

A decorative graphic of a green pipe network with various fittings, valves, and elbows, framing the central text.

Consolidating representations of the physical memory

Mike Rapoport
<rppt@linux.ibm.com>

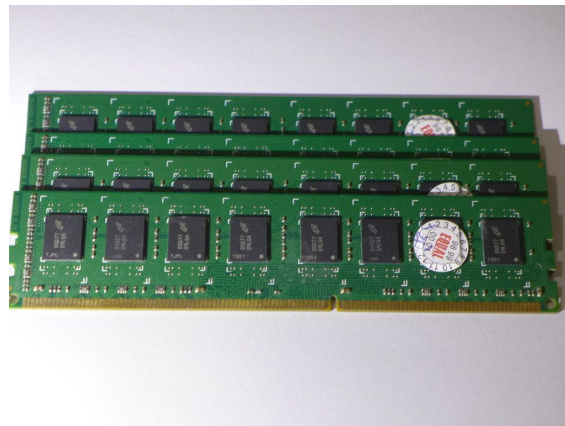


Linux
Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

Memory detection

Firmware

- Detect memory
- Reserve for own use
 - E.g. ACPI tables, EFI services
- Create memory map
 - HWRPBB, ATAGs, DT, e820, ...
- Create NUMA info
 - DT, SRAT, PCI config space, ...



Memory detection

Early boot

- Parse memory map
 - Reserve memory used by firmware
 - Create arch-specific representation
 - Pass memory ranges to memblock
- Setup NUMA topology
 - Verify NUMA tables supplied by firmware
 - Assign memory to nodes
 - Create arch-specific representation



Memory detection

Early boot

- Memory resources initialization
 - Initialize “System RAM” resource
 - Add memory resources for reserved regions
 - Add memory resources for kernel image
- Setup of memory sysfs

```
00000000-00000fff : Reserved
00001000-0009dfff : System RAM
0009e000-0009efff : Reserved
0009f000-0009ffff : System RAM
00100000-63e5dfff : System RAM
70596000-74295fff : Reserved
74296000-744fdfff : ACPI Tables
744fe000-7471efff : ACPI NVS
7471f000-75f1efff : Reserved
75f1f000-75ffefff : Unknown E820 type
75fff000-75ffffff : System RAM
76000000-79ffffff : Reserved
7a600000-7a7ffffff : Reserved
100000000-107f7ffffff : System RAM
    7b3400000-7b400199f : Kernel code
    7b4200000-7b49a9fff : Kernel rodata
    7b4a00000-7b4c8e43f : Kernel data
    7b526c000-7b57ffffff : Kernel bss
```



Run time memory range checks and operations

- Arch-specific representation
 - x86: `e820__mapped_any()`, `phys_to_target_node()`
 - ppc: `for_each_drmem_lmb()`
 - s390: `oldmem`
- Resource tree
 - `page_is_ram()`, `walk_system_ram_range()`
- Memblock
 - `memblock_is_map_memory()`



A decorative graphic of a green pipe network with various fittings, valves, and elbows, running along the top, left, and bottom edges of the slide. The pipes are a vibrant green color with a slight shadow effect.

kexec

- Relies on `/proc/iomem`
 - Find memory to load kernel image
 - Create memory map for the second kernel
- Only x86 can use `/sys/firmware/memory`
 - Except on Xen and some EFI systems



Goal

Create a generic representation of the physical memory

- Consistent with firmware memory map
- Backward compatible with existing uABI
- Flexibility to accommodate architecture-specific needs



Alternatives

- New data structure
- Resource tree
- memblock



Proposal

Use memblock as physical memory representation

- Enable ARCH_KEEP_MEMBLOCK for NUMA architectures
- Generate “System RAM” from memblock
 - <https://lore.kernel.org/all/20210531122959.23499-1-rppt@kernel.org/>
- Close the gap in e820 vs memblock integration
- Rename `memblock_alloc()` back to `bootmem_alloc()`?



A stylized green pipe network graphic is positioned around the edges of the slide. It features various pipe fittings, elbows, and valves, creating a complex network that frames the central text.

Thank you!



Linux
Plumbers Conference | Dublin, Ireland **Sept. 12-14, 2022**