Lecture 11
Storage Allocation

Saman Amarasinghe
2013
Outline

Memory Systems Revisited
Stacks
Fixed-Size Heap Allocation
Variable-Size Heap Allocation
Garbage Collection by Reference Counting
Stop-and-Copy Garbage Collection
Intel® Core™ Microarchitecture – Memory Sub-system

Intel Core 2 Quad Processor

<table>
<thead>
<tr>
<th>L1 Data Cache</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Line Size</td>
<td>Latency</td>
<td>Associativity</td>
</tr>
<tr>
<td>32 KB</td>
<td>64 bytes</td>
<td>3 cycles</td>
<td>8-way</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L1 Instruction Cache</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Line Size</td>
<td>Latency</td>
<td>Associativity</td>
</tr>
<tr>
<td>32 KB</td>
<td>64 bytes</td>
<td>3 cycles</td>
<td>8-way</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L2 Cache</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Line Size</td>
<td>Latency</td>
<td>Associativity</td>
</tr>
<tr>
<td>6 MB</td>
<td>64 bytes</td>
<td>14 cycles</td>
<td>24-way</td>
</tr>
</tbody>
</table>
**Intel® Nehalem™ Microarchitecture – Memory Sub-system**

## Intel 6 Core Processor

<table>
<thead>
<tr>
<th>Cache Level</th>
<th>Size</th>
<th>Line Size</th>
<th>Latency</th>
<th>Associativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Data Cache</td>
<td>32 KB</td>
<td>64 bytes</td>
<td>4 ns</td>
<td>8-way</td>
</tr>
<tr>
<td>L1 Instruction Cache</td>
<td>32 KB</td>
<td>64 bytes</td>
<td>4 ns</td>
<td>4-way</td>
</tr>
<tr>
<td>L2 Cache</td>
<td>256 KB</td>
<td>64 bytes</td>
<td>10 ns</td>
<td>8-way</td>
</tr>
<tr>
<td>L3 Cache</td>
<td>12 MB</td>
<td>64 bytes</td>
<td>50 ns</td>
<td>16-way</td>
</tr>
<tr>
<td>Main Memory</td>
<td>64 bytes</td>
<td>75 ns</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2013 Charles E. Leiserson, Saman P. Amarasinghe and I–Ting Angelina Lee
for(rep=0; rep < REP; rep++)
    for(a=0; a < N ; a++)

Capacity misses if larger than the cache at each level
for(rep=0; rep < REP; rep++)
for(a=0; a < N ; a++)

Amazing prefetcher

Single core cannot saturate the memory system
mask = (1<<n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523)*253573) & mask;
}
mask = (1<<n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523)*253573) & mask;
}
mask = (1 << n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523) * 253573) & mask;
}
mask = (1 << n) - 1;
for(rep = 0; rep < REP; rep++) {
    addr = ((rep + 523) * 253573) & mask;
}
```c
mask = (1 << n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523) * 253573) & mask;
}
```
mask = (1 << n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523) * 253573) & mask;
}
Virtual Memory System

You access virtual memory, your computer has physical memory & disk

- $2^{64}$ virtual memory
- Limited physical memory
- All memory backed up on disk

Virtual to physical mapped by pages

- x86 supports 4K small and 4M large pages

OS Manages Virtual memory

- Allocates virtual pages, maps them to physical
- Backs pages on disk and bring them in and out
- Provides a page table to the hardware

Hardware caches the entries in the TLB

When you access a memory location

- If that page is mapped to physical memory and the mapping is cached in TLB $\rightarrow$ aok ($\sim 1$ cycle)
- If mapping is not in TLB $\rightarrow$ TLB miss. ($\sim 100$ cycles)
  - The HW gets the mapping from the page table and caches it in TLB
- If page is not mapped $\rightarrow$ Page fault. ($\sim 1,000,000$ cycles)
  - The OS has to get involved in bringing in the page to physical memory from disk and updating the page table
mask = (1 << n) - 1;
for(rep=0; rep < REP; rep++) {
    addr = ((rep + 523) * 253573) & mask;
}
TLB

