than br.call to invoke the system-calls, but must ensure that a valid return address resides in the b0 register. For the IPC and LIPC system-calls the application must additionally ensure that message registers are mapped into input registers after invoking the system-call (i.e., the output registers if one were to use a br.call instruction).

The system-call definitions below only specify the contents of the general registers. Except for the KERNELINTERFACE, IPC and LIPC system-calls, the contents of the remaining user accessible registers closely resembles the IA-64 software calling conventions. More precisely, the register contents of these registers are ignored upon system-call entry, and the contents after system-call exit are defined as follows:

### Floating-point Registers:
- f0...f1  fixed
- f2...f5  preserved
- f6...f15 scratch
- f16...f127 preserved

### Predicate Registers:
- p0  fixed
- p1...p5  preserved
- p6...p15 scratch
- p16...p63 preserved

### Branch Registers:
- b0  system-call return address
- b1...b5 preserved
- b6...b7 scratch

The application registers are:

### Application Registers:
- ar.fpsr special (see below)
- ar.rnat preserved
- ar.unat preserved
- ar.pfs scratch
- ar.bsp preserved
- ar.bsstore preserved
- ar.rsc special (see below)
- ar.le preserved
- ar.ec preserved
- ar.ccv scratch
- ar.ite scratch
- ar.k0...k4 scratch
- ar.k5 MyGlobalId
- ar.k6 MyLocalId
- ar.k7 scratch

The ar.fpsr and ar.rsc registers are special. The second and third status fields of ar.fpsr, and the loadrs field of ar.rsc have scratch semantics. The remaining fields have preserved semantics.

### KERNELINTERFACE  [Slow Systemcall]

- \( r1...r7 \)
- \( r8 \)
- \( r9 \)
- \( r10 \)
- \( r11 \)
- \( r12...r31 \)
- in0...in95
- loc0...loc95
- out0...out95

The \( \{ \text{mlx (ap) break.m 0x1face (ap) movl r0 = 0x0 ;} \} \) all other registers remain unchanged. A qualifying predicate, \( qp \), can be used to conditionally execute the KERNELINTERFACE system-call.

---

**SYSTEMCALLS** 105
### EXCHANGREGISTERS [Systemcall]

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>r1</td>
<td>- ExchangeRegisters → r1</td>
</tr>
<tr>
<td>r2...r3</td>
<td>r2...r3</td>
<td>~</td>
</tr>
<tr>
<td>r4...r7</td>
<td>r4...r7</td>
<td>~</td>
</tr>
<tr>
<td>r8...r11</td>
<td>r8...r11</td>
<td>~</td>
</tr>
<tr>
<td>r12...r15</td>
<td>r12...r13</td>
<td>~</td>
</tr>
<tr>
<td>dest</td>
<td>r14</td>
<td>result</td>
</tr>
<tr>
<td>control</td>
<td>r15</td>
<td>control</td>
</tr>
<tr>
<td>SP</td>
<td>r16</td>
<td>SP</td>
</tr>
<tr>
<td>IP</td>
<td>r17</td>
<td>IP</td>
</tr>
<tr>
<td>FLAGS</td>
<td>r18</td>
<td>FLAGS</td>
</tr>
<tr>
<td>UserDefinedHandle</td>
<td>r19</td>
<td>UserDefinedHandle</td>
</tr>
<tr>
<td>pager</td>
<td>r20</td>
<td>pager</td>
</tr>
<tr>
<td>r21...r31</td>
<td>r21...r31</td>
<td>~</td>
</tr>
<tr>
<td>out0...out95</td>
<td>out0...out95</td>
<td>~</td>
</tr>
</tbody>
</table>

### THREADCONTROL [Privileged Systemcall]

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>r1</td>
<td>- ThreadControl → r1</td>
</tr>
<tr>
<td>r2...r3</td>
<td>r2...r3</td>
<td>~</td>
</tr>
<tr>
<td>r4...r7</td>
<td>r4...r7</td>
<td>~</td>
</tr>
<tr>
<td>r8</td>
<td>r8</td>
<td>result</td>
</tr>
<tr>
<td>r9...r11</td>
<td>r9...r11</td>
<td>~</td>
</tr>
<tr>
<td>r12...r13</td>
<td>r12...r13</td>
<td>~</td>
</tr>
<tr>
<td>dest</td>
<td>r14</td>
<td>~</td>
</tr>
<tr>
<td>SpaceSpecifier</td>
<td>r15</td>
<td>~</td>
</tr>
<tr>
<td>Scheduler</td>
<td>r16</td>
<td>~</td>
</tr>
<tr>
<td>Pager</td>
<td>r17</td>
<td>~</td>
</tr>
<tr>
<td>UtcbLocation</td>
<td>r18</td>
<td>~</td>
</tr>
<tr>
<td>r19...r31</td>
<td>r19...r31</td>
<td>~</td>
</tr>
<tr>
<td>out0...out95</td>
<td>out0...out95</td>
<td>~</td>
</tr>
</tbody>
</table>

### SYSTEMCLOCK [Systemcall]

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>r1</td>
<td>- SystemClock → r1</td>
</tr>
<tr>
<td>r2...r3</td>
<td>r2...r3</td>
<td>~</td>
</tr>
<tr>
<td>r4...r7</td>
<td>r4...r7</td>
<td>~</td>
</tr>
<tr>
<td>r8</td>
<td>r8</td>
<td>clock</td>
</tr>
<tr>
<td>r9...r11</td>
<td>r9...r11</td>
<td>~</td>
</tr>
<tr>
<td>r12...r13</td>
<td>r12...r13</td>
<td>~</td>
</tr>
<tr>
<td>r14...r31</td>
<td>r14...r31</td>
<td>~</td>
</tr>
<tr>
<td>out0...out95</td>
<td>out0...out95</td>
<td>~</td>
</tr>
</tbody>
</table>

### THREADSWITCH [Systemcall]

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>r1</td>
<td>- ThreadSwitch → r1</td>
</tr>
<tr>
<td>r2...r3</td>
<td>r2...r3</td>
<td>~</td>
</tr>
<tr>
<td>r4...r7</td>
<td>r4...r7</td>
<td>~</td>
</tr>
<tr>
<td>r8...r11</td>
<td>r8...r11</td>
<td>~</td>
</tr>
<tr>
<td>r12...r13</td>
<td>r12...r13</td>
<td>~</td>
</tr>
<tr>
<td>dest</td>
<td>r14</td>
<td>~</td>
</tr>
<tr>
<td>r15...r31</td>
<td>r15...r31</td>
<td>~</td>
</tr>
<tr>
<td>out0...out95</td>
<td>out0...out95</td>
<td>~</td>
</tr>
</tbody>
</table>
**SCHEDULE [Systemcall]**

- \( r_1 \)
- \( r_{2..r3} \)
- \( r_4..r7 \)
- \( r_8 \)
- \( r_9 \)
- \( r_{10..r11} \)
- \( r_{12..r13} \)
- \( \text{dest} \)
- \( r_{14} \)
- \( \text{time control} \)
- \( r_{15} \)
- \( \text{processor control} \)
- \( r_{16} \)
- \( r_{17} \)
- \( \text{preemption control} \)
- \( r_{18} \)
- \( r_{19..r31} \)
- \( \text{out}_{0..95} \)

\[
\text{br.call } b_0 = \text{Schedule} \\
\]

- \( r_1 \equiv \)
- \( r_{2..r3} \equiv \)
- \( r_4..r7 \equiv \)
- \( r_8 \equiv \)
- \( r_9 \equiv \)
- \( r_{10..r11} \equiv \)
- \( r_{12..r13} \equiv \)
- \( r_{14} \equiv \)
- \( r_{15} \equiv \)
- \( r_{16} \equiv \)
- \( r_{17} \equiv \)
- \( r_{18} \equiv \)
- \( r_{19..r31} \equiv \)
- \( \text{out}_{0..95} \equiv \)

**IPC [Systemcall]**

- \( r_1 \)
- \( r_{2..r8} \)
- \( r_9 \)
- \( r_{10..r11} \)
- \( r_{12} \)
- \( r_{13} \)
- \( \text{to} \)
- \( r_{14} \)
- \( \text{FromSpecifer} \)
- \( r_{15} \)
- \( \text{Timeouts} \)
- \( r_{16} \)
- \( \text{MR}_{0,31} \)
- \( \text{MR}_{2,31} \)
- \( \text{MR}_{3,31} \)
- \( \text{MR}_{4,31} \)
- \( \text{MR}_{5,31} \)
- \( \text{MR}_{6,31} \)
- \( \text{MR}_{7,31} \)
- \( \text{out}_{0..95} \)

\[
\text{br.call } b_0 = \text{Ipc} \\
\]

- \( r_1 \equiv \)
- \( r_{2..r8} \equiv \)
- \( r_9 \equiv \)
- \( r_{10..r11} \equiv \)
- \( r_{12} \equiv \)
- \( r_{13} \equiv \)
- \( r_{14} \equiv \)
- \( r_{15} \equiv \)
- \( r_{16} \equiv \)
- \( r_{17..r31} \equiv \)
- \( \text{out}_{0} \equiv \text{MR}_{0} \)
- \( \text{out}_{1} \equiv \text{MR}_{1} \)
- \( \text{out}_{2} \equiv \text{MR}_{2} \)
- \( \text{out}_{3} \equiv \text{MR}_{3} \)
- \( \text{out}_{4} \equiv \text{MR}_{4} \)
- \( \text{out}_{5} \equiv \text{MR}_{5} \)
- \( \text{out}_{6} \equiv \text{MR}_{6} \)
- \( \text{out}_{7} \equiv \text{MR}_{7} \)
- \( \text{out}_{8..95} \equiv \)

**LIPC [Systemcall]**

- \( r_1 \)
- \( r_{2..r8} \)
- \( r_9 \)
- \( r_{10..r11} \)
- \( r_{12} \)
- \( r_{13} \)
- \( \text{to} \)
- \( r_{14} \)
- \( \text{FromSpecifer} \)
- \( r_{15} \)
- \( \text{Timeouts} \)
- \( r_{16} \)
- \( \text{MR}_{0,31} \)
- \( \text{MR}_{2,31} \)
- \( \text{MR}_{3,31} \)
- \( \text{MR}_{4,31} \)
- \( \text{MR}_{5,31} \)
- \( \text{MR}_{6,31} \)
- \( \text{MR}_{7,31} \)
- \( \text{out}_{0..95} \)

\[
\text{br.call } b_0 = \text{Lipc} \\
\]

- \( r_1 \equiv \)
- \( r_{2..r8} \equiv \)
- \( r_9 \equiv \)
- \( r_{10..r11} \equiv \)
- \( r_{12} \equiv \)
- \( r_{13} \equiv \)
- \( r_{14} \equiv \)
- \( r_{15} \equiv \)
- \( r_{16} \equiv \)
- \( r_{17..r31} \equiv \)
- \( \text{out}_{0} \equiv \text{MR}_{0} \)
- \( \text{out}_{1} \equiv \text{MR}_{1} \)
- \( \text{out}_{2} \equiv \text{MR}_{2} \)
- \( \text{out}_{3} \equiv \text{MR}_{3} \)
- \( \text{out}_{4} \equiv \text{MR}_{4} \)
- \( \text{out}_{5} \equiv \text{MR}_{5} \)
- \( \text{out}_{6} \equiv \text{MR}_{6} \)
- \( \text{out}_{7} \equiv \text{MR}_{7} \)
- \( \text{out}_{8..95} \equiv \)

All remaining registers (including application registers) will have scratch semantics over the IPC system-call. Upon entry to the IPC system-call, the register stack backing store must be able to contain the dirty partition of the register stack.
All remaining registers (including application registers) will have scratch semantics over the LIPC system-call. Upon entry to the LIPC system-call, the register stack backing store must be able to contain the dirty partition of the register stack.

UNMAP [Systemcall]

- $r1$
- $r2...r3$
- $r4...r7$
- $r8...r11$
- r12...r13

control r14
- r16...r31

MR 0 out0
MR 1 out1
MR 2 out2
MR 3 out3
MR 4 out4
MR 5 out5
MR 6 out6
MR 7 out7
- out8...out95

- Unmap →
  - br.call b0 = Unmap

  \[ r1 \equiv r2...r3 \sim r4...r7 \equiv r8...r11 \sim r12...r13 \equiv r14 \sim r16...r31 \sim \]

  \[ \text{out0} \sim \text{MR 0} \]
  \[ \text{out1} \sim \text{MR 1} \]
  \[ \text{out2} \sim \text{MR 2} \]
  \[ \text{out3} \sim \text{MR 3} \]
  \[ \text{out4} \sim \text{MR 4} \]
  \[ \text{out5} \sim \text{MR 5} \]
  \[ \text{out6} \sim \text{MR 6} \]
  \[ \text{out7} \sim \text{MR 7} \]
  \[ \text{out8...out95} \sim \]

SPACECONTROL [Privileged Systemcall]

- $r1$
- $r2...r3$
- $r4...r7$
- $r8$
- $r9$
- $r10...r11$
- r12...r13

SpaceSpecifier control r15
KernelInterfacePageArea r16
UtebArea r17
Redirector r18
- r19...r31
- out0...out95

- Space Control →
  - br.call b0 = SpaceControl

  \[ r1 \equiv r2...r3 \sim r4...r7 \equiv r8 \rightarrow \text{result} \]
  \[ r9 \sim \text{control} \]
  \[ r10...r11 \sim \]
  \[ r12...r13 \sim \]
  \[ r14 \sim \]
  \[ r15 \sim \]
  \[ r16 \sim \]
  \[ r17 \sim \]
  \[ r18 \sim \]
  \[ r19...r31 \sim \]
  \[ \text{out0...out95} \sim \]

PROCESSORCONTROL [Privileged Systemcall]

- $r1$
- $r2...r3$
- $r4...r7$
- $r8$
- $r9...r11$
- r12...r13

ProcessorNo r14
InternalFrequency r15
ExternalFrequency r16
Voltage r17
- r18...r31
- out0...out95

