Using segmentation to build a capability-based single address space operating system

Wuyang Chung

wy-chung@outlook.com

Outline

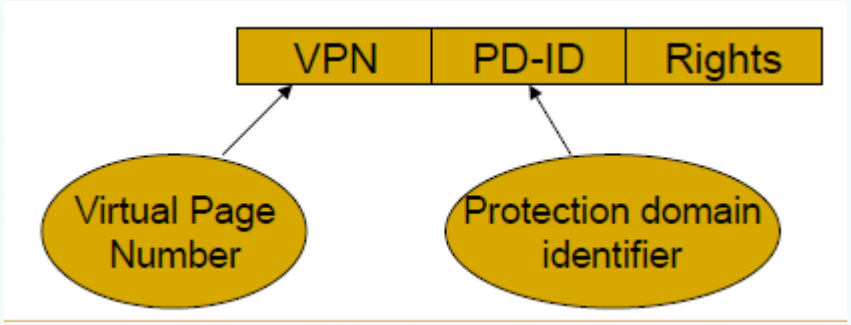
- Single Address Space Operating System
- Capability-based Addressing
- Segmentation
- Benefits of Segmentation
- Compatibility with Old Programs
- Conclusion

Single Address Space OS

- Problems with multiple address space OS
  - Cannot use pointers in shared buffer
  - On platforms without ASID
    - Need to flush TLB during address space switch
  - On platforms with ASID
    - Duplication of translation information in TLB
- Advantages of single address space OS
  - Encourage of sharing memory between protection domains
  - Low context switch overhead
- Challenge of single address space OS
  - Protection

Multiple Protection Domains on a Single Address Space

- Domain page model
  - PLB (Protection Lookaside Buffer)

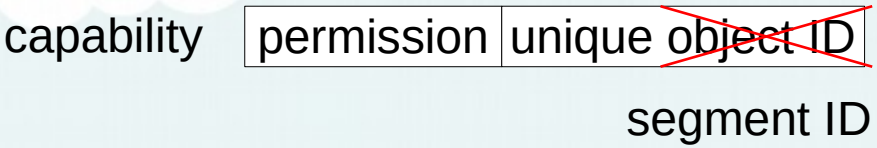


Architecture support for single address space operating systems, ACM SIGPLAN, 1992.

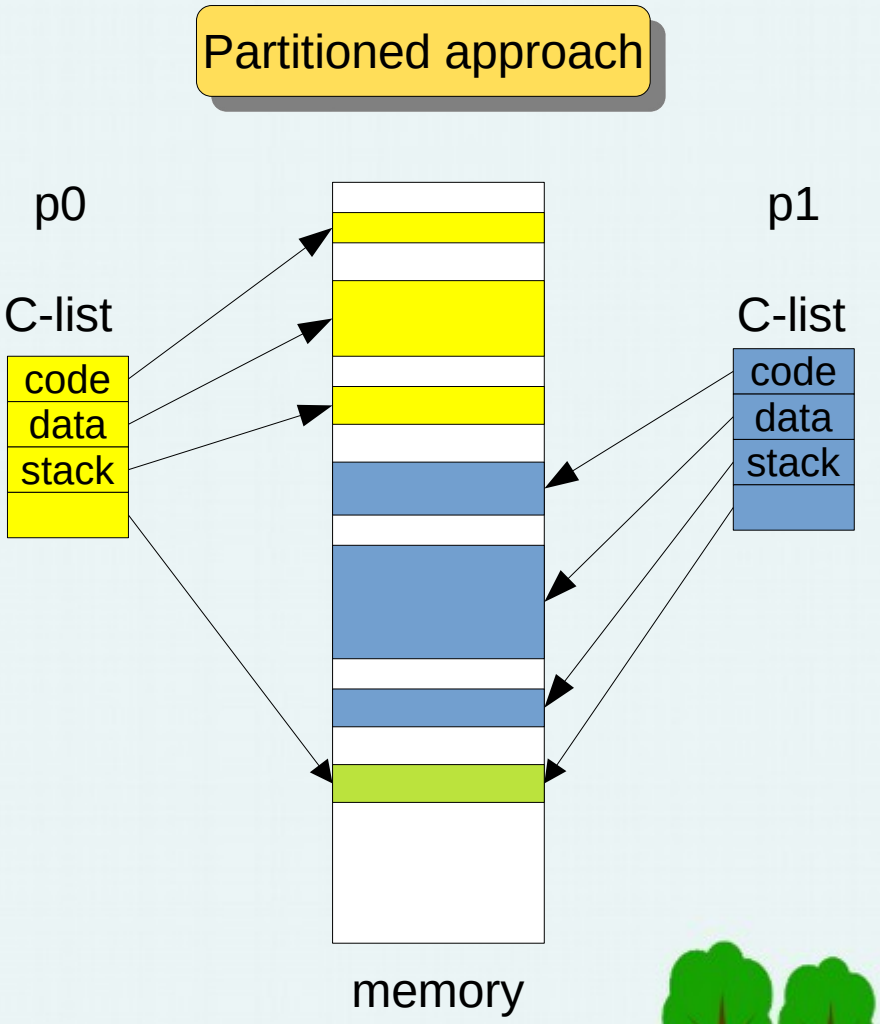
https://cseweb.ucsd.edu/classes/fa11/cse240A-a/Slides1/06_SAOS.pdf