- Page size is 4 KB
- Number of TLB entries is 512
- So, total memory that can be mapped by TLB is 2 MB
- L3 cache is 12 MB!
- TLB misses before L3 cache misses!
STACKS
Stack Allocation

Array and pointer

Allocate $x$ bytes

```
sp += x;
return sp - x;
```
Stack Allocation

Array and pointer

Allocate \( x \) bytes

\[
\begin{align*}
\text{sp} & \leftarrow \text{sp} + x; \\
\text{return sp} - x;
\end{align*}
\]

How does check for stack overflow?
Guard Pages

Surround the page with two unallocated pages (guard pages)

Pros
- Cost of bounds check is zero

Cons
- Bounds violation (a page fault) is very costly
- Data sizes has to be a multiple of available page size
  - 4KB or 4MB
- If offset > guard page size, may not catch the violation
Stack Deallocation

Array and pointer

Allocate $x$ bytes

$$\text{sp} += x;$$
$$\text{return } \text{sp} - x;$$

Free $x$ bytes

$$\text{sp} -= x;$$
Stack Deallocation

Array and pointer

Allocate $x$ bytes

Free $x$ bytes

\[
\text{sp} += x; \\
\text{return } \text{sp} - x;
\]

\[
\text{sp} -= x;
\]

Should check for stack underflow.
Stack Storage

Array and pointer

Allocate x bytes

```
sp += x;
return sp - x;
```

Free x bytes

```
sp -= x;
```

- Allocating and freeing take $\Theta(1)$ time.
- Must free consistent with stack discipline.
- Limited applicability, but great when it works!
- One can allocate on the call stack using `alloca()`, but this function is deprecated, and the compiler is more efficient with fixed-size frames.
FIXED-SIZE HEAP ALLOCATION
Heap Allocation*

C provides `malloc()` and `free()`. C++ provides `new` and `delete`.

Unlike Java and Python, C and C++ provide no garbage collector. Heap storage allocated by the programmer must be freed explicitly. Failure to do so creates a memory leak. Also, watch for dangling pointers and double freeing.

Memory checkers can assist in finding these pernicious bugs:

```
% valgrind --leakcheck=yes ./myprog <arguments>
```

Valgrind is installed on cloud machines. See [http://valgrind.org](http://valgrind.org) for details.

*Do not confuse with a heap data structure.
Fixed-Size Allocation

Free list

A

used
used used

•

used

•

free
Fixed-Size Allocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free}\rightarrow\text{next}; \\
\text{return } x;
\]
Fixed-Size Allocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free} \rightarrow \text{next}; \\
\text{return } x;
\]
Fixed-Size Allocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free->next}; \\
\text{return } x;
\]

Should check \text{free} \neq \text{NULL}.
Fixed-Size Allocation

Free list

Allocate 1 object

\[ x = \text{free}; \]
\[ \text{free} = \text{free}\rightarrow\text{next}; \]
\[ \text{return } x; \]
**Fixed-Size Allocation**

**Free list**

```
Allocate 1 object
```

```c
x = free;
free = free->next;
return x;
```
Fixed-Size Deallocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free} \rightarrow \text{next}; \\
\text{return } x;
\]

Free object \(x\)

\[
x \rightarrow \text{next} = \text{free}; \\
\text{free} = x;
\]
Fixed-Size Deallocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free}->\text{next}; \\
\text{return } x;
\]

Free object \( x \)

\[
x->\text{next} = \text{free}; \\
\text{free} = x;
\]
Fixed-Size Deallocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free}\rightarrow\text{next}; \\
\text{return } x;
\]

Free object x

\[
x\rightarrow\text{next} = \text{free}; \\
\text{free} = x;
\]
Fixed-SizeDeallocation

