Documenting Xenon’s Page_Alloc Module

James Kirby, Jr.
John McDermott
Myong Kang
Bruce Montrose

Center for High Assurance Computer Systems
Information Technology Division

December 10, 2007

Approved for public release; distribution is unlimited.
One of the critical assurance requirements for achieving medium or high assurance is a requirement for significant modularity in design and implementation. As part of the Xenon effort to create a secure Xen with a medium to high degree of assurance, we have embarked on its remodularization, a documented decomposition into well-defined pieces with well-defined relationships among them. This remodularization of Xen is based on the information hiding principle. Associated with an information hiding module may be a provided interface, a set of public programs (e.g., functions, subroutines, macros) that programs outside the module can use to accomplish their work. Documentation of a module’s provided interface serves as a contract between the module’s users and its developers. This report documents the decomposition of the Xen page_alloc module and the specification of the provided interface of each of its submodules.
CONTENTS

Introduction .................................................................................................................................................... 1

1 Environmental Model of Hardware Memory .......................................................................................... 4

2 Page Allocator (was Allocation Bitmap) ................................................................................................. 6

   2.1 init_boot_allocator .......................................................................................................................... 6
   2.2 allocated_in_map ............................................................................................................................. 7
   2.3 map_alloc ........................................................................................................................................ 8
   2.4 map_free .......................................................................................................................................... 9

3 Boot-Time Allocator ................................................................................................................................ 10

   3.1 init_boot_pages ............................................................................................................................... 10
   3.2 alloc_boot_pages_at ........................................................................................................................ 11
   3.3 alloc_boot_pages ............................................................................................................................ 12
   3.4 end_boot_allocator ....................................................................................................................... 13

4 Run-Time Allocator (was Binary Buddy Allocator) ............................................................................... 14

   4.1 init_heap_pages ............................................................................................................................... 14
   4.2 free_heap_pages ............................................................................................................................. 15
   4.3 alloc_heap_pages ........................................................................................................................... 16
   4.4 avail_heap_pages ............................................................................................................................ 17
   4.5 scrub_heap_pages ........................................................................................................................... 18
   4.6 dump_heap .................................................................................................................................... 19

5 Xen Heap Allocator (was Xen-Heap Sub-Allocator) ............................................................................. 20

   5.1 init_xenheap_pages ......................................................................................................................... 20
   5.2 alloc_xenheap_pages ....................................................................................................................... 21
   5.3 free_xenheap_pages ......................................................................................................................... 22

6 Dom Heap Allocator (was Domain-Heap Sub-Allocator) ....................................................................... 23

   6.1 init_domheap_pages ........................................................................................................................ 23
   6.2 assign_pages .................................................................................................................................. 24
   6.3 __alloc_domheap_pages .................................................................................................................. 25
   6.4 alloc_domheap_pages ...................................................................................................................... 28
   6.5 avail_domheap_pages ...................................................................................................................... 31
   6.6 free_domheap_pages ....................................................................................................................... 32

7 Page Scrubbing ....................................................................................................................................... 33

   7.1 page_scrub_softirq ......................................................................................................................... 33
   7.2 avail_scrub_pages ............................................................................................................................ 34
Abstract

One of the critical assurance requirements for achieving medium or high assurance is a requirement for significant modularity in design and implementation. As part of the Xenon effort to create a secure Xen with a medium to high degree of assurance, we have embarked on its remodularization, a documented decomposition into well-defined pieces with well-defined relationships among them. This remodularization of Xen is based on the information hiding principle. Associated with an information hiding module may be a provided interface, a set of public programs (e.g., functions, subroutines, macros) that programs outside the module can use to accomplish their work. Documentation of a module's provided interface serves as a contract between the module's users and its developers. This report documents the decomposition of the Xen page_alloc module and the specification of the provided interface of each of its submodules.

Introduction

As part of the Xenon effort to create a secure Xen with a medium to high degree of assurance, we have embarked on its remodularization—a documented decomposition into well-defined pieces with well-defined relationships among them. When the modularization is complete, its documentation will support certification reviews and will help developers and maintainers identify parts of the Xen they must understand to accomplish some task without looking at irrelevant parts.

One of the critical assurance requirements for achieving medium (EAL 5) or high (EAL 6/7) assurance is a requirement for significant modularity in design and implementation. This increased modularity serves multiple purposes. First, increased modularity makes any security analysis more believable. Second, it reduces the scope of a flaw (or malware in some cases); that is modularity reduces dependencies between parts of the software, so flaws remaining in the code are less likely to be exploitable. Finally, modularity can be used to separate code into security-relevant and security-irrelevant modules. This reduces the amount of code that needs to be built according to high assurance rules.

This remodularization of Xen is based on the information hiding principle, which Dave Parnas described in his well-known paper, On the Criteria to Be Used in Decomposing Systems into Modules [CACM 1972]. A later paper, The Modular Structure of Complex Systems [Parnas, Clements, and Weiss, IEEE TSE, March 1985], which reported results of an NRL project to redevelop the operational flight program (OFP) for the Navy's A-7E attack aircraft, describes techniques that aid in applying the principle to a real system.

∗ This software is a Research Work of the United States Naval Research Laboratory, derived from GPL software. Any distribution of a source code or binary form of this software is prohibited. Release of this software outside the Department of Defense may be a violation of U.S. Law. The derived portion of this software is United States Government Work not protected by U.S. Copyright.

Manuscript approved October 29, 2007.
An information hiding module is a design construct. Each module has a secret, one or more decisions (which might also be thought of as assumptions) that developers judge likely to change or that they judge it is useful not to distribute throughout the system. Associated with a module may be a set of a public programs (e.g., functions, subroutines, macros) that programs outside the module can use to accomplish their work. These public programs constitute a public or provided interface representing decisions and assumptions upon which using programs may depend. When the module is carefully designed, the decisions it hides can be changed without invalidating the decisions and assumptions that the provided interface represents and upon which using programs depend.

A module’s secret is decomposed by its children. For example, a module that hides characteristics of peripheral devices that are likely to change might be decomposed into a set of modules, each of which hides characteristics of a particular class of device that are likely to change. This module structure or information hiding structure for a system is a tree of modules. It is useful to think of each module in the hierarchy as a work assignment for one or more developers or maintainers. The most difficult (and important) parts of designing the module structure are identifying the module secrets and clearly and concisely describing them. As a tree, the module structure can be usefully presented in a variety of ways, including as an indented list and as a UML class diagram.

Documentation of a module’s provided interface serves as a contract between the module’s users and its developers. To be useful, this documentation should be written so that it does not reveal or assume decisions designers intend the module to hide. Some implications of this are that documentation of the behavior of functions on the provided interface should not be written in terms of such implementation details as their algorithms, internal data structures, nor which functions they may call.