- Page group model
  - There is an AID (access identifier) field in page table entry
  - The CPU has several PID (protection identifier) registers
  - A process can access a page if any of the PIDs is equal to the page AID
- Capability-based addressing

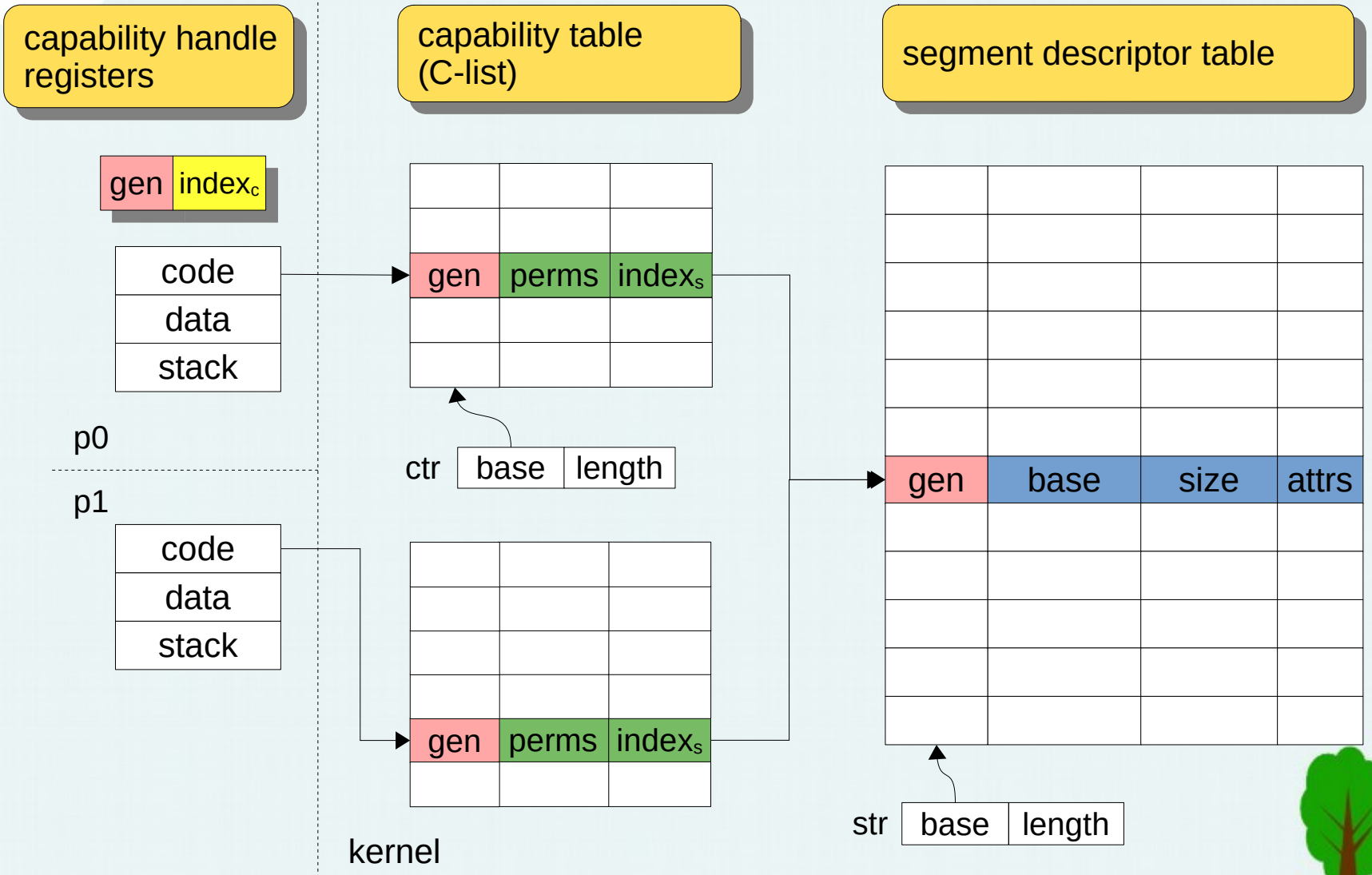
Capability-based Addressing



- The most important thing in a capability system is to prevent capability from being forged
- Two approaches
  - Tagged approach
    - e.g. CHERI
  - Partitioned approach
    - e.g. Plessey 250