- Processor Control →
  - br.call b0 = ProcessorControl

  \[ r1 \equiv r2...r3 \sim r4...r7 \equiv r8 \rightarrow \text{result} \]
  \[ r9...r11 \sim \]
  \[ r12...r13 \sim \]
  \[ r14 \sim \]
  \[ r15 \sim \]
  \[ r16 \sim \]
  \[ r17 \sim \]
  \[ r18...r31 \sim \]
  \[ \text{out0...out95} \sim \]
**MEMORYCONTROL** [Privileged Systemcall]

- r1
- r2...r3
- r4...r7
- r8...r11
- r12...r13

control
attribute0
attribute1
attribute2
attribute3
MR0
MR1
MR2
MR3
MR4
MR5
MR6
MR7
- out0...out95

___

- Memory Control →

br.call b0 = MemoryControl

r4
r2...r3
r4...r7
r8...r11
r12...r13
r14
r15
r16
r17
r18
r19...r31
out0
out1
out2
out3
out4
out5
out6
out7
out8...out95

---

**PAL_CALL** [Architecture Specific Systemcall]

- r1
- r2...r3
- r4...r7
- r8
- r9
- r10
- r11
- r12...r13
- r14...r27

idx
arg1
arg2
arg3
- out0...out95

___

- PAL Call →

br.call b0 = PAL_Call

r1
r2...r3
r4...r7
r8
status
r9
ret1
r10
ret2
r11
ret3
r12...r13
r14...r27
r28
r29
r30
r31
out0...out95

---

**SAL_CALL** [Architecture Specific Systemcall]

- r1
- r2...r3
- r4...r7
- r8
- r9
- r10
- r11
- r12...r13
- r14...r31

idx
arg1
arg2
arg3
arg4
arg5
arg6
- out0...out95

___

- SAL Call →

br.call b0 = SAL_Call

r1
r2...r3
r4...r7
r8
status
r9
ret1
r10
ret2
r11
ret3
r12...r13
r14...r31
out0
out1
out2
out3
out4
out5
out6
out7...out95
**B.4 PCI Configuration Space**  [ia64]

On ia64 processors, the PCI configuration space is handled as fpages. PCI Config fpages can be mapped, granted, and unmapped like memory fpages. Their minimal granularity is 256 (i.e., one single device function). A PCI config fpage of size $2^s$ has a $2^s$-aligned base address $p$, i.e. $p \mod 2^s = 0$. An fpage with base PCI configuration address $p$ and size $2^s$ is denoted as described below.

```
P (48)  s' (6)  s = 2 (6)  0 r w x
```

The execute bit of the PCI config fpage is ignored.

---

**Generic Programming Interface**

```
#include <l4/space.h>

Fpage PCIConfigFpage (Word BaseAddress, int FpageSize \geq 256)
Fpage PCIConfigFpageLog2 (Word BaseAddress, int Log2FpageSize \leq 64)

Delivers a PCI config fpage with the specified location and size.
```
B.5 Cacheability Hints  [ia64]

String items can specify cacheability hints to the kernel (see page 52). For ia64, the cacheability hints have the following semantics.

\[ hh = 00 \]
Use the default cacheability strategy. Temporal locality is assumed for all cache levels. That is, cache lines are allocated on all levels for both data read and written.

\[ hh = 01 \]
No temporal locality is assumed for the first level cache. Temporal locality is assumed for all lower cache levels. That is, cache lines are allocated on all cache levels below L1 for both data read and written.

\[ hh = 10 \]
No temporal locality is assumed for the first and second level caches. Temporal locality is assumed for all lower cache levels. That is, cache lines are allocated on all cache levels below L2 for both data read and written.

\[ hh = 11 \]
No temporal locality is assumed on any cache level. That is, cache lines are not allocated on any cache level.

Note that support for cacheability hints is processor dependent. Refer to the processor specification to see what type of locality hints the processor supports for load and store instructions.

---

**Convenience Programming Interface**

```
#include <l4/ipc.h>

CacheAllocationHint UseDefaultCacheLineAllocation
CacheAllocationHint CacheNonTemporalL1
CacheAllocationHint CacheNonTemporalL2
CacheAllocationHint CacheNonTemporalAllLevels
```
B.6 Memory Attributes [ia64]

The ia64 architecture in general supports the following memory attributes values.

<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0</td>
</tr>
<tr>
<td>Write Back</td>
<td>1</td>
</tr>
<tr>
<td>Write Coalescing</td>
<td>7</td>
</tr>
<tr>
<td>Uncacheable</td>
<td>5</td>
</tr>
<tr>
<td>Uncacheable Exported</td>
<td>6</td>
</tr>
<tr>
<td>NaT Page</td>
<td>8</td>
</tr>
</tbody>
</table>

Note that some attributes are only supported on certain processors. See the “Intel Itanium Architecture Software Developer’s Manual, Volume 2: System Architecture” for the semantics of the memory attributes.

---

Generic Programming Interface

```
#include <l4/misc.h>

Word DefaultMemory
Word WriteBackMemory
Word WriteCoalescingMemory
Word UncacheableMemory
Word UncacheableExportedMemory
Word NaTPageMemory
```
B.7 Memory Descriptors [ia64]

The following memory descriptors (see page 6) are specific to the ia64 architecture.

<table>
<thead>
<tr>
<th>t</th>
<th>type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x1</td>
<td>0xF</td>
<td>ACPI Memory</td>
</tr>
</tbody>
</table>

#include <l4/kip.h>

Word ACPIMemoryType
B.8 Exception Message Format [ia64]

To be defined.
Appendix C

PowerPC Interface
C.1 Virtual Registers

Thread Control Registers (TCRs)

TCRs are mapped to memory locations. They are implemented as part of the PowerPC-specific user-level thread control block (UTCB). The address of the current thread’s UTCB is identical to the thread’s local ID, and is thus immutable. The UTCB address is provided in the general purpose register R2 at application start. The R2 register must contain the UTCB address for every system call invocation. UTCB objects of the current thread can be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible. ThreadWord0 and ThreadWord1 are free to be used by systems software (e.g., IDL compilers). The kernel associates no semantics with these words.

The TCR MyLocalId is not part of the UTCB. On PowerPC it is identical with the UTCB address and can be loaded from register R2.

Message Registers (MRs)

Message registers MR₀ through MR₉ map to the processor’s general purpose register file for IPC and LIPO calls. The remaining message registers map to memory locations in the UTCB. MR₁₀ starts at byte offset 40 in the UTCB, and successive message registers follow in memory.

For the other system calls, message registers map to memory locations in the UTCB, with MR₀ starting at byte offset zero.
### Virtual Registers

#### MR 0...9

<table>
<thead>
<tr>
<th>MR 0</th>
<th>MR 9</th>
<th>R0</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR 8</td>
<td></td>
<td>R10</td>
</tr>
<tr>
<td>MR 7</td>
<td></td>
<td>R9</td>
</tr>
<tr>
<td>MR 6</td>
<td></td>
<td>R8</td>
</tr>
<tr>
<td>MR 5</td>
<td></td>
<td>R7</td>
</tr>
<tr>
<td>MR 4</td>
<td></td>
<td>R6</td>
</tr>
<tr>
<td>MR 3</td>
<td></td>
<td>R5</td>
</tr>
<tr>
<td>MR 2</td>
<td></td>
<td>R4</td>
</tr>
<tr>
<td>MR 1</td>
<td></td>
<td>R3</td>
</tr>
<tr>
<td>MR 0</td>
<td></td>
<td>R14</td>
</tr>
</tbody>
</table>

#### MR 10...63 [UTCB fields]

| MR 63 (32) | +252 |
| :          | :    |
| MR 11 (32) | +44  |
| MR 10 (32) | ← UTCB address + 40 |

### Buffer Registers (BRs)

The buffer registers map to memory locations in the UTCB. BR 0 is at byte offset -64 in the UTCB, BR 1 at byte offset -68, etc.

#### BR 0...32 [UTCB fields]

| ~ (32) | ← UTCB address |
| :     | :              |
| BR 0 (32) | -64 |
| BR 1 (32) | -68 |
| :     | :              |
| BR 32 (32) | -196 |

### UTCB Memory With Undefined Semantics

The kernel will associate no semantics with memory located at UTCB address...UTCB address + 39. The application can use this memory as thread local storage, e.g., for implementing the L4 API. Note, however, that the memory contents within this region may be overwritten during a system-call operating on message registers.

All undefined UTCB memory which is not covered by the above mentioned region may have kernel defined semantics.
C.2 Systemcalls [powerpc]

The system-calls invoked via the `bl` instruction are located in the kernel’s area of the virtual address space. Their precise locations are stored in the kernel interface page (see page 2). One may invoke the system calls with any instruction that branches to the appropriate target, as long as the link-return register (LR) contains the correct return address.

The locations of the system-calls are fixed during the life of an application. But they may change outside of the life of an application. It is not valid to prelink an application against a set of system call locations. The official locations are always provided in the kip.

The registers defined to survive across system-call invocations (unless otherwise noted) are: R1, R2, R30, R31, and the floating point registers. All other registers contain return values, are undefined, or may be preserved according to processor specific rules.

The R2 register must contain the UTCB pointer when invoking all system calls.

---

**KERNELINTERFACE [Slow Systemcall]**

<table>
<thead>
<tr>
<th>UTCB</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R3</td>
</tr>
<tr>
<td></td>
<td>R4</td>
</tr>
<tr>
<td></td>
<td>R5</td>
</tr>
<tr>
<td></td>
<td>R6</td>
</tr>
<tr>
<td></td>
<td>R7</td>
</tr>
<tr>
<td></td>
<td>R8</td>
</tr>
<tr>
<td></td>
<td>R9</td>
</tr>
<tr>
<td></td>
<td>R10</td>
</tr>
</tbody>
</table>

− KernelInterface →

<table>
<thead>
<tr>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
<th>R8</th>
<th>R9</th>
<th>R10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KIP base address</td>
<td>API Version</td>
<td>API Flags</td>
<td>Kernel ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For this system-call, all registers other than the output registers are preserved. The tlbia instruction encoding is 0x7c0002e4.

---

**EXCHANGEREGISTERS [Systemcall]**

<table>
<thead>
<tr>
<th>UTCB</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R3</td>
</tr>
<tr>
<td></td>
<td>dest</td>
</tr>
<tr>
<td></td>
<td>control</td>
</tr>
<tr>
<td></td>
<td>SP</td>
</tr>
<tr>
<td></td>
<td>IP</td>
</tr>
<tr>
<td></td>
<td>FLAGS</td>
</tr>
<tr>
<td></td>
<td>UserDefinedHandle</td>
</tr>
<tr>
<td></td>
<td>pager</td>
</tr>
<tr>
<td></td>
<td>R8</td>
</tr>
<tr>
<td></td>
<td>R9</td>
</tr>
<tr>
<td></td>
<td>R10</td>
</tr>
</tbody>
</table>

− Exchange Registers →

<table>
<thead>
<tr>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
<th>R8</th>
<th>R9</th>
<th>R10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>result</td>
<td>control</td>
<td>SP</td>
<td>IP</td>
<td>FLAGS</td>
<td>UserDefinedHandle</td>
<td>pager</td>
<td></td>
</tr>
</tbody>
</table>

“FLAGS” refers to the user-modifiable PowerPC processor flags that are held in the MSR register. See the PowerPC Processor Mirroring section (page 125).

---

**THREADCONTROL [Privileged Systemcall]**

<table>
<thead>
<tr>
<th>UTCB</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R3</td>
</tr>
<tr>
<td></td>
<td>dest</td>
</tr>
<tr>
<td></td>
<td>SpaceSpecifier</td>
</tr>
<tr>
<td></td>
<td>Scheduler</td>
</tr>
<tr>
<td></td>
<td>Pager</td>
</tr>
<tr>
<td></td>
<td>UtcbLocation</td>
</tr>
<tr>
<td></td>
<td>R7</td>
</tr>
<tr>
<td></td>
<td>R8</td>
</tr>
<tr>
<td></td>
<td>R9</td>
</tr>
<tr>
<td></td>
<td>R10</td>
</tr>
</tbody>
</table>

− Thread Control →

<table>
<thead>
<tr>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
<th>R8</th>
<th>R9</th>
<th>R10</th>
</tr>
</thead>
<tbody>
<tr>
<td>result</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td></td>
</tr>
</tbody>
</table>
**SYSTEMCALLS**

---

**SYSTEMCLOCK** [Systemcall]

\[
\begin{array}{c|c}
\text{UTCB} & \text{R2} \\
\hline
\text{– R3} & \text{– R5} \\
\text{– R4} & \text{– R6} \\
\text{– R7} & \text{– R8} \\
\text{– R9} & \text{– R10} \\
\end{array}
\]

\[\text{– SystemClock} \rightarrow \text{R2} \equiv\]
\[\text{R1} \equiv \text{clock 0 . . . 31} \]
\[\text{R4} \equiv \text{clock 32 . . . 63} \]
\[\text{R5} \sim \]
\[\text{R6} \sim \]
\[\text{R7} \sim \]
\[\text{R8} \sim \]
\[\text{R9} \sim \]
\[\text{R10} \sim \]

---

**THREADSWITCH** [Systemcall]

\[
\begin{array}{c|c}
\text{UTCB} & \text{R2} \\
\hline
\text{dest} & \text{R3} \\
\text{– R4} & \text{– R5} \\
\text{– R6} & \text{– R7} \\
\text{– R8} & \text{– R9} \\
\text{– R10} & \text{– R10} \\
\end{array}
\]

\[\text{– ThreadSwitch} \rightarrow \text{R2} \equiv\]
\[\text{R2} \sim \]
\[\text{R3} \sim \]
\[\text{R4} \sim \]
\[\text{R5} \sim \]
\[\text{R6} \sim \]
\[\text{R7} \sim \]
\[\text{R8} \sim \]
\[\text{R9} \sim \]
\[\text{R10} \sim \]