Free list

Allocate 1 object

\[
x = \text{free}; \\
\text{free} = \text{free} \to \text{next}; \\
\text{return } x;
\]

Free object x

\[
x \to \text{next} = \text{free}; \\
\text{free} = x;
\]
Free Lists

Free list

- Allocating and freeing take $\Theta(1)$ time.
- Good temporal locality.
- Poor spatial locality due to external fragmentation — blocks distributed across virtual memory — which can increase the size of the page table and cause disk thrashing.
- The translation lookaside buffer (TLB) can also be a problem.
Mitigating External Fragmentation

- Keep a free list per disk page.
- Allocate from the free list for the fullest page.
- Free a block of storage to the free list for the page on which the block resides.
- If a page becomes empty (only free-list items), the virtual-memory system can page it out without affecting program performance.

90–10 is better than 50–50:

\[
\text{Probability that 2 random accesses hit the same page} = .9 \times .9 + .1 \times .1 = .82 \text{ versus } .5 \times .5 + .5 \times .5 = .5
\]
VARIABLE-SIZE HEAP ALLOCATION
### Variable-Size Allocation

**Binned free lists**
- Leverage the efficiency of free lists.
- Accept a bounded amount of internal fragmentation.

Bin $k$ holds memory blocks of size $2^k$. 

![Diagram showing binned free lists](image-url)
Allocate \( x \) bytes

- If bin \( k = \lceil \lg x \rceil \) is nonempty, return a block.
- Otherwise, find a block in the next larger nonempty bin \( k' > k \), split it up into blocks of sizes \( 2^{k'-1}, 2^{k'-2}, \ldots, 2^k, 2^k \), and distribute the pieces.

Example

\( x = 3 \Rightarrow \lceil \lg x \rceil = 2 \).

Bin 2 is empty.
Allocate $x$ bytes

- If bin $k = \lceil \lg x \rceil$ is nonempty, return a block.
- Otherwise, find a block in the next larger nonempty bin $k' > k$, split it up into blocks of sizes $2^{k'-1}, 2^{k'-2}, \ldots, 2^k, 2^k$, and distribute the pieces.

Example

$x = 3 \Rightarrow \lceil \lg x \rceil = 2.$
Bin 2 is empty.
Allocate $x$ bytes

- If bin $k = \lceil \log x \rceil$ is nonempty, return a block.
- Otherwise, find a block in the next larger nonempty bin $k' > k$, split it up into blocks of sizes $2^{k'-1}, 2^{k'-2}, \ldots, 2^k, 2^k$, and distribute the pieces.

**Example**

$x = 3 \Rightarrow \lceil \log x \rceil = 2$.

Bin 2 is empty.

```
return
```
Allocation for Binned Free Lists

Allocate \( x \) bytes

- If bin \( k = \lfloor \lg x \rfloor \) is nonempty, return a block.
- Otherwise, find a block in the next larger nonempty bin \( k' > k \), split it up into blocks of sizes \( 2^{k'-1}, 2^{k'-2}, \ldots, 2^k, 2^k \), and distribute the pieces.*

Example

\( x = 3 \Rightarrow \lfloor \lg x \rfloor = 2 \).
Bin 2 is empty.

*If no larger blocks exist, ask the OS to allocate \( x \) more bytes of VM.
Storage Layout of a Program

- **Stack**: Dynamically allocated, initialized to 0 at program start.
- **Heap**: Dynamically allocated.
- **BSS**: Initialized to 0 at program start.
- **Data**: Read from disk.
- **Text**: Code.

**Virtual Memory**

- **High Address**
- **Low Address**
Q. Since a 64-bit address space takes over a century to write at a rate of 4 billion bytes per second, we effectively never run out of virtual memory. Why not just allocate out of virtual memory and never free?