To facilitate describing the behavior of functions on modules’ provided interfaces, we develop an environmental model which provides an application-specific ontology for the system [Kirby, COMPSAC 2006]. The model records the system boundary by identifying objects in the environment of the system (which may include the system itself and components of the system) and attributes of the objects that may be relevant to the system, referred to as environmental attributes. The declaration of an attribute in the model includes its type, which characterizes the values it can assume, and a description of how to interpret its value. Identifying appropriate and relevant attributes—those that describe what the software can sense, control, and affect—is key. Descriptions of software behavior can be written in terms of these attributes.

UML classes represent objects in the system environment. Standard UML class notation may record relationships among the classes of the environmental model, the relative cardinality of the objects abstracted by the classes of the environmental model, and the cardinality of the attributes of each object. The attributes associated with each object are listed in the corresponding class. Attributes whose values the software can sense (either directly or via physical or cyber sensors) are referred to as monitored attributes. Attributes whose values the software can set or affect (either directly or via physical or cyber actuators) are referred to as controlled attributes. Monitored attributes and controlled attributes can be distinguished by assigning the former to a class compartment labeled monitored, and assigning the latter to a class compartment labeled controlled (see Fig. 1). Assigning an attribute to an unnamed compartment indicates that the engineer has not decided whether the attribute is monitored or controlled.

Fig. 1 illustrates an environmental model of hardware memory which consists of a number of hardware pages. The monitored attribute mMaxPage gives the number of pages in memory. The monitored attribute mPageSize gives the size of a page in bytes (in this model all pages have the same size). The figure illustrates attributes of a HardwarePage, e.g., monitored attribute mBad, controlled attributes cAllocated, cZeroized. Xen can sense the values of monitored attributes and set the values of controlled attributes. The tabular declarations of attributes in Tables 1 through 3 include a description of how to interpret attribute values.

Section 2.3 illustrates the specification of the function map_alloc(), which allocates hard-
ware pages. As indicated by the table, the function has two parameters. Both parameters are inputs to the function (indicated by the \textit{I} in the \textit{Mode} column) and are of type \texttt{unsigned long}. The first parameter (\texttt{p1}) gives the linear address of a hardware page. The second parameter specifies a number of hardware pages. Below the table, \textit{Undesired Events} identifies undesired events—requesting pages that $\texttt{map\_alloc()}$ has already allocated and requesting pages before initializing the module \textit{Page Allocator}—which, encountered at run-time, prevent correct operation of the function. \textit{Effects} describes the effect of calling $\texttt{map\_alloc()}$, which is to set to \texttt{true} the \texttt{cAllocated} attribute of all hardware pages with linear addresses in the range \[ [p1, p1 + p2 - 1] \]

Section 2.2 illustrates the specification of the function \texttt{allocated\_in\_map}, which callers can use to determine whether a particular hardware page is allocated. In the parameter table, labeling the first parameter \texttt{p0}, rather than \texttt{p1}, indicates that it specifies the value returned by the function, which the term \texttt{tAllocated} specifies. The \textit{O} in the \textit{Mode} column indicates that the parameter is output from the function to its caller (as would be expected of the function’s return value). The definition of \texttt{tAllocated} in the Dictionary specifies that the value returned by the function (\texttt{p0}) is the value of the \texttt{cAllocated} attribute of the hardware page whose linear address is given by \texttt{p1}.

Some tables in the \textit{Effects} and \textit{Dictionary} subsections of Sections 6.2, 6.3, and 6.4 are more complicated than those discussed above. For example, the first table in \textit{Effects} of Section 6.2 specifies the values on return to the caller of the variables indicated in the leftmost cell below the double line (e.g., \texttt{p.cPageOwner}, \texttt{p.cDomAllocated}, \texttt{p.cRefCount}). The prime appended to the variable name indicates the value of the variables on return. The values of unprimed variables are those established on the call to the function. The cells in the last rows (below the double line) to the right of the double line specify alternative value(s) for the variable(s). The leftmost cells above the double line partition the state space of the function. Exactly one of them is true, which selects the corresponding row of rules for determining the value of the variable(s). These rules are written so that they also partition the state space—exactly one of them is true. If \texttt{p1.mDying} is true—i.e., the domain referred to by the first parameter is dying—then the first row determines which of the alternative set of values in the last row apply. The \texttt{true} in the first column to the right of the double line—which can be thought of as specifying \textit{always}—indicates that the values in the last row to the immediate right of the double line apply. The \texttt{false} in the rightmost column, which can be thought of as \textit{never}, indicates that the values in the rightmost column of the last row do not apply when \texttt{p1.mDying} is true. When \texttt{p1.mDying} is false, the more complex expressions to the right of the double line in the second row determine which set of values in the bottom row apply. Note that the table is quantified by the expression above the table.

The first table in \textit{Effects} in Section 6.3 which has only the vertical line is interpreted differently. The cells to the left of the double line represent alternative conditions. If one of them is true, then the corresponding expression to the right of the double line describes effects of calling the program. If none of the alternative conditions to the right of the double line is true, then the table does not describe any effects of the calling the program.
Module page_alloc

**Secret.** The page_alloc module’s secret is how memory is allocated in Xenon. This includes how Xen keeps track of which pages of memory have been allocated and which have not.

# 1 Environmental Model of Hardware Memory

Fig. 1 provides a graphical view of Xenon’s environmental model of the memory hardware on which it runs. This is a software view of the memory hardware. While the hardware has its own particular addressing scheme based on address lines, this Xenon model uses the linear address scheme. Hardware memory comprises a set of hardware pages. The figure illustrates two attributes of hardware memory, mPageSize and mMaxPage. Being in the *monitored* compartment of the Hardware Memory class indicates that Xenon is able to determine the values of the two attributes, but is unable to change those values.

The Hardware Page class in Fig. 1 indicates that hardware pages have attributes whose values Xenon can sense but not change and that it has attributes whose values Xenon can change (attributes in the *controlled* compartment of the Hardware Page class).