---

**SCHEDULE** [Systemcall]

\[
\begin{array}{c|c}
\text{UTCB} & \text{R2} \\
\hline
\text{dest} & \text{R3} \\
\text{time control} & \text{R4} \\
\text{processor control} & \text{R5} \\
\text{prio} & \text{R6} \\
\text{preemption control} & \text{R7} \\
\text{– R8} & \text{– R9} \\
\text{– R10} & \text{– R10} \\
\end{array}
\]

\[\text{– Schedule} \rightarrow \text{R2} \equiv\]
\[\text{R2} \equiv \]
\[\text{R3} \equiv \text{result} \]
\[\text{R4} \equiv \text{time control} \]
\[\text{R5} \equiv \text{processor control} \]
\[\text{R6} \equiv \text{prio} \]
\[\text{R7} \equiv \text{preemption control} \]
\[\text{R8} \sim \]
\[\text{R9} \sim \]
\[\text{R10} \sim \]

---

**IPC** [Systemcall]

\[
\begin{array}{c|c}
\text{MR} & \text{R0} \\
\hline
\text{– R1} & \text{– R2} \\
\text{UTCB} & \text{R2} \\
\text{MR} & \text{R3} \\
\text{MR} & \text{R4} \\
\text{MR} & \text{R5} \\
\text{MR} & \text{R6} \\
\text{MR} & \text{R7} \\
\text{MR} & \text{R8} \\
\text{MR} & \text{R9} \\
\text{MR} & \text{R10} \\
\text{– R11} & \text{– R12} \\
\text{– R13} & \text{– R14} \\
\text{– R15} & \text{to R15} \\
\text{FromSpecifier} & \text{R16} \\
\text{Timeouts} & \text{R17} \\
\end{array}
\]

\[\text{– Ipc} \rightarrow \text{R0} \equiv \text{MR 9} \]
\[\text{R1} \equiv \text{MR 1} \]
\[\text{R2} \equiv \text{MR 2} \]
\[\text{R3} \equiv \text{MR 3} \]
\[\text{R4} \equiv \text{MR 4} \]
\[\text{R5} \equiv \text{MR 5} \]
\[\text{R6} \equiv \text{MR 6} \]
\[\text{R7} \equiv \text{MR 7} \]
\[\text{R8} \equiv \text{MR 8} \]
\[\text{R9} \equiv \text{MR 9} \]
\[\text{R10} \equiv \text{MR 10} \]
\[\text{R11} \equiv \text{MR 11} \]
\[\text{R12} \equiv \text{MR 12} \]
\[\text{R13} \equiv \text{MR 13} \]
\[\text{R14} \equiv \text{MR 14} \]
\[\text{R15} \equiv \text{MR 15} \]
\[\text{R16} \equiv \text{from} \]
\[\text{R17} \equiv \text{from} \]
LIPC  [Systemcall]

MR_9 R0
− R1
UTCB R2
MR_2 R3
MR_5 R4
MR_4 R5
MR_6 R6
MR_7 R7
MR_8 R8
− R9
− R10
MR_0 R14
to R15
FromSpecifier R16
Timeouts R17

− Lipc →
call Lipc

MR_0 R0
− R1
UTCB R2
MR_1 R3
MR_2 R4
MR_5 R5
MR_4 R6
MR_6 R7
MR_7 R8
MR_8 R9
− R10
− R11
− R12
− R13
MR_0 R14
to R15
FromSpecifier R16
Timeouts R17

R0  MR_9
R1  ≡
R2  ≡
R3  MR_1
R4  MR_2
R5  MR_3
R6  MR_4
R7  MR_5
R8  MR_6
R9  MR_7
R10 MR_8
R11 ～
R12 ～
R13 ～
R14 MR_0
R15 ～
R16 from
R17 ～

UNMAP  [Systemcall]

UTCB R2
− R3
control R4
− R5
− R6
− R7
− R8
− R9
− R10

− Unmap →
call Unmap

R2  ≡
R3 ～
R4 ～
R5 ～
R6 ～
R7 ～
R8 ～
R9 ～
R10 ～

SPACECONTROL  [Privileged Systemcall]

UTCB R2
− R3
SpaceSpecifier R4
control R5
KernelInterfacePageArea R6
UtchArea R7
Redirector R8
− R9
− R10

− Space Control →
call SpaceControl

R2  ≡
R3 result
R4 ～
R5 ～
R6 ～
R7 ～
R8 ～
R9 ～
R10 ～

PROCESSORCONTROL  [Privileged Systemcall]

UTCB R2
− R3
processor no R4
InternalFreq R5
ExternalFreq R6
voltage R7
− R8
− R9
− R10

− Processor Control →
call ProcessorControl

R2  ≡
R3 result
R4 ～
R5 ～
R6 ～
R7 ～
R8 ～
R9 ～
R10 ～
**MEMORYCONTROL**  [Privileged Systemcall]

<table>
<thead>
<tr>
<th>UTCB</th>
<th>R2</th>
<th>Memory Control →</th>
<th>R2</th>
<th>≃</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>R3</td>
<td></td>
<td>R3</td>
<td>~</td>
</tr>
<tr>
<td>attribute₀</td>
<td>R4</td>
<td></td>
<td>R4</td>
<td>~</td>
</tr>
<tr>
<td>attribute₁</td>
<td>R5</td>
<td>call MemoryControl</td>
<td>R5</td>
<td>~</td>
</tr>
<tr>
<td>attribute₂</td>
<td>R6</td>
<td></td>
<td>R6</td>
<td>~</td>
</tr>
<tr>
<td>attribute₃</td>
<td>R7</td>
<td></td>
<td>R7</td>
<td>~</td>
</tr>
<tr>
<td>−</td>
<td>R8</td>
<td></td>
<td>R8</td>
<td>~</td>
</tr>
<tr>
<td>−</td>
<td>R9</td>
<td></td>
<td>R9</td>
<td>~</td>
</tr>
<tr>
<td>−</td>
<td>R10</td>
<td></td>
<td>R10</td>
<td>~</td>
</tr>
</tbody>
</table>
C.3 Memory Attributes [powerpc]

The PowerPC architecture supports the following memory/cache attribute values, to be used with the MEMORYCONTROL system-call:

<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0</td>
</tr>
<tr>
<td>Write-through</td>
<td>1</td>
</tr>
<tr>
<td>Write-back</td>
<td>2</td>
</tr>
<tr>
<td>Caching-inhibited</td>
<td>3</td>
</tr>
<tr>
<td>Caching-enabled</td>
<td>4</td>
</tr>
<tr>
<td>Memory-global (coherent)</td>
<td>5</td>
</tr>
<tr>
<td>Memory-local (not coherent)</td>
<td>6</td>
</tr>
<tr>
<td>Guarded</td>
<td>7</td>
</tr>
<tr>
<td>Speculative</td>
<td>8</td>
</tr>
</tbody>
</table>

The default attributes enable write-back, caching, and speculation. Only if the kernel is compiled with support for multiple processors will memory coherency be enabled by default.

The PowerPC architecture places a variety of restrictions on the usage of the memory/cache attributes. Some combinations are meaningless (such as combining write-through with caching-inhibited), or are not permitted and will lead to undefined behavior (for example, instruction fetching is incompatible with some combinations of attributes). The precise semantics of the memory/cache access attributes are described in the “Programming Environments Manual For 32-Bit Implementations of the PowerPC Architecture.”

Before disabling the cache for a page, the software must ensure that all memory belonging to the target page is flushed from the cache.

---

**Generic Programming Interface**

```c
#include <4/misc.h>

Word DefaultMemory
Word WriteThroughMemory
Word WriteBackMemory
Word CachingInhibitedMemory
Word CachingEnabledMemory
Word GlobalMemory
Word LocalMemory
Word GuardedMemory
Word SpeculativeMemory
```
### C.4 Exception Message Format [powerpc]

**System Call Trap**

**System Call Trap Message to Exception Handler**

<table>
<thead>
<tr>
<th>Flags (32)</th>
<th>MR_{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP (32)</td>
<td>MR_{11}</td>
</tr>
<tr>
<td>IP (32)</td>
<td>MR_{10}</td>
</tr>
<tr>
<td>R0 (32)</td>
<td>MR_{9}</td>
</tr>
<tr>
<td>R10 (32)</td>
<td>MR_{8}</td>
</tr>
<tr>
<td>R9 (32)</td>
<td>MR_{7}</td>
</tr>
<tr>
<td>R8 (32)</td>
<td>MR_{6}</td>
</tr>
<tr>
<td>R7 (32)</td>
<td>MR_{5}</td>
</tr>
<tr>
<td>R6 (32)</td>
<td>MR_{4}</td>
</tr>
<tr>
<td>R5 (32)</td>
<td>MR_{3}</td>
</tr>
<tr>
<td>R4 (32)</td>
<td>MR_{2}</td>
</tr>
<tr>
<td>R3 (32)</td>
<td>MR_{1}</td>
</tr>
</tbody>
</table>

| -6 (16/48) | 0 (4)          | t = 0 (6) | u = 12 (6) | MR_{0} |

When user code executes the PowerPC `sc` instruction, the kernel delivers the system call trap message to the exception handler. The kernel preserves only partial user state when handling an `sc` instruction. State is preserved according to the SVR4 PowerPC ABI for function calls. The non-volatile registers are R1, R2, R13…R31, CR2, CR3, CR4, and FPSCR. The volatile registers are R0, R3…R12, CR0, CR1, CR5…CR7, LR, CTR, and XER.

**Generic Traps**

**Generic Trap Message To Exception Handler**
Some traps are handled by the kernel and therefore do not generate exception messages.
C.5 Processor Mirroring [powerpc]

The kernel will expose the following supervisor instructions to all user level programs via emulation: MFSPR for the PVR, MFSPR and MTSPR for the DABR and other cpu-specific debug registers.

The kernel will emulate the MFSPR and MTSPR instructions for accessing cpu-specific performance monitor registers on behalf of privileged tasks. The performance monitor registers are global, and not per-thread.

The EXCHANGEREGISTERS system-call accesses the flags of the processor. The flags map directly to the PowerPC MSR register. The following bits may be read and modified by user applications: LE, BE, SE, FE0, and FE1. The kernel also exposes additional cpu-specific bits.
C.6 Booting  [powerpc]

Apple New World Compatible Machines

L4 must be loaded into memory at the physical location defined by the kernel’s ELF header. It can be started with virtual addressing enabled or disabled. Execution of L4 must begin at the entry point defined by the kernel’s ELF header.

When entering the kernel, the registers which support in-register file parameter passing, R3–R10 according to the SVR4 ABI, must be cleared for upwards compatibility, except as noted below. All other registers in the register file are undefined at kernel entry.

The kernel may use OpenFirmware for debug console I/O. To support OpenFirmware I/O, the OpenFirmware virtual mode client call-back address must be passed to the kernel in register R5, and OpenFirmware must be prepared to handle client call-backs using virtual addressing. In all other cases, register R5 must be zero.

The boot loader must copy the OpenFirmware device tree to memory, and record its physical location in a memory descriptor of the kernel interface page. The copy of the device tree must include the package handles of the device tree nodes.
D.1 Virtual Registers

Thread Control Registers (TCRs)

TCRs are mapped to memory locations. They are implemented as part of the Alpha-specific user-level thread control block (UTCB). The address of the current thread’s UTCB is identical to the thread’s local ID, and is thus immutable. The UTCB (and hence local ID) is available through the rdunique PAL call. UTCB objects of the current thread can be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible.

<table>
<thead>
<tr>
<th>Address Offset</th>
<th>Field Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>+88</td>
<td>ThreadWord1 (64)</td>
</tr>
<tr>
<td>+80</td>
<td>ThreadWord0 (64)</td>
</tr>
<tr>
<td>+72</td>
<td>VirtualSender/ActualSender (64)</td>
</tr>
<tr>
<td>+64</td>
<td>IntendedReceiver (64)</td>
</tr>
<tr>
<td>+56</td>
<td>ErrorCode (64)</td>
</tr>
<tr>
<td>+48</td>
<td>XferTimeouts (64)</td>
</tr>
<tr>
<td>+40</td>
<td>~ (48)</td>
</tr>
<tr>
<td></td>
<td>cop flags (8)</td>
</tr>
<tr>
<td></td>
<td>preempt flags (8)</td>
</tr>
<tr>
<td>+32</td>
<td>ExceptionHandler (64)</td>
</tr>
<tr>
<td>+24</td>
<td>Pager (64)</td>
</tr>
<tr>
<td>+16</td>
<td>UserDefinedHandle (64)</td>
</tr>
<tr>
<td>+8</td>
<td>ProcessorNo (64)</td>
</tr>
<tr>
<td></td>
<td>MyGlobalId (64)</td>
</tr>
<tr>
<td></td>
<td>MyLocalId = UTCB address (64)</td>
</tr>
</tbody>
</table>

The TCR MyLocalId is not part of the UTCB. On Alpha it is identical with the UTCB address and can be found using the rdunique PAL call.

Message Registers (MRs)

Message registers MR_0 through MR_8 map to the processor’s general purpose register file for IPC and LIPC calls. The remaining message registers map to memory locations in the UTCB. MR_9 starts at byte offset 200 in the UTCB, and successive message registers follow in memory.

For the other system calls, message registers map to memory locations in the UTCB, with MR_0 starting at byte offset 128.
### Virtual Registers

<table>
<thead>
<tr>
<th>MR 0...8</th>
<th>s5</th>
<th>s4</th>
<th>s3</th>
<th>s2</th>
<th>s1</th>
<th>s0</th>
<th>t7</th>
<th>t6</th>
<th>s6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MR 9...63 [UTCB fields]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR 63 (64) +632</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>MR 12 (64) +224</td>
</tr>
<tr>
<td>MR 11 (64) +216</td>
</tr>
<tr>
<td>MR 10 (64) +208</td>
</tr>
<tr>
<td>MR 9 (64) ⤷ UTCB address + 200</td>
</tr>
</tbody>
</table>

### Buffer Registers (BRs)

The buffer registers map to memory locations in the UTCB. BR0 is at byte offset 640 in the UTCB, BR1 at byte offset 648, etc.

<table>
<thead>
<tr>
<th>BR 0...32 [UTCB fields]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR 32 (64) +896</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>BR 1 (64) +648</td>
</tr>
<tr>
<td>BR 0 (64) ⤷ UTCB address + 640</td>
</tr>
</tbody>
</table>

### UTCB Memory With Undefined Semantics

The kernel will associate no semantics with memory located at UTCB address + 128...UTCB address + 199. The application can use this memory as thread local storage, e.g., for implementing the L4 API. Note, however, that the memory contents within this region may be overwritten during a system-call operating on message registers.