A. **External fragmentation** would be horrendous! The performance of the page table would degrade tremendously leading to **disk thrashing**, since all nonzero memory must be backed up on disk in page-sized blocks.

**Goal of storage allocators**
Use as little virtual memory as possible, and try to keep the used portions relatively compact.
Theorem. Suppose that the maximum amount of heap memory in use at any time by a program is M. If the heap is managed by a BFL allocator, the amount of virtual memory consumed by heap storage is \( O(M \lg M) \).

Proof. An allocation request for a block of size \( x \) consumes \( 2^{\lceil \lg x \rceil} \leq 2x \) storage. Thus, the amount of virtual memory devoted to blocks of size \( 2^k \) is at most \( 2M \). Since there are at most \( \lg M \) free lists, the theorem holds. \( \blacksquare \)
Coalescing

Binned free lists can sometimes be heuristically improved by splicing together adjacent small blocks into a larger block.

- Clever schemes exist for finding adjacent blocks efficiently — e.g., the “buddy” system — but the overhead is still greater than simple BFL.
- No good theoretical bounds exist that prove the effectiveness of coalescing.
- Coalescing seems to work in practice, because heap storage tends to be deallocated as a stack (LIFO) or in batches.
GARBAGE COLLECTION
BY REFERENCE COUNTING
Garbage Collectors

Idea
- Free the programmer from freeing objects.
- A garbage collector identifies and recycles the objects that the program can no longer access.
- GC can be built-in (Java, Python) or do-it-yourself.

- Any other Pros and Cons?
Terminology

- **Roots** are objects directly accessible by the program (globals, stack, etc.).
- **Live** objects are reachable from the roots by following pointers.
- **Dead** objects are inaccessible and can be recycled.

How can the GC identify pointers in objects?

- **Strong typing.**
- **Prohibit** pointer arithmetic (which may slow down some programs).
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.

root

root

root

© 2013 Charles E. Leiserson, Saman P. Amarasinghe and I–Ting Angelina Lee
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.

root

root

root
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.
Reference Counting

Keep a count of the number of pointers referencing each object. If the count drops to 0, free the dead object.

What are the pros and cons of reference counting?
Limitation of Reference Counting

Problem
A cycle is never garbage collected!
Limitation of Reference Counting

Problem
A cycle is never garbage collected!
Limitation of Reference Counting

Problem
A cycle is never garbage collected!

- root
- root
- root
Limitation of Reference Counting

Problem
A cycle is never garbage collected!

Nevertheless, reference counting works well for acyclic structures.

Uncollected garbage stinks!
STOP–&–COPY
GARBAGE COLLECTION

SPEED LIMIT

PER ORDER OF 6.172
**Graph Abstraction**

**Idea**
Objects and pointers form a directed graph \( G = (V, E) \). Live objects are reachable from the roots. Use breadth-first search to find the live objects.

\[
\text{FIFO queue } Q
\]

```plaintext
for (\forall v \in V) \{
    if (\text{root}(v)) {
        v.mark = 1;
        enqueue(Q, v);
    } else v.mark = 0;

while (Q != \emptyset) {
    u = dequeue(Q);
    for (\forall v \in V \text{ such that } (u, v) \in E) {
        if (v.mark == 0) {
            v.mark = 1;
            enqueue(Q, v);
        }
    }
}
```

© 2013 Charles E. Leiserson, Saman P. Amarasinghe and I–Ting Angelina Lee
Breadth-First Search

A graph with nodes labeled a through j and arrows connecting them. A queue labeled Q is shown with a head and tail indicator.
Breadth-First Search
Breadth-First Search

Q

head
tail

© 2013 Charles E. Leiserson, Saman P. Amarasinghe and I-Ting Angelina Lee
Breadth-First Search

Q

r b

head tail
Breadth-First Search

Q: [r, b, c]

head tail
Breadth-First Search

A Breadth-First Search (BFS) is an algorithm for traversing or searching tree or graph data structures. It starts at the root node (selecting some arbitrary node as the root node in the case of a graph) and explores the neighbor nodes at the present depth prior to moving on to nodes at the next depth level.