Table 1 declares the attributes of hardware memory which the environmental model in Fig. 1 introduced. From Table 1 we see that mMaxPage denotes the number of pages in hardware memory and that mPageSize denotes the size of hardware pages in bytes. Table 3 declares the attributes of hardware pages which the environmental model introduced.
Table 1: Hardware Memory Attribute Declarations

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Class</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMaxPage</td>
<td>integer</td>
<td>monitored</td>
<td>Denotes the number of pages in hardware memory.</td>
</tr>
<tr>
<td>mPageSize</td>
<td>integer</td>
<td>monitored</td>
<td>Denotes the number of bytes in a hardware page.</td>
</tr>
</tbody>
</table>

Table 2: Hardware Page Attribute Declarations

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Class</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAddress</td>
<td>integer</td>
<td>monitored</td>
<td>Denotes the linear address of the hardware page.</td>
</tr>
<tr>
<td>cAllocated</td>
<td>boolean</td>
<td>controlled</td>
<td>cAllocated = true iff Xen has allocated the hardware page.</td>
</tr>
<tr>
<td>mBad</td>
<td>boolean</td>
<td>monitored</td>
<td>mBad = true iff the hardware page is not to be used.</td>
</tr>
<tr>
<td>cZeroized</td>
<td>boolean</td>
<td>controlled</td>
<td>cZeroized = true iff the hardware page is zeroized.</td>
</tr>
<tr>
<td>cZone</td>
<td>yZoneType</td>
<td>controlled</td>
<td>Denotes zone to which Xen has assigned the hardware page.</td>
</tr>
<tr>
<td>cPageOwner</td>
<td>yDomain</td>
<td>controlled</td>
<td>Denotes a domain to which Xen has assigned the hardware page.</td>
</tr>
<tr>
<td>cRefCount</td>
<td>int</td>
<td>controlled</td>
<td>Count of references to the page.</td>
</tr>
<tr>
<td>cDomAllocated</td>
<td>boolean</td>
<td>controlled</td>
<td>Indicates whether Xen has assigned the hardware page to a domain.</td>
</tr>
<tr>
<td>cScrubMe</td>
<td>boolean</td>
<td>controlled</td>
<td>Indicates a page that Xen needs to zeroize.</td>
</tr>
</tbody>
</table>

Dictionary

- **yDomain** is a handle for a Xen domain.
- **yZoneType** denotes memory zones. Enumerated values are: xen, dom, dma, any (Xen makes limited and inconsistent use of the latter).
- **mfn2Page()**, **page2Mfn()** are functions on addresses.

\[
mfn2Page(\text{linear address}) \rightarrow \text{virtual address} \\
page2Mfn(\text{virtual address}) \rightarrow \text{linear address}
\]

\[
(\forall p \in \text{HardwareMemory})(\exists \text{virtual address } v)(v = mfn2Page(p) \Rightarrow p = page2Mfn(v))
\]

\[
(\forall \text{virtual address } v)(\exists p \in \text{HardwareMemory})(p = page2Mfn(v) \Rightarrow v = mfn2Page(p))
\]

Fig. 2 graphically illustrates the module structure of the Page Alloc module. The remainder of this document describes each of the submodules in turn, describing its secret and specifying the programs on its provided interface.

Table 3: Domain Attribute Declarations

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Class</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mDying</td>
<td>boolean</td>
<td>monitored</td>
<td>Indicates whether the domain is dying.</td>
</tr>
<tr>
<td>mMaxPages</td>
<td>unsigned int</td>
<td>monitored</td>
<td>Indicates the maximum number of pages that Xen assigns to a domain.</td>
</tr>
<tr>
<td>cTotPages</td>
<td>unsigned int</td>
<td>controlled</td>
<td>Indicates the number of pages that Xen has assigned to a domain.</td>
</tr>
<tr>
<td>cDomainID</td>
<td>domid_t</td>
<td>controlled</td>
<td>Indicates the identifier of a domain.</td>
</tr>
</tbody>
</table>
2 Page Allocator (was Allocation Bitmap)

Secret. This module hides how to keep track of which hardware pages of memory have and have not been allocated.

2.1 init_boot_allocator

Initialize boot-time memory allocation mechanism.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>paddr_t</td>
<td>Starting memory location of available hardware memory to manage.</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>paddr_t</td>
<td>Starting memory location of hardware memory to manage.</td>
</tr>
</tbody>
</table>

Undesired Events

- uAllocationAlreadyInitialized. Hardware memory allocation already initialized.
- uAllocationNotInitialized. Hardware memory allocation not initialized.

Effects

$$(\forall p \in HardwareMemory)(p.mAddress \in [p1, mMaxPage - 1] \Rightarrow p.cAllocated' = true)$$

- enables uAllocationAlreadyInitialized
- disables uAllocationNotInitialized

Issues

- Don’t think behavior is quite right. Don’t think the pages containing the bit map, which is at the beginning of the memory pointed to by \textit{p1}, is allocated.
2.2  allocated_in_map

Hardware page already allocated?

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>boolean</td>
<td>tAllocated</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned long</td>
<td>Linear address of a hardware page.</td>
</tr>
</tbody>
</table>

Undesired Events

- uAllocationNotInitialized
- uNotLegalAddress

Effects

None.

Dictionary

- tAllocated boolean

\[(\exists p \in HardwareMemory)(p.mAddress = p1 \Rightarrow tAllocated' = p.cAllocated)\]

Issues

- Does not report undesired events encountered.
### 2.3 map_alloc

Allocate hardware pages.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned long</td>
<td>Linear address of a hardware page.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned long</td>
<td>Count of hardware pages.</td>
</tr>
</tbody>
</table>

**Undesired Events**

- **uHardwarePagesAlreadyAllocated.** The requested hardware pages are already allocated.
- **uAllocationNotInitialized**

**Effects**

\[ (\forall p \in \text{HardwareMemory})(p.mAddress \text{ in } [p1, p1 + p2 - 1] \Rightarrow p.cAllocated' = true) \]

- Enables the undesired event **uHardwarePagesAlreadyAllocated** for \( p2 \) hardware pages starting at page number \( p1 \).
- Disables the undesired event **uHardwarePagesNotAllocated** for \( p2 \) hardware pages starting at page number \( p1 \).

**Issues**

- Does not report undesired events encountered.
2.4 map_free

Return allocated hardware pages to free store.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned long</td>
<td>Hardware page number.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned long</td>
<td>Count of hardware pages.</td>
</tr>
</tbody>
</table>

Undesired Events

- **uHardwarePagesNotAllocated.** Returned hardware pages were not allocated.
- **uAllocationNotInitialized**

Effects

\[ (\forall p \in \text{HardwareMemory})(p.mAddress \text{ in } [p1, p1 + p2 - 1] \Rightarrow p.cAllocated' = false) \]

- Disables the undesired event **uHardwarePagesAlreadyAllocated** for \( p2 \) hardware pages starting at page number \( p1 \).
- Enables the undesired event **uHardwarePagesNotAllocated** for \( p2 \) hardware pages starting at page number \( p1 \).

Issues

- Does not report undesired events encountered.
3 Boot-Time Allocator

The Boot-Time Allocator module hides how the initial allocation of memory is performed.