All undefined UTCB memory which is not covered by the above mentioned region may have kernel defined semantics.
D.2 Systemcalls

The system-calls invoked via the ‘jsr’ instruction are located in the kernel’s area of the virtual address space. Their precise locations are stored in the kernel interface page (see page 2). One may invoke the system calls with any instruction that branches to the appropriate target, as long as the return-address register (RA) contains the correct return address.

The locations of the system-calls are fixed during the life of an application, although they may change outside of the life of an application. It is not valid to prelink an application against a set of system call locations. The official locations are always provided in the kip.

Unless explicitly stated, the kernel follows the Alpha calling convention for the system call interface. This means that arguments are passed in the a0 – a5 registers and the result is placed in the v0 register. All ‘s’ registers are preserved and all ‘t’ registers are undefined. The sp and ra registers are also preserved.

All floating point registers are preserved across a system call.

All other registers contain return values, are undefined, or may be preserved according to processor specific rules.

<table>
<thead>
<tr>
<th>KERNELINTERFACE</th>
<th>[Slow Systemcall]</th>
</tr>
</thead>
<tbody>
<tr>
<td>- v0</td>
<td>- KernelInterface →</td>
</tr>
<tr>
<td>0x4c:34754b4b495034</td>
<td>call_pal cserve</td>
</tr>
<tr>
<td>- a0</td>
<td>- v0 KIP base address</td>
</tr>
<tr>
<td>- a1</td>
<td>a0 API Version</td>
</tr>
<tr>
<td>- a2</td>
<td>a1 API Flags</td>
</tr>
<tr>
<td>- a3</td>
<td>a2 Kernel ID</td>
</tr>
<tr>
<td>- a4</td>
<td>a3 ~</td>
</tr>
<tr>
<td>- a5</td>
<td>a4 ~</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXCHANGEREGISTERS</th>
<th>[Systemcall]</th>
</tr>
</thead>
<tbody>
<tr>
<td>- v0</td>
<td>- Exchange Registers →</td>
</tr>
<tr>
<td>dest</td>
<td>jsr ra, ExchangeRegisters</td>
</tr>
<tr>
<td>control</td>
<td>- Exchange Registers →</td>
</tr>
<tr>
<td>SP</td>
<td>- result</td>
</tr>
<tr>
<td>IP</td>
<td>a0 control</td>
</tr>
<tr>
<td>FLAGS</td>
<td>a1 SP</td>
</tr>
<tr>
<td>UserDefinedHandle</td>
<td>a2 IP</td>
</tr>
<tr>
<td>pager</td>
<td>a3 FLAGS</td>
</tr>
<tr>
<td>t1</td>
<td>a4 UserDefinedHandle</td>
</tr>
<tr>
<td></td>
<td>a5 pager</td>
</tr>
<tr>
<td></td>
<td>t1 ~</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>THREADCONTROL</th>
<th>[Privileged Systemcall]</th>
</tr>
</thead>
<tbody>
<tr>
<td>- v0</td>
<td>- Thread Control →</td>
</tr>
<tr>
<td>dest</td>
<td>jsr ra, ThreadControl</td>
</tr>
<tr>
<td>SpaceSpecifier</td>
<td>- Thread Control →</td>
</tr>
<tr>
<td>Scheduler</td>
<td>- v0 result</td>
</tr>
<tr>
<td>Pager</td>
<td>a0 ~</td>
</tr>
<tr>
<td>UtcbLocation</td>
<td>a1 ~</td>
</tr>
<tr>
<td>- a2</td>
<td>a2 ~</td>
</tr>
<tr>
<td>- a3</td>
<td>a3 ~</td>
</tr>
<tr>
<td>- a4</td>
<td>a4 ~</td>
</tr>
<tr>
<td>- a5</td>
<td>a5 ~</td>
</tr>
</tbody>
</table>
SYSTEMCALLS

SYSTEMCLOCK  [Systemcall]

<table>
<thead>
<tr>
<th>v0</th>
<th>Clock</th>
</tr>
</thead>
<tbody>
<tr>
<td>a0</td>
<td></td>
</tr>
<tr>
<td>a1</td>
<td></td>
</tr>
<tr>
<td>a2</td>
<td></td>
</tr>
<tr>
<td>a3</td>
<td></td>
</tr>
<tr>
<td>a4</td>
<td></td>
</tr>
<tr>
<td>a5</td>
<td></td>
</tr>
</tbody>
</table>

- SystemClock → jsr ra, SystemClock
- SystemClock → a0 ~
- SystemClock → a1 ~
- SystemClock → a2 ~
- SystemClock → a3 ~
- SystemClock → a4 ~
- SystemClock → a5 ~

Note that the SystemClock system call is currently UNIMPLEMENTED on Alpha.

THREADSWITCH  [Systemcall]

<table>
<thead>
<tr>
<th>v0</th>
<th>Dest</th>
</tr>
</thead>
<tbody>
<tr>
<td>a0</td>
<td></td>
</tr>
<tr>
<td>a1</td>
<td></td>
</tr>
<tr>
<td>a2</td>
<td></td>
</tr>
<tr>
<td>a3</td>
<td></td>
</tr>
<tr>
<td>a4</td>
<td></td>
</tr>
<tr>
<td>a5</td>
<td></td>
</tr>
</tbody>
</table>

- ThreadSwitch → jsr ra, ThreadSwitch
- ThreadSwitch → a0 ~
- ThreadSwitch → a1 ~
- ThreadSwitch → a2 ~
- ThreadSwitch → a3 ~
- ThreadSwitch → a4 ~
- ThreadSwitch → a5 ~

SCHEDULE  [Systemcall]

<table>
<thead>
<tr>
<th>v0</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>a0</td>
<td>TimeControl</td>
</tr>
<tr>
<td>a1</td>
<td></td>
</tr>
<tr>
<td>a2</td>
<td></td>
</tr>
<tr>
<td>a3</td>
<td></td>
</tr>
<tr>
<td>a4</td>
<td></td>
</tr>
<tr>
<td>a5</td>
<td></td>
</tr>
</tbody>
</table>

- Schedule → jsr ra, Schedule
- Schedule → a0 ~
- Schedule → a1 ~
- Schedule → a2 ~
- Schedule → a3 ~
- Schedule → a4 ~
- Schedule → a5 ~

IPC  [Systemcall]

<table>
<thead>
<tr>
<th>v0</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>a0</td>
<td>T5</td>
</tr>
<tr>
<td>a1</td>
<td>MR1</td>
</tr>
<tr>
<td>a2</td>
<td>MR2</td>
</tr>
<tr>
<td>a3</td>
<td>MR3</td>
</tr>
<tr>
<td>a4</td>
<td>MR4</td>
</tr>
<tr>
<td>a5</td>
<td>MR5</td>
</tr>
<tr>
<td>s6</td>
<td>MR0</td>
</tr>
<tr>
<td>s7</td>
<td>MR1</td>
</tr>
<tr>
<td>s8</td>
<td>MR2</td>
</tr>
<tr>
<td>s9</td>
<td>MR3</td>
</tr>
<tr>
<td>s10</td>
<td>MR4</td>
</tr>
<tr>
<td>s11</td>
<td>MR5</td>
</tr>
<tr>
<td>s12</td>
<td>MR6</td>
</tr>
<tr>
<td>s13</td>
<td>MR7</td>
</tr>
<tr>
<td>s14</td>
<td>MR8</td>
</tr>
</tbody>
</table>

- Ipc → jsr ra, Ipc
- Ipc → a0 ~
- Ipc → a1 ~
- Ipc → a2 ~
- Ipc → a3 ~
- Ipc → a4 ~
- Ipc → a5 ~
- Ipc → s6 MR0
- Ipc → s7 MR1
- Ipc → s8 MR2
- Ipc → s9 MR3
- Ipc → s10 MR4
- Ipc → s11 MR5
- Ipc → s12 MR6
- Ipc → s13 MR7
- Ipc → s14 MR8
### LIpc [Systemcall]

<table>
<thead>
<tr>
<th>dest a0</th>
<th>source a1</th>
<th>timeout a2</th>
</tr>
</thead>
<tbody>
<tr>
<td>v0</td>
<td>a0</td>
<td>a1</td>
</tr>
<tr>
<td>v0</td>
<td>a0</td>
<td>a1</td>
</tr>
<tr>
<td>v0</td>
<td>a2</td>
<td>a2</td>
</tr>
<tr>
<td>v0</td>
<td>a3</td>
<td>a3</td>
</tr>
<tr>
<td>v0</td>
<td>a4</td>
<td>a4</td>
</tr>
<tr>
<td>v0</td>
<td>a5</td>
<td>a5</td>
</tr>
</tbody>
</table>

- **Lipc** → 
  - `jsr ra, Lipc`

Note that on Alpha LIpc is not implemented: use IPC instead.

### UNmap [Systemcall]

<table>
<thead>
<tr>
<th>control a0</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
</tr>
</thead>
<tbody>
<tr>
<td>v0</td>
<td>a0</td>
<td>a1</td>
<td>a2</td>
<td>a3</td>
<td>a4</td>
</tr>
</tbody>
</table>

- **Unmap** → 
  - `jsr ra, Unmap`

### SpaceControl [Privileged Systemcall]

| SpaceSpecifier a0 | control a1 | a2 | a3 | a4 | a5 |
|SpaceControl a0 | control a1 | a2 | a3 | a4 | a5 |

- **Space Control** → 
  - `jsr ra, SpaceControl`

### ProcessorControl [Privileged Systemcall]

<table>
<thead>
<tr>
<th>ProcessorNo a0</th>
<th>control a1</th>
<th>InternalFreq. a2</th>
<th>ExternalFreq. a3</th>
<th>voltage a4</th>
<th>a5</th>
</tr>
</thead>
<tbody>
<tr>
<td>v0</td>
<td>a0</td>
<td>a1</td>
<td>a2</td>
<td>a3</td>
<td>a4</td>
</tr>
</tbody>
</table>

- **Processor Control** → 
  - `jsr ra, ProcessorControl`

Note that on Alpha the ProcessorControl system call is not implemented.
MEMORY\textsc{CONTROL} \hspace{1cm} [Privileged Systemcall]

\begin{tabular}{|c|c|}
\hline
control & v0  \\
attribute\textsc{0} & a0  \\
attribute\textsc{1} & a1  \\
attribute\textsc{2} & a2  \\
attribute\textsc{3} & a3  \\
\hline
\end{tabular}

\begin{tabular}{|c|c|}
\hline
\multicolumn{2}{|c|}{\textit{− Memory Control} $\rightarrow$ $\text{v0 result}$}  \\
\multicolumn{2}{|c|}{\hspace{1cm} $\text{jst ra, MemoryControl}$}  \\
\multicolumn{2}{|c|}{a0 $\sim$}  \\
\multicolumn{2}{|c|}{a1 $\sim$}  \\
\multicolumn{2}{|c|}{a2 $\sim$}  \\
\multicolumn{2}{|c|}{a3 $\sim$}  \\
\multicolumn{2}{|c|}{a4 $\sim$}  \\
\multicolumn{2}{|c|}{a5 $\sim$}  \\
\hline
\end{tabular}

Note that on Alpha the MemoryControl system call is not implemented: the memory attributes for a page are defined by the system, and cannot be controlled by the application (or kernel).
D.3 Booting [alpha]

All SRM based machines

L4 must be loaded at the virtual address defined in the ELF header (corresponding to the physical region of the virtual address space). The kernel also requires the bootloader to initialise some kernel data structures, so the supplied bootloader is recommended.

The preferred method for booting the kernel is via BootP. Consult the SRM documentation for instructions on setting up SRM to boot a file from a remote host.
Appendix E

MIPS64 Interface
E.1 Virtual Registers

Thread Control Registers (TCRs)

TCRs are mapped to memory locations. They are implemented as part of the Mips64-specific user-level thread control block (UTCB). The address of the current thread’s UTCB is identical to the thread’s local ID, and is thus immutable. The UTCB (and hence local ID) is available through the break instruction. UTCB objects of the current thread can be accessed as any other memory object. UTCBs of other threads must not be accessed, even if they are physically accessible.

<table>
<thead>
<tr>
<th>Field</th>
<th>Offset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreadWord 1</td>
<td>+88</td>
<td></td>
</tr>
<tr>
<td>ThreadWord 0</td>
<td>+80</td>
<td></td>
</tr>
<tr>
<td>VirtualSender/ActualSender</td>
<td>+72</td>
<td></td>
</tr>
<tr>
<td>IntendedReceiver</td>
<td>+64</td>
<td></td>
</tr>
<tr>
<td>ErrorCode</td>
<td>+56</td>
<td></td>
</tr>
<tr>
<td>XferTimeouts</td>
<td>+48</td>
<td></td>
</tr>
<tr>
<td>~ (48)</td>
<td>+40</td>
<td>cop flags (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>preempt flags (8)</td>
</tr>
<tr>
<td>ExceptionHandler</td>
<td>+32</td>
<td></td>
</tr>
<tr>
<td>Pager</td>
<td>+24</td>
<td></td>
</tr>
<tr>
<td>UserDefinedHandle</td>
<td>+16</td>
<td></td>
</tr>
<tr>
<td>ProcessorNo</td>
<td>+8</td>
<td></td>
</tr>
<tr>
<td>MyGlobalId</td>
<td>←−</td>
<td>UTCB address</td>
</tr>
</tbody>
</table>

MyLocalId = UTCB address

The TCR MyLocalId is not part of the UTCB. On Mips64 it is identical with the UTCB address and can be found using the utcb syscall. The utcb syscall is a break instruction with an argument of 3 in the AT register. Result is in v0. This is a preliminary solution and going to be changed using a faster access method such as a general purpose register.

Message Registers (MRs)

Message registers MR_0 through MR_7 map to the processor’s general purpose register file for IPC and LIPC calls. The remaining message registers map to memory locations in the UTCB. MR_8 starts at byte offset 192 in the UTCB, and successive message registers follow in memory.