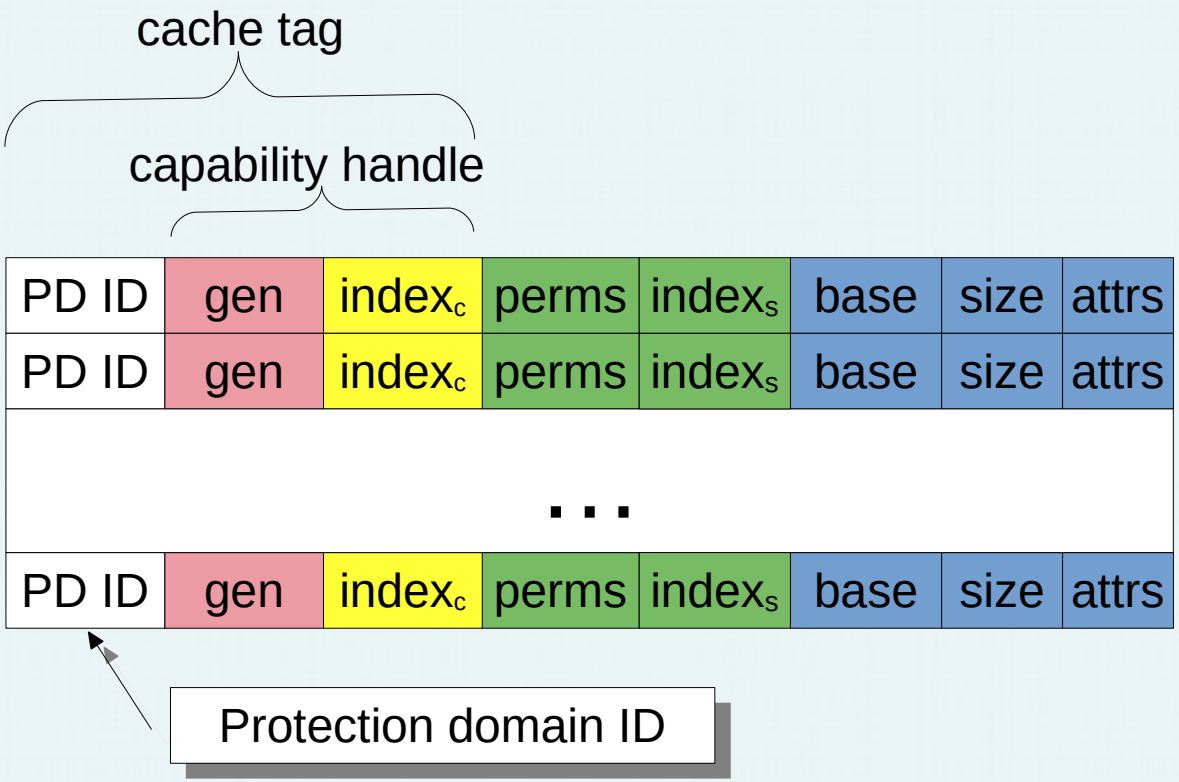


Capability Using Segmentation



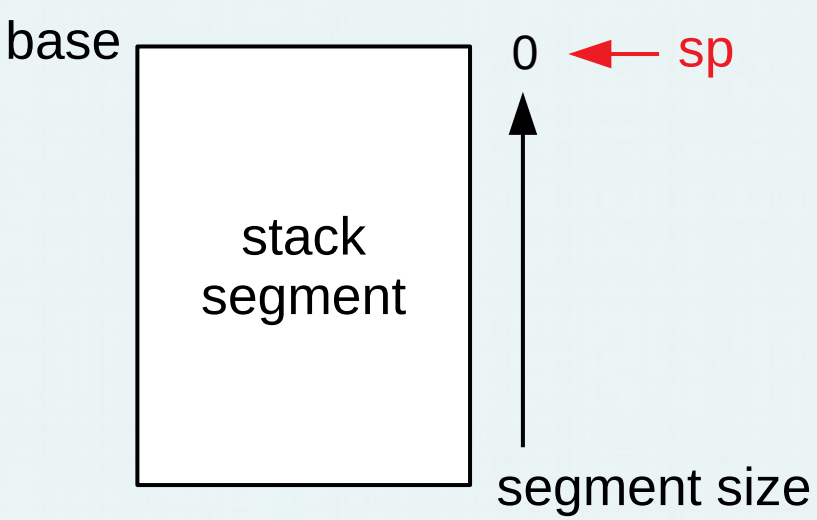
Segmentation Hardware

- CLB (Capability Lookaside Buffer)
  - Cache permissions and segment descriptors

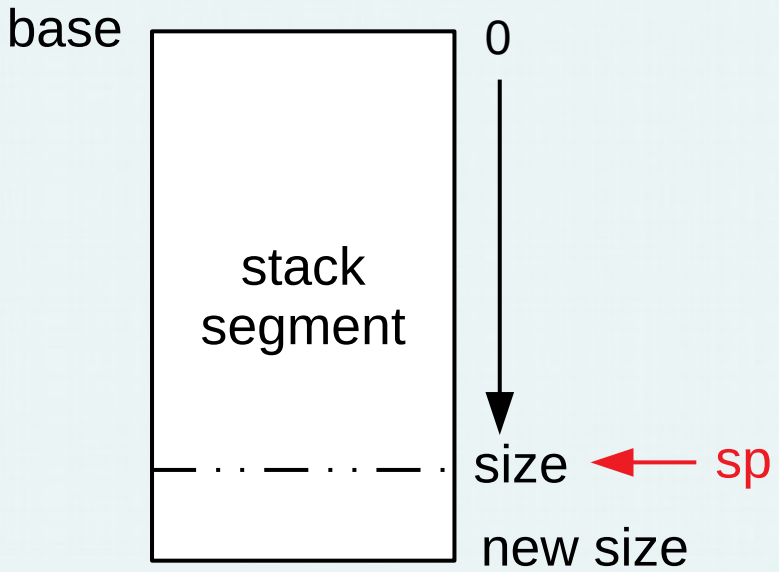


Stack Growth Direction

- Stack should grow towards ∞ instead of 0
  - Can expand the stack segment when it reaches its maximum size