3.1 init_boot_pages

Initial allocation of pages.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>paddr_t</td>
<td>Linear address of a hardware page.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>paddr_t</td>
<td>Linear address of a hardware page.</td>
</tr>
</tbody>
</table>

Effects

\[(\forall p \in \text{HardwareMemory})(p.mAddress \in [p1, p2] \land p.mBad = true \Rightarrow p.cAllocated' = true)\]

Issues

- Why are there both init_boot_pages and init_boot_allocator? Why not combine?
- What if
  \[p1 \geq p2?\]
3.2 alloc_boot_pages_at

Allocate specified number of free pages starting at specified linear address.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>unsigned long</td>
<td>tFirstAllocatedPage</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned long</td>
<td>Number of hardware pages.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned long</td>
<td>Linear address of a hardware page.</td>
</tr>
</tbody>
</table>

**Effects**

\[(\forall p \in \text{HardwareMemory})(p.mAddress \in [p2, p1 + p2 - 1] \Rightarrow p.cAllocated = false) \Rightarrow \\
(\forall p \in \text{HardwareMemory})(p.mAddress \in [p2, p1 + p2 - 1] \Rightarrow p.cAllocated' = true)\]

**Dictionary**

| tFirstAllocatedPage | unsigned long |

\[(\forall p \in \text{HardwareMemory})(p.mAddress \in [p2, p1 + p2 - 1] \Rightarrow p.cAllocated = false) \Rightarrow \\
tFirstAllocatedPage' = p2\]

\[(\exists p \in \text{HardwareMemory})(p.mAddress \in [p2, p1 + p2 - 1] \land \text{HardwarePage}[i].cAllocated = true) \Rightarrow \\
tFirstAllocatedPage' = 0\]
### 3.3 alloc_boot_pages

Allocate specified number of free pages.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>unsigned long</td>
<td>tFirstAllocatedPage</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned long</td>
<td>Number of hardware pages.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned long</td>
<td>Hardware page alignment.</td>
</tr>
</tbody>
</table>

**Effects**

\[
(\exists p \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(
\begin{align*}
& \hat{p}.mAddress \in [p.mAddress, p.mAddress - 1] \land \hat{p}.c\text{Allocated} = false \\
& \Rightarrow \hat{p}.c\text{Allocated}' = true
\end{align*}
\]

**Dictionary**

\[
tFirstAllocatedPage \quad \text{unsigned long}
\]

\[
(\exists p \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(
\begin{align*}
& \hat{p}.mAddress \in [p.mAddress, p.mAddress - 1] \land \hat{p}.c\text{Allocated} = false \\
& \Rightarrow tFirstAllocatedPage' = p.mAddress
\end{align*}
\]

\[
(\exists \hat{p} \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(
\begin{align*}
& \hat{p}.mAddress \in [p.mAddress, p.mAddress - 1] \land \hat{p}.c\text{Allocated} = false \\
& \Rightarrow tFirstAllocatedPage' = 0
\end{align*}
\]

**Issues**

- Not handling hardware page alignment (parameter p2) correctly. Yet.
- Assume there are mMaxPage locations, so memory runs from 0 to mMaxPage - 1. Why?
- This effects section needs more thought. The calculation of j doesn’t look right.
3.4 end_boot_allocator

Assign remaining free pages to domain.

Effects

\[(\forall p \in HardwareMemory)(p.mAddress \in [0, mMaxDmaPfn] \land p.cAllocated = false \Rightarrow p.cAllocated' = true \land p.cZone' = dma)\]

\[(\forall p \in HardwareMemory)(p.mAddress \in [mMaxDmaPfn + 1, mMaxPage - 1] \land p.cAllocated = false \Rightarrow p.cAllocated' = true \land p.cZone' = dom)\]

Issues

• Looks like we need to distinguish before and after state (as with the primes, above).
4 Run-Time Allocator (was Binary Buddy Allocator)

Secret. Xen partitions memory into a number of zones. The Run-Time Allocator module manages the allocation of blocks of pages of memory from, and the deallocation of blocks of pages of memory to these zones. This module hides the algorithms and data structures used to implement the allocation and deallocation of memory.

4.1 init_heap_pages

Initialize heap pages.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned int</td>
<td>Zone.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>struct page_info *</td>
<td>Page.</td>
</tr>
<tr>
<td>p3</td>
<td>I</td>
<td>unsigned long</td>
<td>Number of pages.</td>
</tr>
</tbody>
</table>

Effects

\( (\forall p \in \text{HardwareMemory})(p.mAddress \in [\text{page2M fn}(p2), \text{page2M fn}(p2) + p3 - 1]) \Rightarrow \)
\( p.cAllocated' = false \land p.cZone' = p1 \)
4.2 free_heap_pages

Put block of pages in free space for specified zone.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned int</td>
<td>Zone.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>struct page_info *</td>
<td>Page.</td>
</tr>
<tr>
<td>p3</td>
<td>I</td>
<td>unsigned int</td>
<td>Order of pages.</td>
</tr>
</tbody>
</table>

Undesired Events

- uRequestTooLarge
- uBadZone

Effects

\[(\forall p \in \text{HardwareMemory})(p.mAddress \text{ in } [\text{page2M fn}(p2), \text{page2M fn}(p2) + 2^{p3} - 1] \Rightarrow p.cAllocated' = false) \land p.cZone' = p1)\]

Issues

- Function does not detect either UE.
- When can the zone a block belongs to change?
- I assume that the user of this function is not concerned with the merging of blocks of memory into larger blocks, nor with the algorithms and data structures involved in managing and implementing such merging.
  Of course, the implementor is.
### 4.3 alloc_heap_pages

Allocate block of pages from specified zone.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>struct page_info *</td>
<td>tAllocatedPageBlock</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned int</td>
<td>Zone.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned int</td>
<td>Order of pages.</td>
</tr>
</tbody>
</table>

**Undesired Events**
- uRequestTooLarge
- uNoSuitableBlocks

**Effects**

\[
(\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory)(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p2} - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = p1) \Rightarrow \\
\hat{p}.cAllocated' = true
\]

**Dictionary**

- **tAllocatedPageBlock** struct page_info *

\[
(\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory)(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p2} - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = p1) \Rightarrow \\
tAllocatedPageBlock' = mfn2Page(p.mAddress)
\]

\[
(\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory)(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p2} - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = p1) \Rightarrow \\
tAllocatedPageBlock' = null
\]

\[
p2 > mMaxOrder \Rightarrow tAllocatedPageBlock' = null
\]

**Issues**
- The function detects both UEs, but reports either by returning null.
4.4 **avail_heap_pages**

How many unused pages?

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>unsigned long</td>
<td>tNumAvailPages</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>int</td>
<td>Identify one zone or all zones.</td>
</tr>
</tbody>
</table>

**Undesired Events**

- uBadZone
- uNotInitialized

**Effects**

None.

**Dictionary**

* tNumAvailPages unsigned long

\[
tNumAvailPages' = \left\| (\forall p \in \text{HardwareMemory})(p.\text{Allocated} = \text{false} \land (p1 = p.\text{cZone} \lor p1 = -1)) \right\|
\]

**Issues**

- Elsewhere, zones is declared an unsigned int. Here, zones is declared an integer, presum-
  ably to allow -1 to be used to indicate all zones.
- UEIs neither detected nor reported.
4.5 scrub_heap_pages

Scrub unallocated pages from all heap zones.

