CERTIFIED INTERRUPTIBLE OS KERNELS AND DEVICE DRIVERS

Zhong Shao
Joint work with Hao Chen, Newman Wu, Joshua Lockerman, and Ronghui Gu
Yale University

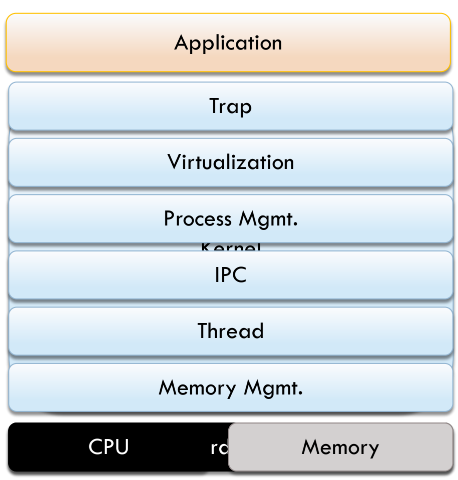
CertiKOS

CPSC 422 Operating Systems October 17, 2022

1

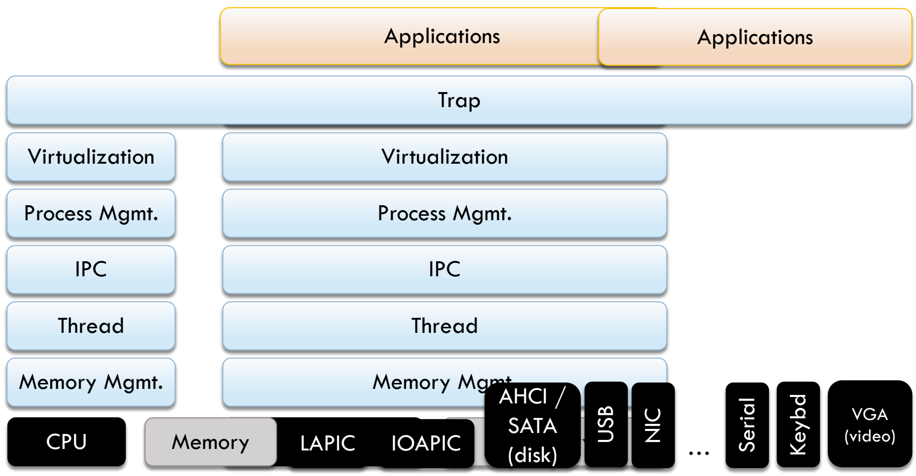
Formal Verification of OS Kernel

- seL4
- CertiKOS
- Verve



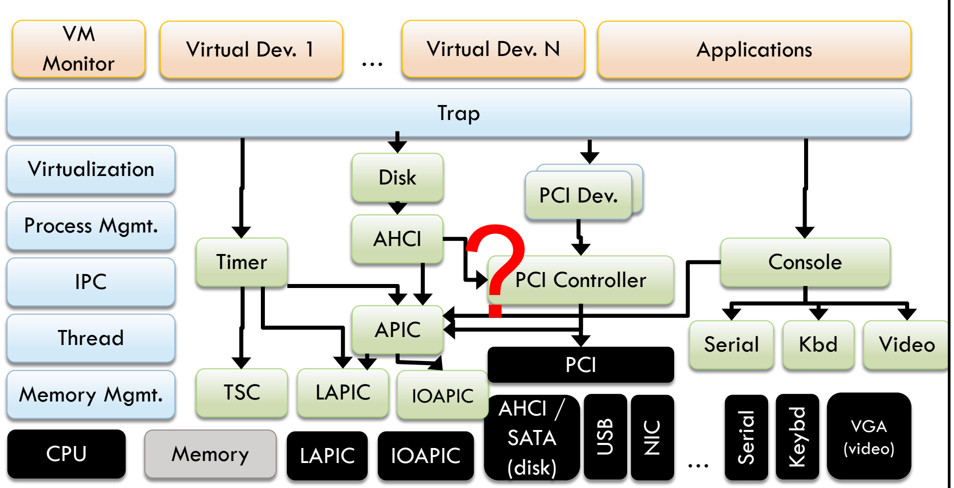
2

Formal Verification of OS Kernel



3

Formal Verification of OS Kernel



4

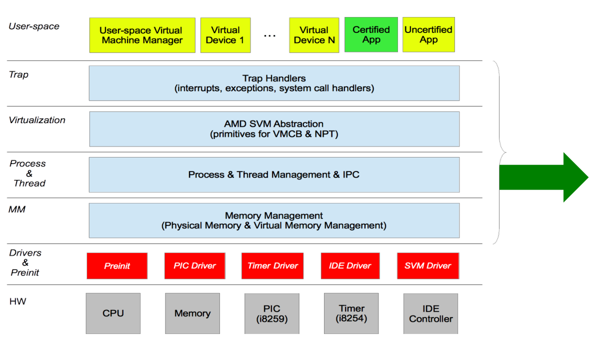
Device Drivers in Mainstream OS

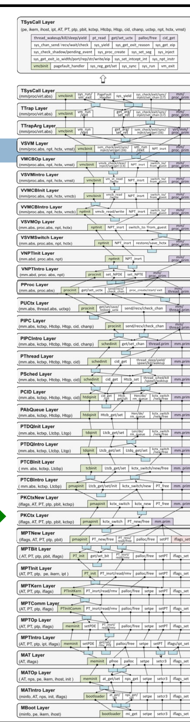
- 70% of Linux 2.4.1 kernel are device drivers.
- 70% of Windows crash are caused by third-party driver code.