towards 0



towards ∞

Segment Types

- Segmentation translates logical address to linear address
- Segment types
  - Virtual segment
    - Linear address is virtual address
  - Physical segment
    - Linear address is physical address
  - I/O segment
    - Linear address is I/O address

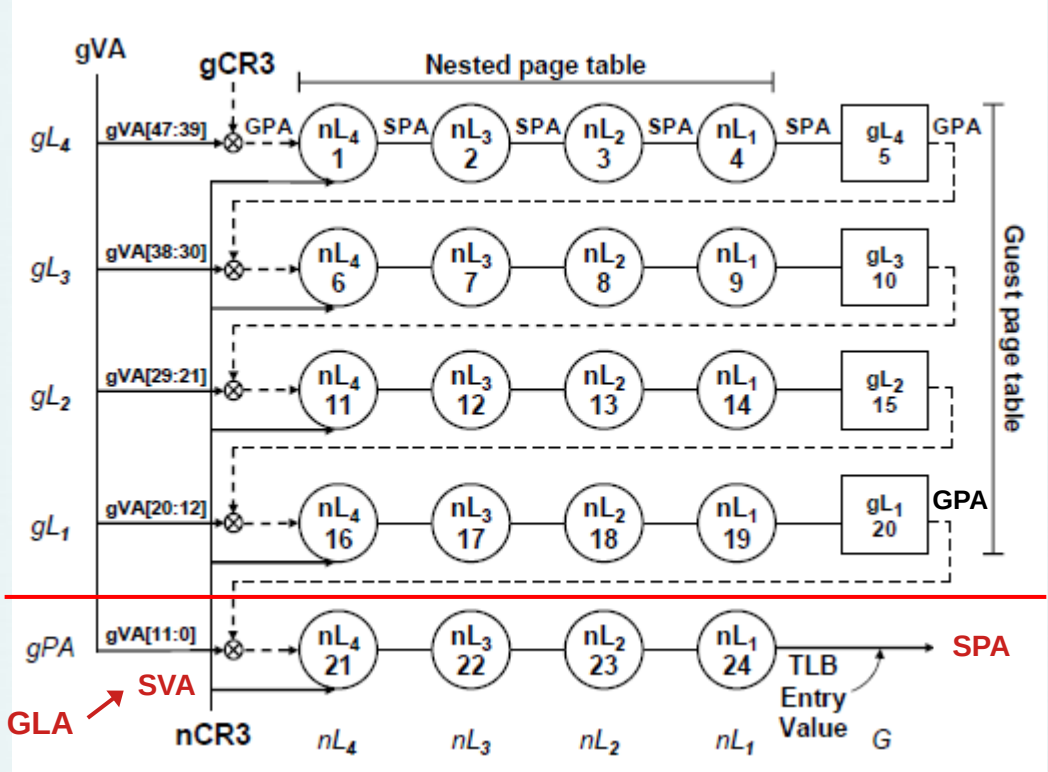
logical address is
{capability handle, offset within the segment}