Undesired Events

•

Effects

\[(\forall p \in HardwareMemory)(p.cAllocated = false \Rightarrow p.cZeroized' = true)\]

Issues

• There may be details of visible behavior this does not yet address, e.g., progress dots, process pending timers.
4.6 dump_heap

Print allocation information on heap zones.

Undesired Events
•

Effects
None.

Issues
• Not capturing printouts.
• Printing is not captured in environmental model.
5 Xen Heap Allocator (was Xen-Heap Sub-Allocator)

Secret. The Xen Heap Allocator module manages the allocation of blocks of pages of memory from, and the deallocation of blocks of pages of memory to the xen heap zone. This module hides the algorithms and data structures used to implement the allocation and deallocation of memory.

5.1 init_xenheap_pages

Initialize xen heap pages.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>paddr_t</td>
<td>Address of first page of xen heap.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>paddr_t</td>
<td>Address of last page of xen heap.</td>
</tr>
</tbody>
</table>

Undesired Events
- Detects but does not report $p2 \leq p1$.

Effects

$$(\forall p \in HardwareMemory)(p.mAddress \in [page2Mfn(p1), page2Mfn(p2) - 1] \Rightarrow (p.cAllocated' = false \land p.cZone' = xen))$$

Issues
- This doesn’t yet deal with “rounding” addresses up and down, nor with leaving one page buffer between xen and dom zones.
- Whose responsibility is it to know the location of the xen heap?
5.2 alloc_xenheap_pages

Allocate block of pages from the xen heap zone.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>void *</td>
<td>tAllocatedPageBlock</td>
</tr>
<tr>
<td>p1</td>
<td>I</td>
<td>unsigned int</td>
<td>Order of pages.</td>
</tr>
</tbody>
</table>

Undesired Events

- uRequestTooLarge
- uNoSuitableBlocks

Effects

\[(\exists p \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^p1 - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = \text{xen}) \Rightarrow \hat{p}.cAllocated' = true)\]

Dictionary

\[(\exists p \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^p1 - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = \text{xen}) \Rightarrow tAllocatedPageBlock' = \text{mfn2Page}(p.mAddress)\]

\[(\exists p \in \text{HardwareMemory})(\forall \hat{p} \in \text{HardwareMemory})(\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^p1 - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = \text{xen}) \Rightarrow tAllocatedPageBlock' = \text{null})\]

\[p1 > \text{mMaxOrder} \Rightarrow tAllocatedPageBlock' = \text{null}\]

Issues

- The function detects both UEs, but reports either by returning null.
5.3 free_xenheap_pages

Put block of pages in free space for xen heap zone.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>void *</td>
<td>Virtual address of block of pages.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned int</td>
<td>Order of pages.</td>
</tr>
</tbody>
</table>

Undesired Events

- uRequestTooLarge
- uNoAddress

Effects

\[(\forall p \in \text{HardwareMemory})(p.mAddress \in [\text{page2Mfn}(p1), \text{page2Mfn}(p1) + 2^{p2} - 1]) \Rightarrow (p.cAllocated' = false) \land \text{HardwarePage}[i].cZone = \text{xen})\]

Issues

- I assume that I don’t need to set all the p.cZone’ = xen, since they should be already set.
- Function does not detect nor report UE.
- Can the zone a block belongs to change? So the zone should have been and should remain xen, eh?
- I assume that the user of this function is not concerned with the merging of blocks of memory into larger blocks, nor with the algorithms and data structures involved in managing and implementing such merging.
  Of course, the implementor is.
6 Dom Heap Allocator (was Domain-Heap Sub-Allocator)

Secret. The Dom Heap Allocator module manages the allocation of blocks of pages of memory from, and the deallocation of blocks of pages of memory to the domain heap zone. This module hides the algorithms and data structures used to implement the allocation and deallocation of memory.

6.1 init_domheap_pages

Initialize domain heap.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>paddr</td>
<td>Linear address of page.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>paddr</td>
<td>Linear address of page.</td>
</tr>
</tbody>
</table>

Undesired Events

- Detects but does not report \( p2 \leq p1 \).

Effects

\[
s_{dma} < e_{dma} \Rightarrow (\forall i \in [s_{dma}, e_{dma}]) (HardwarePage[i].cAllocated = false \land HardwarePage[i].cZone = dma)
\]

\[
s_{dom} < e_{dom} \Rightarrow (\forall i \in [s_{dom}, e_{dom}]) (HardwarePage[i].cAllocated = false \land HardwarePage[i].cZone = dom)
\]

Dictionary

\[
s_{dma} = \min(p1, mMaxDmaPfn)
\]

\[
e_{dma} = \min(p2, mMaxDmaPfn)
\]

\[
s_{dom} = \max(p1, mMaxDmaPfn)
\]

\[
e_{dom} = \max(p2, mMaxDmaPfn)
\]

Issues

- Both init_domheap_pages and end_boot_allocator (in distinct modules) know that the dma zone goes below mMaxDmaPfn and dom zone goes above it. Why can’t this knowledge be restricted to one module?
- This doesn’t yet deal with “rounding” addresses up and down, nor with leaving one page buffer between xen and dom zones.
- Which module has the responsibility to know the location of the xen heap?
6.2 **assign_pages**

Assign pages to domain.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>int</td>
<td><em>tReturnValue</em></td>
</tr>
<tr>
<td>p1</td>
<td>IO</td>
<td>struct domain *</td>
<td>Guest domain.</td>
</tr>
<tr>
<td>p2</td>
<td>IO</td>
<td>struct page_info[]</td>
<td>Pages of virtual memory.</td>
</tr>
<tr>
<td>p3</td>
<td>I</td>
<td>int</td>
<td>Order</td>
</tr>
<tr>
<td>p4</td>
<td>I</td>
<td>int</td>
<td>Memory flags.</td>
</tr>
</tbody>
</table>