The first eight message registers are mapped to the registers s0 to s7. MR_8...63 are mapped to memory in the UTCB.
### Virtual Registers

<table>
<thead>
<tr>
<th>MR 0...7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MR 0</td>
<td>s0</td>
</tr>
<tr>
<td>MR 1</td>
<td>s1</td>
</tr>
<tr>
<td>MR 2</td>
<td>s2</td>
</tr>
<tr>
<td>MR 3</td>
<td>s3</td>
</tr>
<tr>
<td>MR 4</td>
<td>s4</td>
</tr>
<tr>
<td>MR 5</td>
<td>s5</td>
</tr>
<tr>
<td>MR 6</td>
<td>s6</td>
</tr>
<tr>
<td>MR 7</td>
<td>s7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MR 0...63 [UTCB fields]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MR 63 (64)</td>
<td>+632</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>MR 8 (64)</td>
<td>← UTCB address + 192</td>
</tr>
</tbody>
</table>

### Buffer Registers (BRs)

The buffer registers map to memory locations in the UTCB. BR 0 is at byte offset 640 in the UTCB, BR 1 at byte offset 648, etc.

<table>
<thead>
<tr>
<th>BR 0...32 [UTCB fields]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BR 32 (64)</td>
<td>+896</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>BR 1 (64)</td>
<td>+648</td>
</tr>
<tr>
<td>BR 0 (64)</td>
<td>← UTCB address + 640</td>
</tr>
</tbody>
</table>

### UTCB Memory With Undefined Semantics

The kernel will associate no semantics with memory located at UTCB address + 128...UTCB address + 191. The application can use this memory as thread local storage, e.g., for implementing the L4 API. Note, however, that the memory contents within this region may be overwritten during a system-call operating on message registers.

All undefined UTCB memory which is not covered by the above mentioned region may have kernel defined semantics.
The system-calls invoked via the `jal` instruction are located in the kernel’s area of the virtual address space. Their precise locations are stored in the kernel interface page (see page 2). One may invoke the system calls with any instruction that branches to the appropriate target, as long as the return-address register `RA` contains the correct return address.

The locations of the system-calls are fixed during the life of an application, although they may change outside of the life of an application. It is not valid to prelink an application against a set of system call locations. The official locations are always provided in the KIP.

In general, the kernel follows the MIPS ABI64 calling convention for the system call boundary. This means that arguments are passed in the a0 – a7 registers, and the result is placed in the v0 register. All floating point registers are preserved across a system call. All other registers contain return values, are undefined, or may be preserved according to processor specific rules.

### KERNEL INTERFACE [Slow Systemcall]

<table>
<thead>
<tr>
<th>opcode 0x07FFFFFF</th>
<th>– KernelInterface →</th>
</tr>
</thead>
<tbody>
<tr>
<td>at</td>
<td>at</td>
</tr>
<tr>
<td>v0, v1</td>
<td>v0, v1</td>
</tr>
<tr>
<td>a0...a3</td>
<td>a0...a3</td>
</tr>
<tr>
<td>t1</td>
<td>KIP base address</td>
</tr>
<tr>
<td>t2</td>
<td>API Version</td>
</tr>
<tr>
<td>t3</td>
<td>API Flags</td>
</tr>
<tr>
<td>t4...t7</td>
<td>Kernel ID</td>
</tr>
<tr>
<td>t8, t9</td>
<td></td>
</tr>
<tr>
<td>gp, sp</td>
<td></td>
</tr>
<tr>
<td>t10</td>
<td></td>
</tr>
</tbody>
</table>

For this system-call, all registers other than the output registers are preserved.

### EXCHANGE REGISTERS [Systemcall]

<table>
<thead>
<tr>
<th>– Exchange Registers →</th>
</tr>
</thead>
<tbody>
<tr>
<td>at</td>
</tr>
<tr>
<td>v0</td>
</tr>
<tr>
<td>v1</td>
</tr>
<tr>
<td>a0</td>
</tr>
<tr>
<td>a1</td>
</tr>
<tr>
<td>a2</td>
</tr>
<tr>
<td>a3</td>
</tr>
<tr>
<td>t1</td>
</tr>
<tr>
<td>t2</td>
</tr>
<tr>
<td>t3</td>
</tr>
<tr>
<td>t4...t7</td>
</tr>
<tr>
<td>t8, t9</td>
</tr>
<tr>
<td>gp</td>
</tr>
<tr>
<td>s0...s7</td>
</tr>
<tr>
<td>t8, t9</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>t8</td>
</tr>
<tr>
<td>s8</td>
</tr>
<tr>
<td>ra</td>
</tr>
</tbody>
</table>

For this system-call, all registers other than the output registers are preserved.
### SYSTEMCALLS

#### THREADCONTROL

<table>
<thead>
<tr>
<th>Privileged Systemcall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thread Control →</td>
</tr>
<tr>
<td>at</td>
</tr>
<tr>
<td>v0</td>
</tr>
<tr>
<td>v1</td>
</tr>
<tr>
<td>a0</td>
</tr>
<tr>
<td>a1</td>
</tr>
<tr>
<td>a2</td>
</tr>
<tr>
<td>a3</td>
</tr>
<tr>
<td>a4</td>
</tr>
<tr>
<td>v1...v3</td>
</tr>
<tr>
<td>a4...a7</td>
</tr>
<tr>
<td>a0...a7</td>
</tr>
<tr>
<td>a8, a9</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>s8</td>
</tr>
<tr>
<td>ra</td>
</tr>
</tbody>
</table>

#### SYSTEMCLOCK

<table>
<thead>
<tr>
<th>Systemcall</th>
</tr>
</thead>
<tbody>
<tr>
<td>SystemClock →</td>
</tr>
<tr>
<td>at</td>
</tr>
<tr>
<td>v0</td>
</tr>
<tr>
<td>v1</td>
</tr>
<tr>
<td>a0...a3</td>
</tr>
<tr>
<td>a4...a7</td>
</tr>
<tr>
<td>a4...a7</td>
</tr>
<tr>
<td>a8, a9</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>s8</td>
</tr>
<tr>
<td>ra</td>
</tr>
</tbody>
</table>

#### THREADSWITCH

<table>
<thead>
<tr>
<th>Systemcall</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreadSwitch →</td>
</tr>
<tr>
<td>at</td>
</tr>
<tr>
<td>v0, v1</td>
</tr>
<tr>
<td>a0</td>
</tr>
<tr>
<td>a1...a3</td>
</tr>
<tr>
<td>a4...a7</td>
</tr>
<tr>
<td>a4...a7</td>
</tr>
<tr>
<td>a8, a9</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>sp</td>
</tr>
<tr>
<td>s8</td>
</tr>
<tr>
<td>ra</td>
</tr>
</tbody>
</table>
### Schedule

<table>
<thead>
<tr>
<th>dest</th>
<th>a0</th>
<th>time control</th>
<th>a1</th>
<th>priority</th>
<th>a2</th>
<th>preemption control</th>
<th>a3</th>
</tr>
</thead>
<tbody>
<tr>
<td></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>
</tr>
<tr>
<td></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>
</tr>
</tbody>
</table>

\[
\text{jal Schedule} \rightarrow \quad a0 \sim v0 \quad \text{result} \quad v1 \sim a0 \sim \text{time control} \quad a1 \sim a2 \sim a3 \sim a4 \sim a5 \ldots a7 \sim a8 \ldots a17 \sim t0 \sim t1 \ldots t3 \sim t4 \ldots t7 \sim s0 \ldots s17 \sim \text{gp} \sim \text{sp} \equiv s8 \equiv ra \sim
\]

### IPC

<table>
<thead>
<tr>
<th>to</th>
<th>a0</th>
<th>FromSpecifier</th>
<th>a1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Timeouts</td>
<td>a2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Ipc} \rightarrow \quad a0 \sim v0 \quad \text{result} \quad v1 \sim a0 \sim a1 \sim a2 \sim a3 \sim a4 \ldots a7 \sim a8 \ldots a17 \sim s0 \sim s1 \sim s2 \sim s3 \sim s4 \sim s5 \sim s6 \sim s7 \sim \text{gp} \sim \text{sp} \equiv s8 \equiv ra \sim
\]
### System Calls

#### LPC [Systemcall]

- **FromSpecifier**
  - a1
  - a2
  - a3
  - v0
  - v1

- **Timeouts**
  - a0

- **To**
  - a0

#### Lipc [Systemcall]

- jal Lipc
- a0 ~
- a1 ~
- a2 ~
- a3 ~
- a4...a7 ~
- a8...a9 ~

#### Unmap [Systemcall]

- **control**
  - a0
  - a1...a3
  - a4...a7
  - v0, v1

- jal Unmap
- a0 ~
- a1...a3 ~
- a4...a7 ~
- a8...a9 ~

#### SpaceControl [Privileged Systemcall]

- **SpaceSpecifier**
  - a1
  - a2
  - a3

- **control**
  - a0

- **Redirector**
  - a0...a3

- **KernelInterfacePageArea**
  - a1...a3

- **UserArea**
  - a4...a7

- **t1...t3**
- **t4...t7**
- **a8...a9**
- **gp**
- **sp**
- **ra**

- jal SpaceControl
- a0 ~
- a1 ~
- a2 ~
- a3 ~
- a4 ~
- a5...a7 ~
- a8...a9 ~
- gp ~
- sp ~
- ra ~
- s8 ~
- s9 ~
- t0..t3
- t4..t7
- s0...s7
- t8, t9
- gp
- sp
- ra
### PROCESSOR CONTROL

- `at`
- `v0`
- `v1`
- `a0`
- `a1`
- `a2`
- `a3`
- `t0`...`t3`
- `t4`...`t7`
- `s0`...`s7`
- `a4`...`a7`
- `gp`
- `sp`
- `ra`

---

### MEMORY CONTROL

- `at`
- `v0`
- `v1`
- `a0`
- `a1`
- `a2`
- `a3`
- `t0`...`t3`
- `t4`...`t7`
- `s0`...`s7`
- `a4`...`a7`
- `gp`
- `sp`
- `ra`
E.3 Memory Attributes [mips64]

The Mips64 architecture supports the following memory/cache attribute values, to be used with the MEMORYCONTROL system-call:

<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0</td>
</tr>
<tr>
<td>Uncached</td>
<td>1</td>
</tr>
<tr>
<td>Write-back</td>
<td>2</td>
</tr>
<tr>
<td>Write-through</td>
<td>3</td>
</tr>
<tr>
<td>Write-through (no allocate)</td>
<td>4</td>
</tr>
<tr>
<td>Flush (Cache flush)</td>
<td>31</td>
</tr>
</tbody>
</table>

The default attributes depend on the platform and not all modes are defined for all processors.

Before disabling the cache for a page, the software must ensure that all memory belonging to the target page is flushed from the cache.
E.4 Booting [mips64]

The kernel is provided as an ELF file and must be loaded according to the load addresses defined in the ELF header (corresponding to the physical region of the virtual address space). The kernel must be started in 64bit mode.
Development Remarks

These remarks illuminate the design process from version 2 to version 4.

F.1 Exception Handling

The current model decided upon for exception handling in L4 is to associate an exception handler thread with each thread in the system (see page 66). This model was chosen because it allowed us to handle exceptions generically without introducing any new concepts into the API. It also closely resembles the current page fault handling model.

Another model for exception handling is to use callbacks. Using this model an instruction pointer for a callback function and a pointer to an exception state save area is associated with each thread. Upon catching an exception the kernel stores the cause of the exception into the save area and transfers execution to the exception callback function.

It is evident that the callback model can be faster than the IPC model because the callback model may require only one control transfer into the kernel whereas the IPC model will require at least two. Nevertheless, the IPC model was chosen because it introduces no new mechanisms into the kernel, and we are currently not aware of any real life scenario where the extra performance gain you very much. There exists a challenge to prove these claims wrong. See http://l4hq.org/fun/ for the rules of the challenge.
Table of Procs, Types, and Constants

<table>
<thead>
<tr>
<th>used system call</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>!= (CacheAllocationHint l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>!= (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>!= (MsgTag l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>!= (ThreadId l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>!= (Time l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Acceptor l, r) Acceptor</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Clock l, int r) Clock</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Clock l, Word64 r) Clock</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Fpage f, Word AccessRights) Fpage</td>
<td>none–</td>
</tr>
<tr>
<td>+ (MsgTag l, Word label) MsgTag</td>
<td>none–</td>
</tr>
<tr>
<td>+ (StringItem s, CacheAllocationHint h) StringItem</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Time l, r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Time l, Word r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Acceptor l, r) Acceptor</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Fpage f, Word AccessRights) Fpage</td>
<td>none–</td>
</tr>
<tr>
<td>+ (MsgTag l, Word label) MsgTag</td>
<td>none–</td>
</tr>
<tr>
<td>+ (StringItem&amp; dest, StringItem AdditionalSubstring) StringItem &amp;</td>
<td>none–</td>
</tr>
<tr>
<td>+ (StringItem&amp; dest,VOID* AdditionalSubstringAddress) StringItem &amp;</td>
<td>none–</td>
</tr>
<tr>
<td>+ (StringItem s, CacheAllocationHint h) StringItem</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Time l, r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>+ (Time l, Word r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>- (Acceptor l, r) Acceptor</td>
<td>none–</td>
</tr>
<tr>
<td>- (Clock l, int r) Clock</td>
<td>none–</td>
</tr>
<tr>
<td>- (Clock l, Word64 r) Clock</td>
<td>none–</td>
</tr>
<tr>
<td>- (Fpage f, Word AccessRights) Fpage</td>
<td>none–</td>
</tr>
<tr>
<td>- (Time l, r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>- (Time l, Word r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>- (Acceptor l, r) Acceptor</td>
<td>none–</td>
</tr>
<tr>
<td>- (Fpage f, Word AccessRights) Fpage</td>
<td>none–</td>
</tr>
<tr>
<td>- (Time l, r) Time</td>
<td>none–</td>
</tr>
<tr>
<td>&lt; (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&lt; (Time l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&lt;= (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&lt;= (Time l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>== (CacheAllocationHint l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>== (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>== (MsgTag l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>== (ThreadId l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>== (Time l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&gt; (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&gt; (Time l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&gt;= (Clock l, r) bool</td>
<td>none–</td>
</tr>
<tr>
<td>&gt;= (Time l, r) bool</td>
<td>none–</td>
</tr>
</tbody>
</table>