Virtual Segment

- Simplification of segment move
  - Don't need to copy the segment data, only need to copy the virtual to physical mapping
  - Don't need to fix any pointers within the segment

Simplification of page table walk in Virtual Machine

- On a virtual machine, hardware needs to do two virtual address translations for TLB miss
  - Guest virtual address to guest physical address
  - System virtual address to system physical address
- With segmentation only one virtual address translation is needed
  - Guest logical address can be translated directly to system virtual address



Two-Dimensional Page Table Walk

Accelerating two-dimensional page walks for virtualized systems, acm ASPLOS, 2008

logical address is {capability handle, offset within the segment}

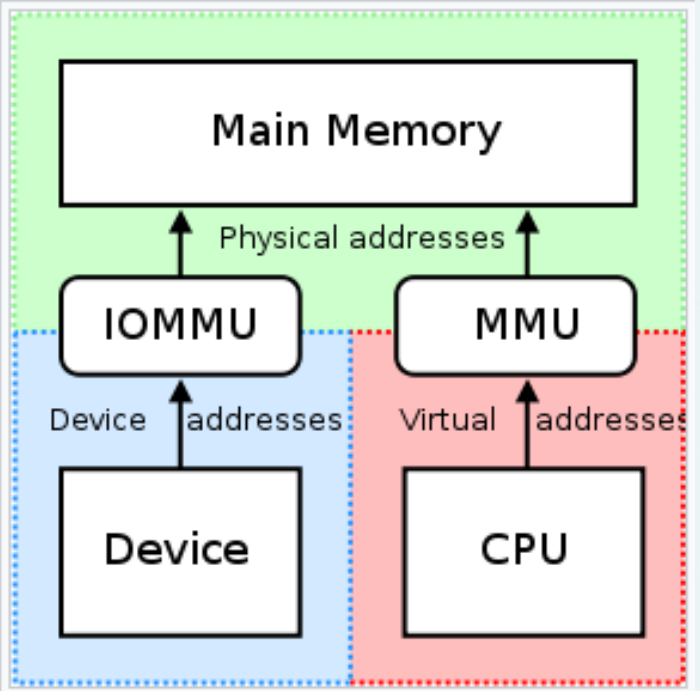
Physical Segment

- Benefits
  - Software-manged TLB
    - Must guarantee no TLB miss exceptions when the kernel is running the TLB miss handling routine
  - Performance improvement
    - Arkaprava Basu, Jayneel Gandhi, Jichuan Chang, Mark D. Hill, Michael M. Swift. Efficient virtual memory for big memory servers. In Proc. ISCA, 2013.
      - TLB misses consume up to 51% of execution cycles for big memory servers
      - Direct segment
  - IOMMU simplification
    - No page translation is needed
    - Simple authorization
    - IOMMU is just a CLB that caches physical segment descriptors

Simplification of IOMMU

Steps for device doing DMA via IOMMU

1. Device sends the capability, offset and data length to IOMMU
2. IOMMU verifies that the capability is valid and it points to a physical segment
3. IOMMU then checks that the offset and length is within segment size
4. IOMMU calculates the physical address by adding segment base address and the offset
5. The calculated physical address is used to access system main memory



Comparison of the I/O memory management unit (IOMMU) to the memory management unit (MMU).