**Effects**

\[(\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage)( (p = page2Mfn(p2[i])) \Rightarrow \]

\[
\begin{align*}
&\neg p_1.mDying \quad \begin{cases} \text{true} & p_1.mTotPages + 2^{p_3} > p_1.mMaxPages \\ \land \text{MEMF\_no\_refcount} \notin p_4 \end{cases} \\
&\quad \begin{cases} \text{false} & p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages \\ \lor \text{MEMF\_no\_refcount} \in p_4 \end{cases} \\
&\quad (p.cPageOwner' = p.cPageOwner) \\
&\quad (p.cDomAllocated' = p.cDomAllocated) \\
&\quad (p.cRefCount' = p.cRefCount) \\
&\quad (\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage)( (p = page2Mfn(p2[i])) \Rightarrow \]

\[
\begin{align*}
&\neg p_1.mDying \quad \begin{cases} \text{true} & p_1.mTotPages + 2^{p_3} > p_1.mMaxPages \\ \land \text{MEMF\_no\_refcount} \notin p_4 \end{cases} \\
&\quad \begin{cases} \text{false} & p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages \\ \lor \text{MEMF\_no\_refcount} \in p_4 \end{cases} \\
&\quad (p1.mTotPages' = p1.mTotPages) \\
&\quad (\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage)( (p = page2Mfn(p2[i])) \Rightarrow \]

\[
\begin{align*}
&\neg p_1.mDying \quad \begin{cases} \text{true} & p_1.mTotPages + 2^{p_3} > p_1.mMaxPages \\ \land \text{MEMF\_no\_refcount} \notin p_4 \end{cases} \\
&\quad \begin{cases} \text{false} & p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages \\ \lor \text{MEMF\_no\_refcount} \in p_4 \end{cases} \\
&\quad (p1.mTotPages' = p1.mTotPages + 2^{p_3}) \\
&\quad (\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage)( (p = page2Mfn(p2[i])) \Rightarrow \]

\[
\begin{align*}
&\neg p_1.mDying \quad \begin{cases} \text{true} & p_1.mTotPages + 2^{p_3} > p_1.mMaxPages \\ \land \text{MEMF\_no\_refcount} \notin p_4 \end{cases} \\
&\quad \begin{cases} \text{false} & p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages \\ \lor \text{MEMF\_no\_refcount} \in p_4 \end{cases} \\
&\quad (tReturnValue' = -1) \\
&\quad (\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage)( (p = page2Mfn(p2[i])) \Rightarrow \]

**Dictionary**

<table>
<thead>
<tr>
<th>tReturnValue</th>
<th>true</th>
<th>false</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1.mDying</td>
<td>\neg p1.mDying</td>
<td>p1.mDying</td>
</tr>
<tr>
<td>p1.mTotPages</td>
<td>p1.mTotPages + 2^{p_3} &gt; p1.mMaxPages &amp; MEMF_no_refcount \notin p4</td>
<td>p1.mTotPages + 2^{p_3} \leq p1.mMaxPages \lor MEMF_no_refcount \in p4</td>
</tr>
</tbody>
</table>

**Issues**

- What does the call to wmb(), a macro defined in system.h, do?
- Not quite capturing !(memflags & MEMF\_no\_refcount)
6.3 _alloc_domheap_pages

Allocate heap pages for a domain.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>struct page_info *</td>
<td>tDomHeapPages</td>
</tr>
<tr>
<td>p1</td>
<td>IO</td>
<td>struct domain *</td>
<td>Guest domain.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned int</td>
<td>CPU</td>
</tr>
<tr>
<td>p3</td>
<td>I</td>
<td>unsigned int</td>
<td>Order</td>
</tr>
<tr>
<td>p4</td>
<td>I</td>
<td>unsigned int</td>
<td>Memory flags.</td>
</tr>
</tbody>
</table>

Undesired Events

Effects

\[
\neg p_1.mDying \land
\begin{aligned}
& (MEMF_{dma} \notin p4 \lor \\
& p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages)
\end{aligned}
\Rightarrow
\begin{aligned}
& (\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory) ( \\
& (p.mAddress in [p.mAddress, p.mAddress + 2^{p_3} - 1] \\
& \land \hat{p}.cAllocated = false \land \hat{p}.cZone = dom) \Rightarrow \\
& \hat{p}.cAllocated' = true)
\end{aligned}
\]

\[
\neg p_1.mDying \land
\begin{aligned}
& (MEMF_{dma} \notin p4 \lor \\
& p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages)
\end{aligned}
\Rightarrow
\begin{aligned}
& (\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory) ( \\
& (p.mAddress in [p.mAddress, p.mAddress + 2^{p_3} - 1] \\
& \land \hat{p}.cAllocated = false \land \hat{p}.cZone = dom) \land \\
& p_3 \leq mMaxOrder \\
& \land \\
& tNumAvailDmaPages \geq DmaEmergencyPoolPages + 2^{p_3}
\end{aligned}
\]

\[
\neg p_1.mDying \land
\begin{aligned}
& MEMF_{dma} \in p4 \land \\
& p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages
\end{aligned}
\Rightarrow
\begin{aligned}
& (\exists p \in HardwareMemory)(\forall \hat{p} \in HardwareMemory) ( \\
& (p.mAddress in [p.mAddress, p.mAddress + 2^{p_3} - 1] \\
& \land \hat{p}.cAllocated = false \land \hat{p}.cZone = dma) \Rightarrow \\
& \hat{p}.cAllocated' = true)
\end{aligned}
\]

\[(\forall i \in [0, 2^{p_3} - 1])(\exists p \in HardwarePage) ( \\
(tDomHeapPage' \neq null \land p = page2Mfn(tDomHeapPages'[i])) \Rightarrow \\
\begin{array}{|c|c|c|}
\hline
p_1.mDying & true & false \\
\hline
\neg p_1.mDying & \begin{aligned}
& p_1.mTotPages + 2^{p_3} > p_1.mMaxPages \\
& \land MEMF_{no_refcount} \notin p4
\end{aligned} & \begin{aligned}
& p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages \\
& \lor MEMF_{no_refcount} \in p4
\end{aligned} \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|
\hline
p.cPageOwner' = & p.cPageOwner & p1 \\
p.cDomAllocated' = & p.cDomAllocated & true \\
p.cRefCount' = & p.cRefCount & 1 \\
\hline
\end{array}
\)
2

[p1.mDying] true false

¬ p1.mDying p1.mTotPages + 2^{p3} > p1.mMaxPages p1.mTotPages + 2^{p3} ≤ p1.mMaxPages

∨ MEMF_no refcount ∈ p4 ∧ MEMF_no refcount /∈ p4

[p1.cTotPages] = p1.cTotPages p1.cTotPages + 2^{p3}

Dictionary

tNumAvailDmaPages unsigned long

Does this have to be redundant with the definition in avail_heap_pages()?

\[
\text{tNumAvailDmaPages} = \{ (\forall p \in \text{HardwareMemory})(p.cAllocated = false \land (p.cZone = dma)) \}\]

tDomHeapPages struct page_info *

\[
(p1 = \text{null} \lor (\neg p1.mDying \land \text{MEMF_no refcount} \in p4 \lor p1.mTotPages + 2^{p3} \leq p1.mMaxPages)) \land \text{MEMF_dma} \notin p4 \lor p3 \leq mMaxOrder) \lor