AbortIpc and stop (ThreadId t) ThreadState  EXCHANGEREGISTERS  21
AbortIpc and stop (ThreadId t, Word& sp, ip, flags) ThreadState  EXCHANGEREGISTERS  21
AbortReceive and stop (ThreadId t) ThreadState  EXCHANGEREGISTERS  21
AbortReceive and stop (ThreadId t, Word& sp, ip, flags) ThreadState  EXCHANGEREGISTERS  21
AbortSend and stop (ThreadId t) ThreadState  EXCHANGEREGISTERS  21
<table>
<thead>
<tr>
<th>Function/Type</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbortSend_and_stop (ThreadId t, Word &amp; sp, ip, flags) ThreadState</td>
<td>used system call</td>
<td>21</td>
</tr>
<tr>
<td>Accepted () Acceptor</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>Acceptor data type</td>
<td>-n/a-</td>
<td>55</td>
</tr>
<tr>
<td>Accept (Acceptor a) Void</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>Accept (Acceptor a, MsgBuffer &amp; b) Void</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>ACPIMemoryType Word const</td>
<td>-n/a-</td>
<td>113</td>
</tr>
<tr>
<td>ActualSender () ThreadId</td>
<td>-none-</td>
<td>17</td>
</tr>
<tr>
<td>ActualSender () ThreadId</td>
<td>-none-</td>
<td>63</td>
</tr>
<tr>
<td>Address (Fpage f) Word</td>
<td>-none-</td>
<td>37</td>
</tr>
<tr>
<td>AllocateNewCacheLines CacheAllocationHint const</td>
<td>-n/a-</td>
<td>95</td>
</tr>
<tr>
<td>AllocateOnlyNewL1CacheLines CacheAllocationHint const</td>
<td>-n/a-</td>
<td>95</td>
</tr>
<tr>
<td>anylocalthread ThreadId const</td>
<td>-n/a-</td>
<td>15</td>
</tr>
<tr>
<td>anythread ThreadId const</td>
<td>-n/a-</td>
<td>15</td>
</tr>
<tr>
<td>ApiFlags () Word</td>
<td>-none-</td>
<td>8</td>
</tr>
<tr>
<td>Append (MsgBuffer &amp; b, StringItem * s) Void</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>Append (MsgBuffer &amp; b, StringItem s) Void</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>Append (Msg &amp; msg, GrantItem g) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>Append (Msg &amp; msg, MapItem m) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>Append (Msg &amp; msg, StringItem &amp; s) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>Append (Msg &amp; msg, StringItem s) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>Append (Msg &amp; msg, Word w) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>ArchitectureSpecificMemoryType Word const</td>
<td>-n/a-</td>
<td>9</td>
</tr>
<tr>
<td>AssociateInterrupt (ThreadId InterruptThread, InterruptHandler) Word</td>
<td>-none-</td>
<td>24</td>
</tr>
<tr>
<td>BootInfo (Void* KernelInterface) Word</td>
<td>-none-</td>
<td>9</td>
</tr>
<tr>
<td>BootLoaderSpecificMemoryType Word const</td>
<td>-n/a-</td>
<td>9</td>
</tr>
<tr>
<td>CacheAllocationHint (StringItem s) CacheAllocationHint</td>
<td>-n/a-</td>
<td>54</td>
</tr>
<tr>
<td>CacheAllocationHint data type</td>
<td>-n/a-</td>
<td>53</td>
</tr>
<tr>
<td>CacheNonTemporalAllLevels CacheAllocationHint const</td>
<td>-n/a-</td>
<td>111</td>
</tr>
<tr>
<td>CacheNonTemporalL1 CacheAllocationHint const</td>
<td>-n/a-</td>
<td>111</td>
</tr>
<tr>
<td>CacheNonTemporalL2 CacheAllocationHint const</td>
<td>-n/a-</td>
<td>111</td>
</tr>
<tr>
<td>CachingEnabledMemory Word const</td>
<td>-n/a-</td>
<td>122</td>
</tr>
<tr>
<td>CachingInhibitedMemory Word const</td>
<td>-n/a-</td>
<td>122</td>
</tr>
<tr>
<td>Call (ThreadId to) MsgTag</td>
<td>IPC</td>
<td>61</td>
</tr>
<tr>
<td>Call (ThreadId to, Time SndTimeout, RcvTimeout) MsgTag</td>
<td>IPC</td>
<td>62</td>
</tr>
<tr>
<td>Clear (MsgBuffer &amp; b) Void</td>
<td>-none-</td>
<td>56</td>
</tr>
<tr>
<td>Clear (Msg &amp; msg) Void</td>
<td>-none-</td>
<td>47</td>
</tr>
<tr>
<td>Clock data type</td>
<td>-n/a-</td>
<td>26</td>
</tr>
<tr>
<td>Clr_CopFlag (Word n) Void</td>
<td>-none-</td>
<td>17</td>
</tr>
<tr>
<td>Clr_CopFlag (Word n) Void</td>
<td>-none-</td>
<td>67</td>
</tr>
<tr>
<td>CompleteAddressSpace Fpage const</td>
<td>-n/a-</td>
<td>37</td>
</tr>
<tr>
<td>CompoundString (StringItem&amp; s) bool</td>
<td>-none-</td>
<td>53</td>
</tr>
<tr>
<td>ConventionalMemoryType Word const</td>
<td>-n/a-</td>
<td>9</td>
</tr>
<tr>
<td>DeassociateInterrupt (ThreadId InterruptThread) Word</td>
<td>-none-</td>
<td>24</td>
</tr>
<tr>
<td>DedicatedMemory Type Word const</td>
<td>-n/a-</td>
<td>9</td>
</tr>
<tr>
<td>DefaultMemory Word const</td>
<td>-n/a-</td>
<td>112</td>
</tr>
<tr>
<td>DefaultMemory Word const</td>
<td>-n/a-</td>
<td>122</td>
</tr>
<tr>
<td>DefaultMemory Word const</td>
<td>-n/a-</td>
<td>70</td>
</tr>
<tr>
<td>DefaultMemory Word const</td>
<td>-n/a-</td>
<td>96</td>
</tr>
<tr>
<td>DisablePreemptionFaultException () bool</td>
<td>-none-</td>
<td>34</td>
</tr>
<tr>
<td>DisablePreemption () bool</td>
<td>-none-</td>
<td>34</td>
</tr>
<tr>
<td>DoNotAllocateNewCacheLines CacheAllocationHint const</td>
<td>-n/a-</td>
<td>95</td>
</tr>
<tr>
<td>EnablePreemptionFaultException () bool</td>
<td>-none-</td>
<td>34</td>
</tr>
<tr>
<td>EnablePreemption () bool</td>
<td>-none-</td>
<td>34</td>
</tr>
<tr>
<td>ErrorCode () Word</td>
<td>-none-</td>
<td>17</td>
</tr>
<tr>
<td>ErrorCode () Word</td>
<td>-none-</td>
<td>62</td>
</tr>
<tr>
<td>ExceptionHandler () ThreadId</td>
<td>-none-</td>
<td>17</td>
</tr>
<tr>
<td>ExceptionHandler () ThreadId</td>
<td>-none-</td>
<td>66</td>
</tr>
<tr>
<td>ExchangeRegisters (ThreadId dest, Word control, sp, ip, flags, UserDefinedHandle, ThreadId pager, Word &amp; old_control, old_sp, old_ip, old_flags, old_UserDefinedHandle, ThreadId &amp; old_pager) ThreadId</td>
<td>used system call</td>
<td>20</td>
</tr>
<tr>
<td>eXecutable Word const</td>
<td>-n/a-</td>
<td>37</td>
</tr>
<tr>
<td>ExternalFreq (ProcDesc &amp; p) Word</td>
<td>-none-</td>
<td>10</td>
</tr>
<tr>
<td>Feature</td>
<td>Value</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------</td>
<td>------</td>
</tr>
<tr>
<td>(Void*, KernelInterface, Word num) Char*</td>
<td>–none–</td>
<td>9</td>
</tr>
<tr>
<td>Flush (Fpage f) Void</td>
<td>UNMAP</td>
<td>40</td>
</tr>
<tr>
<td>Flush (Word n, Fpage &amp;) [n] fpages Void</td>
<td>UNMAP</td>
<td>40</td>
</tr>
<tr>
<td>FpageLog2 (Word BaseAddress, int Log2FpageSize &lt; 64) Fpage</td>
<td>–none–</td>
<td>37</td>
</tr>
<tr>
<td>Fpage (Word BaseAddress, int FpageSize ≥ 1K) Fpage</td>
<td>–none–</td>
<td>37</td>
</tr>
<tr>
<td>Fpage data type</td>
<td>–none–</td>
<td>36</td>
</tr>
<tr>
<td>FullyAccessible Word const</td>
<td>–none–</td>
<td>37</td>
</tr>
<tr>
<td>GetStatus (Fpage f) Fpage</td>
<td>–none–</td>
<td>40</td>
</tr>
<tr>
<td>Get (Msg&amp; msg, Word&amp; ut, {MapItem, GrantItem, StringItem} &amp; Items) Void</td>
<td>–none–</td>
<td>47</td>
</tr>
<tr>
<td>Get (Msg&amp; msg, Word t, GrantItem&amp; g) Word</td>
<td>–none–</td>
<td>48</td>
</tr>
<tr>
<td>Get (Msg&amp; msg, Word t, MapItem&amp; m) Word</td>
<td>–none–</td>
<td>48</td>
</tr>
<tr>
<td>Get (Msg&amp; msg, Word t, StringItem&amp; s) Word</td>
<td>–none–</td>
<td>48</td>
</tr>
<tr>
<td>GlobalId (ThreadId t) ThreadId</td>
<td>EXCHANGEREGISTERS</td>
<td>15</td>
</tr>
<tr>
<td>GlobalId (ThreadId t) ThreadId</td>
<td>EXCHANGEREGISTERS</td>
<td>20</td>
</tr>
<tr>
<td>GlobalId (Word threadno, version) ThreadId</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>GlobalMemory Word const</td>
<td>–none–</td>
<td>122</td>
</tr>
<tr>
<td>GrantItem (Fpage f, Word SndBase) GrantItem</td>
<td>–none–</td>
<td>51</td>
</tr>
<tr>
<td>GrantItem (GrantItem g) bool</td>
<td>–none–</td>
<td>51</td>
</tr>
<tr>
<td>GrantItem data type</td>
<td>–none–</td>
<td>51</td>
</tr>
<tr>
<td>GuardedMemory Word const</td>
<td>–none–</td>
<td>122</td>
</tr>
<tr>
<td>High (MemoryDesc&amp; m) Word</td>
<td>–none–</td>
<td>9</td>
</tr>
<tr>
<td>IntendedReceiver () ThreadId</td>
<td>–none–</td>
<td>17</td>
</tr>
<tr>
<td>IntendedReceiver () ThreadId</td>
<td>–none–</td>
<td>62</td>
</tr>
<tr>
<td>IoFpageLog2 (Word BaseAddress, int Log2FpageSize &lt; 64) Fpage</td>
<td>–none–</td>
<td>93</td>
</tr>
<tr>
<td>IoFpage (Word BaseAddress, int FpageSize) Fpage</td>
<td>–none–</td>
<td>93</td>
</tr>
<tr>
<td>IpcFailed (Msg Tag t) bool</td>
<td>–none–</td>
<td>62</td>
</tr>
<tr>
<td>IpcRedirected (Msg Tag t) bool</td>
<td>–none–</td>
<td>62</td>
</tr>
<tr>
<td>IpcSucceeded (Msg Tag t) bool</td>
<td>–none–</td>
<td>62</td>
</tr>
<tr>
<td>IpcXcpu (Msg Tag t) bool</td>
<td>–none–</td>
<td>62</td>
</tr>
<tr>
<td>Ipc (ThreadId to, FromSpecifier, Word Timeouts, ThreadId&amp; from) MsgTag</td>
<td>IPC</td>
<td>61</td>
</tr>
<tr>
<td>IsGlobalId (ThreadId t) bool</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>IsLocalId (ThreadId t) bool</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>IsNilFpage (Fpage f) bool</td>
<td>–none–</td>
<td>37</td>
</tr>
<tr>
<td>IsNilThread (ThreadId t) bool</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>IsVirtual (MemoryDesc&amp; m) bool</td>
<td>–none–</td>
<td>9</td>
</tr>
<tr>
<td>KernelGenDate (Void*, KernelInterface, Word&amp; year, month, day) Void</td>
<td>–none–</td>
<td>8</td>
</tr>
<tr>
<td>KernelId () Word</td>
<td>–none–</td>
<td>8</td>
</tr>
<tr>
<td>KernelInterface () Void*</td>
<td>KERNELINTERFACE</td>
<td>8</td>
</tr>
<tr>
<td>KernelInterface (Word&amp; ApiVersion, ApiFlags, KernelId) Void*</td>
<td>KERNELINTERFACE</td>
<td>8</td>
</tr>
<tr>
<td>Label (Msg&amp; msg) Word</td>
<td>–none–</td>
<td>47</td>
</tr>
<tr>
<td>Label (Msg Tag t) Word</td>
<td>–none–</td>
<td>46</td>
</tr>
<tr>
<td>LargeSpace Word const</td>
<td>–none–</td>
<td>94</td>
</tr>
<tr>
<td>Lcall (ThreadId to) MsgTag</td>
<td>LIPC</td>
<td>62</td>
</tr>
<tr>
<td>Lipc (ThreadId to, FromSpecifier, Word Timeouts, ThreadId&amp; from) MsgTag</td>
<td>LIPC</td>
<td>61</td>
</tr>
<tr>
<td>LoadBRs (int i, k, Word[&amp; k]) Void</td>
<td>–none–</td>
<td>11</td>
</tr>
<tr>
<td>LoadBRs (int i, k, Word[&amp; k]) Void</td>
<td>–none–</td>
<td>56</td>
</tr>
<tr>
<td>LoadBR (int i, Word w) Void</td>
<td>–none–</td>
<td>11</td>
</tr>
<tr>
<td>LoadBR (int i, Word w) Void</td>
<td>–none–</td>
<td>56</td>
</tr>
<tr>
<td>LoadMRs (int i, k, Word[&amp; k] w) Void</td>
<td>–none–</td>
<td>11</td>
</tr>
<tr>
<td>LoadMRs (int i, k, Word[&amp; k] w) Void</td>
<td>–none–</td>
<td>48</td>
</tr>
<tr>
<td>LoadMR (int i, Word w) Void</td>
<td>–none–</td>
<td>11</td>
</tr>
<tr>
<td>LoadMR (int i, Word w) Void</td>
<td>–none–</td>
<td>48</td>
</tr>
<tr>
<td>Load (Msg&amp; msg) Void</td>
<td>–none–</td>
<td>47</td>
</tr>
<tr>
<td>LocalId (ThreadId t) ThreadId</td>
<td>EXCHANGEREGISTERS</td>
<td>15</td>
</tr>
<tr>
<td>LocalId (ThreadId t) ThreadId</td>
<td>EXCHANGEREGISTERS</td>
<td>20</td>
</tr>
<tr>
<td>Function/Type</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>LocalMemory</td>
<td>Word const</td>
<td>122</td>
</tr>
<tr>
<td>Low</td>
<td>(MemoryDesc&amp; m) Word</td>
<td>9</td>
</tr>
<tr>
<td>LreplyWait</td>
<td>(ThreadId to, ThreadId&amp; from) MsgTag</td>
<td>62</td>
</tr>
<tr>
<td>MapGrantItems</td>
<td>(Acceptor a) bool</td>
<td>56</td>
</tr>
<tr>
<td>MapGrantItems</td>
<td>(Fpage RcvWindow) Acceptor</td>
<td>55</td>
</tr>
<tr>
<td>MapItem</td>
<td>(Fpage f, Word SndBase) MapItem</td>
<td>49</td>
</tr>
<tr>
<td>MapItem</td>
<td>(MapItem m) bool</td>
<td>50</td>
</tr>
<tr>
<td>MemoryControl</td>
<td>(Word control, Word&amp; attributes[4]) Void</td>
<td>MEMORYCONTROL 70</td>
</tr>
<tr>
<td>MemoryDesc</td>
<td>(Void* KernelInterface, Word num) MemoryDesc*</td>
<td>9</td>
</tr>
<tr>
<td>MemoryDesc</td>
<td>data type</td>
<td>8</td>
</tr>
<tr>
<td>MsgBuffer</td>
<td>data type</td>
<td>56</td>
</tr>
<tr>
<td>MsgTag</td>
<td>() MsgTag</td>
<td>46</td>
</tr>
<tr>
<td>MsgTag</td>
<td>(Msg&amp; msg) MsgTag</td>
<td>47</td>
</tr>
<tr>
<td>MSG data type</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>MyGlobalId</td>
<td>() ThreadId</td>
<td>15</td>
</tr>
<tr>
<td>MyLocalId</td>
<td>() ThreadId</td>
<td>17</td>
</tr>
<tr>
<td>MyOwn</td>
<td>() ThreadId</td>
<td>17</td>
</tr>
<tr>
<td>MapItem</td>
<td>data type</td>
<td>8</td>
</tr>
<tr>
<td>MemoryDesc</td>
<td>data type</td>
<td>9</td>
</tr>
<tr>
<td>MSG data type</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>NILpage</td>
<td>Fpage const</td>
<td>37</td>
</tr>
<tr>
<td>NILtag</td>
<td>MSG Tag const</td>
<td>46</td>
</tr>
<tr>
<td>nilthread</td>
<td>ThreadId const</td>
<td>15</td>
</tr>
<tr>
<td>NoAccess</td>
<td>Word const</td>
<td>37</td>
</tr>
<tr>
<td>NumMemoryDesc</td>
<td>(Void* KernelInterface) Word</td>
<td>8</td>
</tr>
<tr>
<td>PageRights</td>
<td>(Void* KernelInterface) Word</td>
<td>8</td>
</tr>
<tr>
<td>Pager</td>
<td>() ThreadId</td>
<td>17</td>
</tr>
<tr>
<td>Pager</td>
<td>(ThreadId t) ThreadId</td>
<td>EXCHANGEREGISTERS 20</td>
</tr>
<tr>
<td>PageSizeMask</td>
<td>(Void* KernelInterface) Word</td>
<td>8</td>
</tr>
<tr>
<td>PAL_Call</td>
<td>(Word idx, a1, a2, a3, Word&amp; r1, r2, r3) Word</td>
<td>PAL_CALL 104</td>
</tr>
<tr>
<td>PCICfgPageLog2</td>
<td>(Word BaseAddress, int Log2FpageSize &lt; 64) Fpage</td>
<td>110</td>
</tr>
<tr>
<td>PCICfgPage</td>
<td>(Word BaseAddress, int FpageSize ≥ 256) Fpage</td>
<td>110</td>
</tr>
<tr>
<td>PreemptionPending</td>
<td>bool</td>
<td>34</td>
</tr>
<tr>
<td>ProcDesc</td>
<td>(Void* KernelInterface, Word num) ProcDesc*</td>
<td>9</td>
</tr>
<tr>
<td>ProcDesc</td>
<td>data type</td>
<td>8</td>
</tr>
<tr>
<td>ProcessorControl</td>
<td>(Word ProcessorNo, control, InternalFrequency, ExternalFrequency, voltage) Word</td>
<td>69</td>
</tr>
<tr>
<td>ProcessorNo</td>
<td>() int</td>
<td>17</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word l, int u, Word&amp; [u] ut, int t, [MapItem GrantItem StringItem ]&amp; Items) Void</td>
<td>47</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word t, GrantItem g) Void</td>
<td>48</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word t, MapItem m) Void</td>
<td>47</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word t, StringItem s) Void</td>
<td>48</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word t, StringItem s) Void</td>
<td>48</td>
</tr>
<tr>
<td>Put</td>
<td>(Msg&amp; msg, Word u, Word w) Void</td>
<td>47</td>
</tr>
<tr>
<td>RcvWindow</td>
<td>(Acceptor a) Fpage</td>
<td>56</td>
</tr>
<tr>
<td>Readable</td>
<td>Word const</td>
<td>36</td>
</tr>
<tr>
<td>ReadExclusive</td>
<td>Word const</td>
<td>37</td>
</tr>
<tr>
<td>ReadPrecision</td>
<td>(Void* KernelInterface) Word</td>
<td>9</td>
</tr>
<tr>
<td>Receive</td>
<td>(ThreadId from) MsgTag</td>
<td>IPC 62</td>
</tr>
<tr>
<td>Receive</td>
<td>(ThreadId from, Time RcvTimeout) MsgTag</td>
<td>IPC 62</td>
</tr>
<tr>
<td>ReplyWait</td>
<td>(ThreadId to, ThreadId&amp; from) MsgTag</td>
<td>IPC 62</td>
</tr>
<tr>
<td>Reply</td>
<td>(ThreadId to, Time RcvTimeout, ThreadId&amp; from) MsgTag</td>
<td>IPC 62</td>
</tr>
<tr>
<td>ReservedMemoryType</td>
<td>Word const</td>
<td>9</td>
</tr>
<tr>
<td>Rights</td>
<td>(Fpage f) Word</td>
<td>37</td>
</tr>
<tr>
<td>SAL_Call</td>
<td>(Word idx, a1, a2, a3, a4, a5, a6, Word&amp; r1, r2, r3) Word</td>
<td>SAL_CALL 104</td>
</tr>
<tr>
<td>SAL_PCI_ConfigRead</td>
<td>(Word address, size, Word&amp; value) Word</td>
<td>SAL_CALL 104</td>
</tr>
</tbody>
</table>
### TABLE OF PROCS, TYPES, AND CONSTANTS