From Wikipedia

I/O Segment

- Purpose
  - Let memory address space for memory only and move all the other stuff to I/O address space, such as CPU's CSRs, PCI configuration space, video frame buffer, boot ROM, etc.
- Advantages
  - No memory holes in memory address space
  - Segment is more fine-grained then page
  - Only one CLB entry is needed for a large I/O segment
  - Same load/store instruction can be used to load from or store to either I/O or memory address space
  - Device driver can access its hardware even when it runs at the user level without kernel intervention

Simplification of Shared Library

- No need for PIC (Position Independent Code), GOT (Global Offset Table) and PLT (Procedure Linkage Table)
  - Shared library has its own code and data segments and segment always starts with offset 0
- Shared libraries can be partially linked
  - The offset part of the library's global variables and functions can be statically linked at compile time
  - The segment part of the library's global variables and functions are linked at program load time

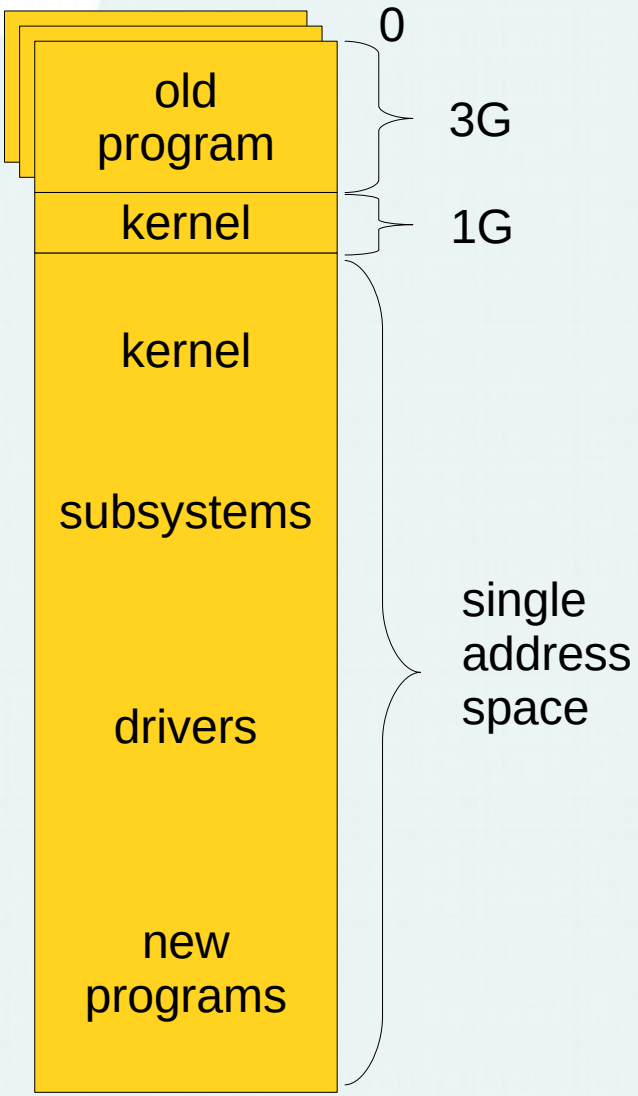
Cross-process Call

- Cross-process call has many names: cross-domain call, migrating thread model, protected control transfer, passive object model
- Currently thread is confined in two protection domains
- Thread should be able to travel around processes
  - Encourage modularity and improve security
  - Resource accounting can be made more accurate
  - Preferable for real-time systems

Simplification of Cross-process Call

- The same stack segment can be used in both client and server domain
- To protect server from client
  - Move the stack capability from client domain to server domain
- To protect client from server – stack domain boundary (sdb) register
  - A privileged register
  - Server can only access part of the stack with the address above the sdb register

Compatibility with Old Programs



- Ideal platform
  - 32-bit program on 64-bit virtual address space
- The first 3G is reserved for old programs
- The reset of the virtual address space is for kernel, device drivers, subsystems and single-address-space programs

Conclusion

- Lots of benefits we can get from single address space, capability and segmentation
- Changes
  - Hardware
    - Capability and segmentation architecture
  - Compiler support
    - Far pointer, far function call
    - New shared library implementation
  - OS
    - Segment, Capability-based single address space

The End

Questions?