(\exists p \in \text{HardwareMemory}) ((\forall p \in \text{HardwareMemory}) (p.mAddress in [p.mAddress, p.mAddress + 2^{p3} - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = dom) \Rightarrow tDomHeapPages' = mfn2Page(p.mAddress))

\[
(p1 = \text{null} \lor (\neg p1.mDying \land \text{MEMF_no refcount} \in p4 \lor p1.mTotPages + 2^{p3} \leq p1.mMaxPages)) \land \text{MEMF_dma} \in p4 \land p3 \leq mMaxOrder) \lor

(\exists p \in \text{HardwareMemory}) ((\forall p \in \text{HardwareMemory}) (p.mAddress in [p.mAddress, p.mAddress + 2^{p3} - 1] \land \hat{p}.cAllocated = false \land \hat{p}.cZone = dom) \Rightarrow tDomHeapPages' = mfn2Page(p.mAddress))

\[
(p1.mDying \lor p3 > mMaxOrder) \lor

(\text{MEMF_no refcount} \notin p4 \land p1.mTotPages + 2^{p3} \leq p1.mMaxPages) \lor

(tNumAvailDmaPages < DmaEmergencyPoolPages + 2^{p3})

\Rightarrow tDomHeapPages' = null

\]

26
Alternative representation of the value of $t_{DomHeapPages}$.

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(p_1 = \text{null}) \lor (-p_1.mDying \land \text{MEMF}_{\text{no refcount}} \in p_4 \lor p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages)) \land p_3 \leq mMaxOrder</td>
<td>(\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_3} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dom}) \Rightarrow t_{DomHeapPages}' = mfn2Page(p.mAddress)</td>
</tr>
<tr>
<td>MEMF_{\text{dma}} \not\in p_4</td>
<td>(\forall p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_3} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dom}) \land t_{\text{NumAvailDmaPages}} \geq DmaEmergencyPoolPages + 2^{p_3} \land (\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_3} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dma}) \Rightarrow t_{DomHeapPages}' = mfn2Page(p.mAddress)</td>
</tr>
</tbody>
</table>

MEMF_{\text{dma}} \in p_4 | (\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_3} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dma}) \Rightarrow t_{DomHeapPages}' = mfn2Page(p.mAddress) |

$p_1 \not= \text{null} \land p_1.mDying \lor p_3 > mMaxOrder \lor (\text{MEMF}_{\text{no refcount}} \not\in p_4 \land p_1.mTotPages + 2^{p_3} \leq p_1.mMaxPages) \lor (t_{\text{NumAvailDmaPages}} < DmaEmergencyPoolPages + 2^{p_3}) | \text{true} |

$t_{DomHeapPages}' = \text{null}$

Issues
6.4 alloc_domheap_pages

Allocate heap pages for a domain.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>struct page_info</td>
<td>tDomHeapPages</td>
</tr>
<tr>
<td>p1</td>
<td>IO</td>
<td>struct domain</td>
<td>Guest domain</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned int</td>
<td>Order</td>
</tr>
<tr>
<td>p3</td>
<td>I</td>
<td>unsigned int</td>
<td>Memory flags</td>
</tr>
</tbody>
</table>

Undesired Events

Effects

\[
\neg p_1.mDying \land (MEMF_dma \notin p3 \lor p_1.mTotPages + 2^{p_2} \leq p_1.mMaxPages) \Rightarrow (\exists p \in HardwareMemory)(\forall p \in HardwareMemory) (\neg p.mDying \land p.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land p.cAllocated = false \land p.cZone = dom) \Rightarrow \neg p.cAllocated' = true
\]

\[
\neg p_1.mDying \land (MEMF_dma \notin p3 \lor p_1.mTotPages + 2^{p_2} \leq p_1.mMaxPages) \Rightarrow (\exists p \in HardwareMemory)(\forall p \in HardwareMemory) (\neg p.mDying \land p.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land p.cAllocated = false \land p.cZone = dom) \land p_2 \leq mMaxOrder \land tNumAvailDmaPages \geq DmaEmergencyPoolPages + 2^{p_2} \land (\exists p \in HardwareMemory)(\forall p \in HardwareMemory) (\neg p.mDying \land p.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land p.cAllocated = false \land p.cZone = dma) \Rightarrow \neg p.cAllocated' = true
\]

\[
\neg p_1.mDying \land MEMF_dma \in p3 \land p_1.mTotPages + 2^{p_2} \leq p_1.mMaxPages \Rightarrow (\exists p \in HardwareMemory)(\forall p \in HardwareMemory) (\neg p.mDying \land p.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land p.cAllocated = false \land p.cZone = dma) \Rightarrow p.cAllocated' = true
\]

\[
(\forall i \in [0, 2^{p_2} - 1])(\exists p \in HardwarePage) (tDomHeapPage' \neq null \land p = page2Mfn(tDomHeapPages'[i])) \Rightarrow
\]

\[
\begin{array}{|c|c|c|}
\hline
p_1.mDying & true & false \\
\neg p_1.mDying & p_1.mTotPages + 2^{p_2} > p_1.mMaxPages \land MEMF_noRefCount \notin p3 & p_1.mTotPages + 2^{p_2} \leq p_1.mMaxPages \lor MEMF_noRefCount \in p3 \\
p.cPageOwner' = & p.cPageOwner & p_1 \\
p.cDomAllocated' = & p.cDomAllocated & true \\
p.cRefCount' = & p.cRefCount & 1 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
p_1.mDying & true & false \\
\neg p_1.mDying & p_1.mTotPages + 2^{p_2} > p_1.mMaxPages \lor MEMF_noRefCount \in p3 & p_1.mTotPages + 2^{p_2} \leq p_1.mMaxPages \land MEMF_noRefCount \notin p3 \\
p_1.cTotPages' = & p_1.cTotPages & p_1.cTotPages + 2^{p_2} \\
\hline
\end{array}
\]
Dictionary

\[ t_{\text{NumAvailDmaPages}} \] unsigned long

Does this have to be redundant with the definition in `avail_heap_pages()`?