<table>
<thead>
<tr>
<th>proc</th>
<th>type</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAL_PCI_ConfigWrite</td>
<td>(Word address, size, value) Word</td>
<td>SAL_CALL</td>
</tr>
<tr>
<td>SameThreads</td>
<td>(ThreadId t, r) bool</td>
<td>ExchangeRegisters</td>
</tr>
<tr>
<td>SchedulePrecision</td>
<td>(Void* KernelInterface) Word</td>
<td>Schedule</td>
</tr>
<tr>
<td>Schedule</td>
<td>(ThreadId dest, Word TimeControl, ProcessorControl, prio, PreemptionControl, Word&amp; old_TimeControl) Word</td>
<td>Schedule</td>
</tr>
<tr>
<td>Send</td>
<td>(ThreadId to, MsgTag) Void</td>
<td>IPC</td>
</tr>
<tr>
<td>Send</td>
<td>(ThreadId to, Time SndTimeout) MsgTag</td>
<td>IPC</td>
</tr>
<tr>
<td>Set_CopFlag</td>
<td>(Word n) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Set_CopFlag</td>
<td>(Word n) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Set_ExceptionHandler</td>
<td>(ThreadId new) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_ExceptionHandler</td>
<td>(ThreadId NewHandler) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Label</td>
<td>(Msg&amp; msg, Word label) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_MsgTag</td>
<td>(Msg&amp; msg, MsgTag t) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_PageAttribute</td>
<td>(Fpage f, Word attribute) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Pager</td>
<td>(ThreadId NewPager) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Pager</td>
<td>(ThreadId t, p) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_PreemptionDelay</td>
<td>(ThreadId dest, Word sensitivePrio, Word maxDelay) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Priority</td>
<td>(ThreadId dest, Word prio) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_ProcessorNo</td>
<td>(ThreadId dest, Word ProcessorNo) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Propagation</td>
<td>(Msg&amp; Tag t) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Rights</td>
<td>(Fpage f, Word AccessRights) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_Timeslice</td>
<td>(ThreadId dest, Time ts, Time tq) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_UserDefinedHandle</td>
<td>(ThreadId t, Word handle) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_UserDefinedHandle</td>
<td>(Word NewValue) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_VirtualSender</td>
<td>(ThreadId t) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_VirtualSender</td>
<td>(ThreadId t) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Set_XferTimeouts</td>
<td>(Word NewValue) Void</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>SharedMemoryType</td>
<td>Word const</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>SizeLog2</td>
<td>(Fpage f) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Size</td>
<td>(Fpage f) Word</td>
<td>PreemptionDelay</td>
</tr>
<tr>
<td>Sleep</td>
<td>(Time t) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SmallSpace</td>
<td>(Word location, size) Word</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SndBase</td>
<td>(GrantItem g) Word</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SndBase</td>
<td>(MapItem m) Word</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SndFpage</td>
<td>(GrantItem g) Fpage</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SndFpage</td>
<td>(MapItem m) Fpage</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SpaceControl</td>
<td>(ThreadId SpaceSpecifier, Word control, Fpage KernelInterfaceFacePageArea, UtcbArea, ThreadId Redirector, Word&amp; old_Control) Word</td>
<td>SpaceControl</td>
</tr>
<tr>
<td>SpeculativeMemory</td>
<td>Word const</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Start</td>
<td>(ThreadId t) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Start</td>
<td>(ThreadId t, Word sp, ip) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Start</td>
<td>(ThreadId t, Word sp, ip, flags) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Stop</td>
<td>(ThreadId t) ThreadState</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Stop</td>
<td>(ThreadId t, Word&amp; sp, ip, flags) ThreadState</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreBRs</td>
<td>(int i, k, Word&amp; [k]) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreBRs</td>
<td>(int i, k, Word&amp; [k]) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreBR</td>
<td>(int i, Word&amp; w) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreMRs</td>
<td>(int i, k, Word&amp; [k] w) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreMRs</td>
<td>(int i, k, Word&amp; [k] w) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreMR</td>
<td>(int i, Word&amp; w) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StoreMR</td>
<td>(int i, Word&amp; w) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Store</td>
<td>(MsgTag t, Msg&amp; msg) Void</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StringItemsAcceptor</td>
<td>Acceptor const</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StringItems</td>
<td>(Acceptor a) bool</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StringItem</td>
<td>(int size, Void* address) StringItem</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StringItem</td>
<td>(StringItem&amp; s) bool</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>StringItem</td>
<td>data type</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Substrings</td>
<td>(StringItem&amp; s) Word</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>Substring</td>
<td>(StringItem&amp; s, Word n) Void*</td>
<td>ProcessorNo</td>
</tr>
<tr>
<td>SystemClock</td>
<td>() Clock</td>
<td>ProcessorNo</td>
</tr>
</tbody>
</table>