In the diagram, the algorithm starts at node r and explores its neighbors in breadth-first order: b, c, d, e, f, g, h, i, and j. The queue Q contains the nodes in the order they are visited, with the head at the front and the tail at the back.
Breadth-First Search

Graph with nodes a, b, c, d, e, f, g, h, i, j, and edges connecting them. At the bottom, there is a queue Q with elements r, b, c, and arrows indicating the head and tail positions.
Breadth-First Search

The diagram illustrates a Breadth-First Search (BFS) algorithm. BFS is a graph traversal algorithm that explores all the vertices of a graph in breadth-first order. The algorithm starts at the root node (or any arbitrary starting node) and explores all the neighboring nodes before moving to the next level of nodes. The nodes are visited in the order of their level, starting from the root node and proceeding outward level by level. The diagram shows the nodes and edges of a graph, with a queue (Q) indicating the order in which nodes are processed. The queue initially contains the root node (r) and its direct neighbors (b, c, d) as they are explored in the first level.
Breadth-First Search
Breadth-First Search

\[
\begin{array}{c}
\text{Q} \\
\text{r b c d e} \\
\text{head} \quad \text{tail}
\end{array}
\]
Breadth-First Search

Q

r b c d e

head tail
Breadth-First Search

A diagram of a graph with nodes labeled a, b, c, d, e, f, g, h, i, j, and connections between them. The nodes are organized in a breadth-first manner.

A queue labeled Q contains the nodes r, b, c, d, e, f, and is shaded green to indicate that these are the nodes that have been visited so far. The head and tail of the queue are marked with arrows.
Breadth-First Search

Q

r b c d e f

head tail

© 2013 Charles E. Leiserson, Saman P. Amarasinghe and I-Ting Angelina Lee
Breadth-First Search
Breadth-First Search
Breadth-First Search

Q

\[
\begin{array}{cccccccc}
\text{r} & \text{b} & \text{c} & \text{d} & \text{e} & \text{f} & \text{g} & \text{h} & \text{i} & \text{j}
\end{array}
\]

Done!

head
tail
Breadth-First Search

Observation
All live vertices are placed in contiguous storage in Q.
Copying Garbage Collector

FROM space

dead
live
unused

next allocation
Copying Garbage Collector

FROM space

live
dead
unused

next allocation
FROM space

next allocation

live
dead
unused
Copying Garbage Collector

FROM space

next allocation

live

dead

unused
Copying Garbage Collector

FROM space

next allocation

live

dead

unused
Copying Garbage Collector

FROM space

next allocation

live

dead

unused
When the FROM space is “full,” copy live storage using BFS with the TO space as the FIFO queue.
Copying Garbage Collector

When the FROM space is “full,” copy live storage using BFS with the TO space as the FIFO queue.
Updating Pointers

Since the FROM address of an object is not generally equal to the TO address of the object, pointers must be updated.

- When an object is copied to the TO space, store a forwarding pointer in the FROM object, which implicitly marks it as moved.
- When an object is removed from the FIFO queue in the TO space, update all its pointers.
Example

Remove an item from the queue.
Remove an item from the queue.
Example

Enqueue adjacent vertices.
Enqueue adjacent vertices.
Place forwarding pointers in FROM vertices.
Update the pointers in the removed item to refer to its adjacent items in the TO space.
Update the pointers in the removed item to refer to its adjacent items in the TO space.
Example

FROM

TO

head
tail

Linear time to copy and update all vertices.
Dynamic Storage Allocation

Lots more is known and unknown about dynamic storage allocation. Strategies include

- buddy system,
- mark–and–sweep garbage collection,
- generational garbage collection,
- real–time garbage collection,
- multithreaded storage allocation,
- parallel garbage collection,
- etc.