\[ t_{\text{NumAvailDmaPages}} = |\{ \forall p \in \text{HardwareMemory} (p.c\text{Allocated} = \text{false} \land (p.c\text{Zone} = \text{dma})) \}| \]

\[ t_{\text{DomHeapPages}} \] struct page_info *

<table>
<thead>
<tr>
<th>Equation 1</th>
<th>Equation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p1 = null ∨ (¬p1.mDying ∧ MEMF_no_refcount ∈ p3 ∨ p1.mTotPages + 2^{\text{order}} ≤ p1.mMaxPages)) ∧ MEMF_dma ∉ p3 ∧ p2 ≤ mMaxOrder</td>
<td>(\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress in [p.mAddress, p.mAddress + 2^{\text{order}} - 1] ∧ \hat{p}.c\text{Allocated} = \text{false} ∧ \hat{p}.c\text{Zone} = \text{dom}) ⇒ t_{\text{DomHeapPages}}' = \text{mfn2Page}(p.mAddress)</td>
</tr>
<tr>
<td>(p1 = null ∨ (¬p1.mDying ∧ MEMF_no_refcount ∈ p3 ∨ p1.mTotPages + 2^{\text{order}} ≤ p1.mMaxPages)) ∧ MEMF_dma ∉ p3 ∧ p2 ≤ mMaxOrder</td>
<td>(\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress in [p.mAddress, p.mAddress + 2^{\text{order}} - 1] ∧ \hat{p}.c\text{Allocated} = \text{false} ∧ \hat{p}.c\text{Zone} = \text{dom}) ∧ t_{\text{NumAvailDmaPages}} ≥ DmaEmergencyPoolPages + 2^{\text{order}} ∧ (\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress in [p.mAddress, p.mAddress + 2^{\text{order}} - 1] ∧ \hat{p}.c\text{Allocated} = \text{false} ∧ \hat{p}.c\text{Zone} = \text{dma}) ⇒ t_{\text{DomHeapPages}}' = \text{mfn2Page}(p.mAddress)</td>
</tr>
<tr>
<td>p1.mDying ∨ p2 &gt; mMaxOrder ∨ (MEMF_no_refcount ∉ p3 ∧ p1.mTotPages + 2^{\text{order}} ≤ p1.mMaxPages) ∨ (t_{\text{NumAvailDmaPages}} &lt; DmaEmergencyPoolPages + 2^{\text{order}})</td>
<td>(\exists p \in \text{HardwareMemory}) (\forall \hat{p} \in \text{HardwareMemory}) (\hat{p}.mAddress in [p.mAddress, p.mAddress + 2^{\text{order}} - 1] ∧ \hat{p}.c\text{Allocated} = \text{false} ∧ \hat{p}.c\text{Zone} = \text{dma}) ⇒ t_{\text{DomHeapPages}}' = \text{mfn2Page}(p.mAddress)</td>
</tr>
</tbody>
</table>

| t_{\text{DomHeapPages}}' = null | 
Alternative representation of the value of \( t_{DomHeapPages} \).

<table>
<thead>
<tr>
<th>( p_1 = \text{null} \lor )</th>
<th>( \neg p_1.mDying \land )</th>
<th>( \text{MEMF,no_refcount} \notin \text{p3} \lor )</th>
<th>( p_1.mTotPages + 2^{p_2} \leq )</th>
<th>( p_1.mMaxPages )</th>
<th>( \text{MEMF,dma} \notin \text{p3} )</th>
<th>( (\exists p \in \text{HardwareMemory}) )</th>
<th>( (\forall \hat{p} \in \text{HardwareMemory}) )</th>
<th>( (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dom}) )</th>
<th>( \Rightarrow )</th>
<th>( t_{DomHeapPages}' = \text{mfn2Page}(p.mAddress) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 \neq \text{null} \land )</td>
<td>( p_1.mDying \lor )</td>
<td>( p_2 &gt; \text{mMaxOrder} \lor )</td>
<td>( (\text{MEMF,no_refcount} \notin \text{p3} \land )</td>
<td>( p_1.mTotPages + 2^{p_2} \leq )</td>
<td>( p_1.mMaxPages )</td>
<td>( \text{MEMF,dma} \in \text{p3} \land )</td>
<td>( (\exists p \in \text{HardwareMemory}) )</td>
<td>( (\forall \hat{p} \in \text{HardwareMemory}) )</td>
<td>( (\hat{p}.mAddress \in [p.mAddress, p.mAddress + 2^{p_2} - 1] \land \hat{p}.cAllocated = \text{false} \land \hat{p}.cZone = \text{dma}) )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \text{true} )</td>
<td>( \text{true} )</td>
<td>( t_{DomHeapPages}' = \text{null} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Issues**
6.5 avail_domheap_pages

How many unused pages?

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>unsigned long</td>
<td>tNumAvailDomHeapPages</td>
</tr>
</tbody>
</table>

Undesired Events

- uBadZone
- uNotInitialized

Effects

None.

Dictionary

tNumAvailDomHeapPages unsigned long

tNumAvailDomHeapPages' = |{(∀p ∈ HardwareMemory)(p.cAllocated = false ∧ (p1.cZone = dom ∨ p1.cZone = dma))}|  

Issues

- UEs neither detected nor reported.
- This specification does not address dma_emergency_pool_pages.
6.6  free_domheap_pages

Put block of domain heap pages in free space.

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>I</td>
<td>struct page_info *</td>
<td>Page.</td>
</tr>
<tr>
<td>p2</td>
<td>I</td>
<td>unsigned int</td>
<td>Order of pages.</td>
</tr>
</tbody>
</table>

Undesired Events

• uRequestTooLarge
• uBadZone

Effects

\[-(p1.cPageOwner.mDying) \Rightarrow \\
(\forall p \in HardwareMemory) (p.mAddress in [page2M fn(p1), page2M fn(p1) + 2^{p2} - 1] \Rightarrow \ p.cAllocated' = false)\]

\[(p1.cPageOwner.mDying) \Rightarrow \\
(\forall p \in HardwareMemory) (p.mAddress in [page2M fn(p1), page2M fn(p1) + 2^{p2} - 1] \Rightarrow \ p.cScrubMe' = true)\]

Issues

• free_domheap_pages() and page_scrub_softirq() share a data structure, scrub_page_list and scrub_pages, which keeps track of pages freed by dying domains which Xen needs to scrub.
• Function does not detect either UE.
• When can the zone a block belongs to change?
• I assume that the user of this function is not concerned with the merging of blocks of memory into larger blocks, nor with the algorithms and data structures involved in managing and implementing such merging.

Of course, the implementor is.
7 Page Scrubbing

The Page Scrubbing module hides when pages are cleared.

7.1 page_scrub_softirq

Zeroize some pages.

Effects

\[(\forall p \in \text{HardwareMemory})(p.cScrubMe = true) \Rightarrow (p.cScrubMe' = false \land p.cAllocated' = false \land p.cZeroized' = true))\]

Issues

- free_domheap_pages() and page_scrub_softirq() share a data structure, scrub_page_list and scrub_pages, which keeps track of pages freed by dying domains which Xen needs to scrub.
7.2 avail_scrub_pages

How many pages to zeroize?

<table>
<thead>
<tr>
<th>Parameter #</th>
<th>Mode</th>
<th>Type</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p0</td>
<td>O</td>
<td>unsigned long</td>
<td>tNumPagesToScrub</td>
</tr>
</tbody>
</table>

Effects

None.

Dictionary

tNumPagesToScrub unsigned long

\[
t\text{NumPagesToScrub}' = |\{p \in \text{HardwareMemory} | (p.cScrubMe = \text{true})\}|
\]