---

**used system call page**
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ThreadControl (ThreadId dest, SpaceSpecifier, Scheduler, Pager, Void* UtcbLocation) Word</td>
<td>THREADCONTROL</td>
<td>23</td>
</tr>
<tr>
<td>ThreadIdBits (Void* KernelInterface) Word</td>
<td>–none–</td>
<td>8</td>
</tr>
<tr>
<td>ThreadIdSystemBase (Void* KernelInterface) Word</td>
<td>–none–</td>
<td>8</td>
</tr>
<tr>
<td>ThreadIdUserBase (Void* KernelInterface) Word</td>
<td>–none–</td>
<td>9</td>
</tr>
<tr>
<td>ThreadId data type</td>
<td>–n/a–</td>
<td>15</td>
</tr>
<tr>
<td>ThreadNo (ThreadId t) Word</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>ThreadState data type</td>
<td>–n/a–</td>
<td>21</td>
</tr>
<tr>
<td>ThreadSwitch (ThreadId dest) Void</td>
<td>THREADSWITCH</td>
<td>30</td>
</tr>
<tr>
<td>ThreadWasHalted (ThreadState s) bool</td>
<td>–none–</td>
<td>21</td>
</tr>
<tr>
<td>ThreadWasIpcing (ThreadState s) bool</td>
<td>–none–</td>
<td>21</td>
</tr>
<tr>
<td>ThreadWasReceiving (ThreadState s) bool</td>
<td>–none–</td>
<td>21</td>
</tr>
<tr>
<td>Timeouts (Time SndTimeout, RcvTimeout) Word</td>
<td>–none–</td>
<td>63</td>
</tr>
<tr>
<td>TimePeriod (Word64 microseconds) Time</td>
<td>–none–</td>
<td>28</td>
</tr>
<tr>
<td>TimePoint (Clock at) Time</td>
<td>–none–</td>
<td>29</td>
</tr>
<tr>
<td>Timeslice (ThreadId dest, Time &amp; ts, Time &amp; tq) Word</td>
<td>–none–</td>
<td>33</td>
</tr>
<tr>
<td>Time data type</td>
<td>–n/a–</td>
<td>28</td>
</tr>
<tr>
<td>TypedWords (Msg Tag t) Word</td>
<td>–none–</td>
<td>46</td>
</tr>
<tr>
<td>Type (MemoryDesc&amp; m) Word</td>
<td>–none–</td>
<td>9</td>
</tr>
<tr>
<td>UncacheableExportedMemory Word const</td>
<td>–n/a–</td>
<td>112</td>
</tr>
<tr>
<td>UncacheableMemory Word const</td>
<td>–n/a–</td>
<td>112</td>
</tr>
<tr>
<td>UncacheableMemory Word const</td>
<td>–n/a–</td>
<td>96</td>
</tr>
<tr>
<td>UndefinedMemoryType Word const</td>
<td>–n/a–</td>
<td>9</td>
</tr>
<tr>
<td>Unmap (Fpage f) Void</td>
<td>UNMAP</td>
<td>39</td>
</tr>
<tr>
<td>Unmap (Word n, Fpage &amp; [n] Fpages) Void</td>
<td>UNMAP</td>
<td>39</td>
</tr>
<tr>
<td>Unmap (Word control) Void</td>
<td>UNMAP</td>
<td>39</td>
</tr>
<tr>
<td>UntypedWordsAcceptor Acceptor const</td>
<td>–n/a–</td>
<td>55</td>
</tr>
<tr>
<td>UntypedWords (Msg Tag t) Word</td>
<td>–none–</td>
<td>46</td>
</tr>
<tr>
<td>UseDefaultCacheLineAllocation CacheAllocationHint const</td>
<td>–n/a–</td>
<td>111</td>
</tr>
<tr>
<td>UseDefaultCacheLineAllocation CacheAllocationHint const</td>
<td>–n/a–</td>
<td>53</td>
</tr>
<tr>
<td>UseDefaultCacheLineAllocation CacheAllocationHint const</td>
<td>–n/a–</td>
<td>95</td>
</tr>
<tr>
<td>UserDefinedHandle () Word</td>
<td>EXCHANGEREGISTERS</td>
<td>20</td>
</tr>
<tr>
<td>UserDefinedHandle (ThreadId t) Word</td>
<td>–none–</td>
<td>17</td>
</tr>
<tr>
<td>Version (ThreadId t) Word</td>
<td>–none–</td>
<td>15</td>
</tr>
<tr>
<td>Wait (ThreadId&amp; from) MsgTag</td>
<td>IPC</td>
<td>62</td>
</tr>
<tr>
<td>Wait (Time RcvTimeout, ThreadId&amp; from) MsgTag</td>
<td>IPC</td>
<td>62</td>
</tr>
<tr>
<td>WaseXecuted (Fpage f) bool</td>
<td>–none–</td>
<td>40</td>
</tr>
<tr>
<td>WasReferenced (Fpage f) bool</td>
<td>–none–</td>
<td>40</td>
</tr>
<tr>
<td>WasWritten (Fpage f) bool</td>
<td>–none–</td>
<td>40</td>
</tr>
<tr>
<td>Writable Word const</td>
<td>–n/a–</td>
<td>36</td>
</tr>
<tr>
<td>WriteBackMemory Word const</td>
<td>–n/a–</td>
<td>112</td>
</tr>
<tr>
<td>WriteBackMemory Word const</td>
<td>–n/a–</td>
<td>122</td>
</tr>
<tr>
<td>WriteBackMemory Word const</td>
<td>–n/a–</td>
<td>96</td>
</tr>
<tr>
<td>WriteCoalescingMemory Word const</td>
<td>–n/a–</td>
<td>112</td>
</tr>
<tr>
<td>WriteCombiningMemory Word const</td>
<td>–n/a–</td>
<td>96</td>
</tr>
<tr>
<td>WriteProtectedMemory Word const</td>
<td>–n/a–</td>
<td>96</td>
</tr>
<tr>
<td>WriteThroughMemory Word const</td>
<td>–n/a–</td>
<td>122</td>
</tr>
<tr>
<td>WriteThroughMemory Word const</td>
<td>–n/a–</td>
<td>96</td>
</tr>
<tr>
<td>XferTimeouts () Word</td>
<td>–none–</td>
<td>17</td>
</tr>
<tr>
<td>Yield () Void</td>
<td>THREADSWITCH</td>
<td>30</td>
</tr>
<tr>
<td>ZeroTime Time const</td>
<td>–n/a–</td>
<td>28</td>
</tr>
</tbody>
</table>
Index

! =, 15, 26, 29
+ =, 26, 29, 37, 46, 54, 55
+ =, 29, 37, 46, 53–55
− =, 26, 29, 37, 55
− (ignored), vii
− =, 29, 37, 55
< =, 26, 29
< =, 26, 29
≡ (unchanged), vii
≡ =, 15, 26, 29, 46, 54
> =, 26, 29
> =, 26, 29
∼ (undefined), vii

σ_0, see sigma0

AbortIpC_and_stop, 21
AbortReceive_and_stop, 21
AbortSend_and_stop, 21
Accept, 56
Accepted, 56
acceptor, 55
ACPI_MEMORY_TYPE, 113
ActualSender, 17, 63
Address, 37
address space
creation/deletion, 41
initial, 79
AllocateNewCacheLines, 95
AllocateOnlyNewL1CacheLines, 95
anylocalthread, 15
anythread, 15
ApiFlags, 8
ApiVersion, 8
Append, 47, 56
ArchitectureSpecificMemoryType, 9
AssociateInterrupt, 24

BootInfo, 9
booting, 82–84
alpha, 134
ia32, 99
mips64, 144
powerpc, 126

BootLoaderSpecificMemoryType, 9
BR, see buffer registers
buffer registers, 55
alpha, 129
ia32, 87–88
ia64, 103
mips64, 137
powerpc, 117
cacheability, 52, 95, 96, 111, 112, 122, 143
CacheAllocationHint, 54
CacheNonTemporalAllLevels, 111
CacheNonTemporalL1, 111
CacheNonTemporalL2, 111
CachingEnabledMemory, 122
CachingInhibitedMemory, 122
Call, 61, 62
Clear, 47, 56
clock, 26
reading, 27
Clr_CopFlag, 17, 67
CompleteAddressSpace, 37
CompoundString, 53
convenience programming interface, vi
ConventionalMemoryType, 9
coprocessors, 67

DeassociateInterrupt, 24
ddebug registers, 98
DedicatedMemoryType, 9
DefaultMemory, 70, 96, 112, 122
DisablePreemption, 34
DisablePreemptionFaultException, 34
DoNotAllocateNewCacheLines, 95
EnablePreemption, 34
EnablePreemptionFaultException, 34
endian, 3
ErrorCode, 17, 62
exception
handling, 66
message
ia32, 97
ia64, 114
powerpc, 123
protocol, 78
ExceptionHandler, 17, 66
ExchangeRegisters, 20
eXecutable, 37
ExternalFreq, 10

Feature, 9
Flush, 40
Fpage, 37
fpage, 36–37
mapping, 57
receiving, 55
unmapping, 36, 38–40
FpageLog2, 37
FullyAccessible, 37
generic binary interface, vi
generic programming interface, vi
Get, 47, 48
GetStatus, 40
global thread ID, 14
GlobalId, 15, 20
GlobalMemory, 122
INDEX

GrantItem, 51
GuardedMemory, 122

High, 9

include files, viii
IntendedReceiver, 17, 62
InternalFreq, 10
interrupt
association, 22
thread ID, 14
I/O fpage, 93
IoFpage, 93
IoFpageLog2, 93
IPC, 57–63
aborting, 18
cross cpu, 60
propagation, 58
Ipc, 61
IpcFailed, 62
IpcPropagated, 62
IpcRedirected, 62
IpcSucceeded, 62
IsGlobalId, 15
IsLocalId, 15
IsNilFpage, 37
IsNilThread, 15
IsVirtual, 9

kernel features, 5
ia32, 92
kernel interface page, 2–10
data structure, 2–6
location, 41
retrieving, 7–10
KernelGenDate, 8
KernelId, 8
KernelInterface, 8
KernelSupplier, 8
KernelVersion, 8
KernelVersionString, 9
KipAreaSizeLog2, 9

Label, 46, 47
LargeSpace, 94
LeaIl, 62
lipc, 57
Load, 47
LoadBR, 11, 56
LoadBRs, 11, 56
LoadMR, 11, 48
LoadMRS, 11, 48
local ipc, 57
local thread ID, 14
LocalId, 15, 20
LocalMemory, 122
logical interface, vi
Low, 9
RecvWindow, 62

MapGrantItems, 55, 56
MapItem, 49, 50
memory descriptor, 6, 83–84
ia64, 113
MemoryControl, 70

MemoryDesc, 9
message registers, 44–45
alpha, 128–129
ia32, 87
ia64, 102–103
mips64, 136–137
powerpc, 116–117
messages
generating, 44–48
model specific registers, 98
MR, see message registers
MsgTag, 46, 47
MyGlobalId, 15, 17
MyLocalId, 15, 17
Myself, 15, 17
NaTPageMemory, 112
Never, 28
Nilpage, 37
Niltag, 46
nithread, 15
NoAccess, 37
NumMemoryDescriptors, 8
NumProcessors, 8

page
access rights, 4, 36, 49, 51, 76, 80
changing, 38, 49, 51
inspecting, 39
attributes, 80
ia32, 96
ia64, 112
mips64, 143
powerpc, 122
size, 3
pagefault
protocol, 76
Pager, 17, 20
pager, 76
changing, 17, 20, 23
PageRights, 8
PageSizeMask, 8
PAL procedure calls, 104
PAL_Call, 104
PCI Config fpage, 110
PCI Configuration Space
ia64, 104, 110
PCIConfigFpage, 110
PCIConfigFpageLog2, 110
preemption, 31, 34
protocol, 77
PreemptionPending, 34
privileged threads, vii
ProcDesc, 9
processor-specific binary interface, vi
ProcessorControl, 69
ProcessorNo, 16
ProcessorNo, 17
propagation, 58
Put, 47, 48

RecvWindow, 56
RDMSR, 98
Readable, 36
ReadExecOnly, 37
ReadPrecision, 9
Receive, 62
redirection, 42, 58
Reply, 62
ReplyWait, 62
ReservedMemoryType, 9
Rights, 37

SAL procedure calls, 104
SAL_Call, 104
SAL_PCI_ConfigRead, 104
SAL_PCI_ConfigWrite, 104
SameThreads, 15
Schedule, 33
SchedulePrecision, 9
segments, 98
Send, 62
send base, 49
sensitive prio, 31
Set_CopFlag, 17, 67
Set_ExceptionHandler, 17, 66
Set_Label, 47
Set_MsgTag, 46, 47
Set_PageAttribute, 71
Set_Page, 17, 20
Set_PagesAttributes, 71
Set_PreemptionDelay, 33
Set_Priority, 33
Set_ProcessorNo, 33
Set_Propagation, 63
Set_Rights, 37
Set_Timeslice, 33
Set_UserDefinedHandle, 17, 20
Set_VirtualSender, 17, 63
Set_XferTimeouts, 17
SharedMemoryType, 9
sigma0, 79
protocol, 79–81
Size, 37
SizeLog2, 37
Sleep, 62
small spaces, 94
SmallSpace, 94
SndBase, 50, 51
SndPage, 50, 51
SpaceControl, 42
SpeculativeMemory, 122
Start, 20
Stop, 21
Store, 47
StoreBR, 11, 56
StoreBRS, 11, 56
StoreMR, 11, 48
StoreMRs, 11, 48
StringItem, 53
StringItems, 56
StringItemsAcceptor, 55
strings, 52–54
receiving, 55
Substring, 53
Substrings, 53
system thread, 14, 58, 62
system-call links, 5
alpha, 130–133
ia32, 89
ia64, 105
mips64, 138–142
powerpc, 118–121

SystemBase, 4
SystemClock, 27

TCR, see thread control registers
thread
creation, 22
halting, 18
ID, 14
id, 15, see thread ID
migration, 32
priority, 31
privileged, vii
startup protocol, 74
state, 21, 32
version, 14, 22

thread control registers, 16–17
alpha, 128
ia32, 86
ia64, 102
mips64, 136
powerpc, 116
thread ID, 14–15
retrieving, 17, 20
ThreadControl, 23
ThreadIdBits, 8
ThreadIdSystemBase, 8
ThreadIdUserBase, 9
ThreadId, 15
ThreadSwitch, 30
ThreadWasHalted, 21
ThreadWas IPCing, 21
ThreadWasReceiving, 21
time, 28–29
time quantum, 31
Timeouts, 63
TimePeriod, 28
TimePoint, 29
Timeslice, 33
timeslice, 31
donation, 30
Type, 9
TypedWords, 46

UncacheableExportedMemory, 112
UncacheableMemory, 96, 112
UndefinedMemoryType, 9
Unmap, 39
UntypedWords, 46
UntypedWordsAcceptor, 55
upward compatibility, vii
UseDefaultCacheLineAllocation, 53, 95, 111
UserBase, 4
UserDefinedHandle, 16, 19
UserDefinedHandler, 17, 20
using the API, viii
UTC

UTC<br />
location, 41
size, 4, 23, 41
UtcbAlignmentLog2, 9
UtcbAreaSizeLog2, 9
UtcbSize, 9

Version, 15
virtual registers, 11

Wait, 62
INDEX

WaseXecuted, 40
WasReferenced, 40
WasWritten, 40
Word, vii
Word16, vii
Word32, vii
Word64, vii
Writable, 36
WriteBackMemory, 96, 112, 122
WriteCoalescingMemory, 112
WriteCombiningMemory, 96
WriteProtectedMemory, 96
WriteThroughMemory, 96, 122
WRMSR, 98

XferTimeouts, 17

Yield, 30

ZeroTime, 28