5

mCertiKOS Overview [POPL'15]

- Single-core version of CertiKOS.
- 3k LOC, can boot Linux as guest.
- Aggressive use of abstraction over deep specification (37 layers).

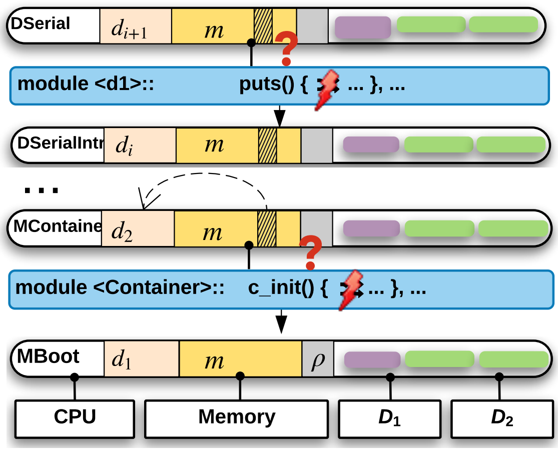




6

Main Challenge

Every fine-grained processor step could be interrupted.



7

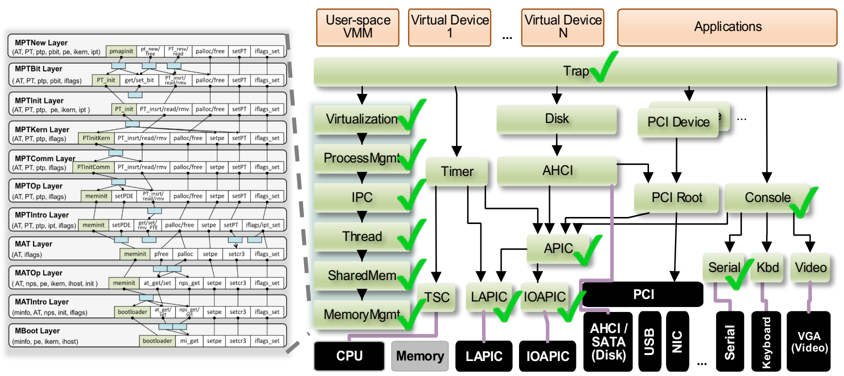
Other Challenges

- Interrupt hardware can be *dynamically* configured.
- Devices and CPU run in *parallel*.
- Device drivers are written in both C and *assembly*.
- The correctness results of different components should be *linked formally*.

8

Our Contributions [PLDI'16]

The *first* formally verified *interruptible* OS kernel with *device drivers.*



9

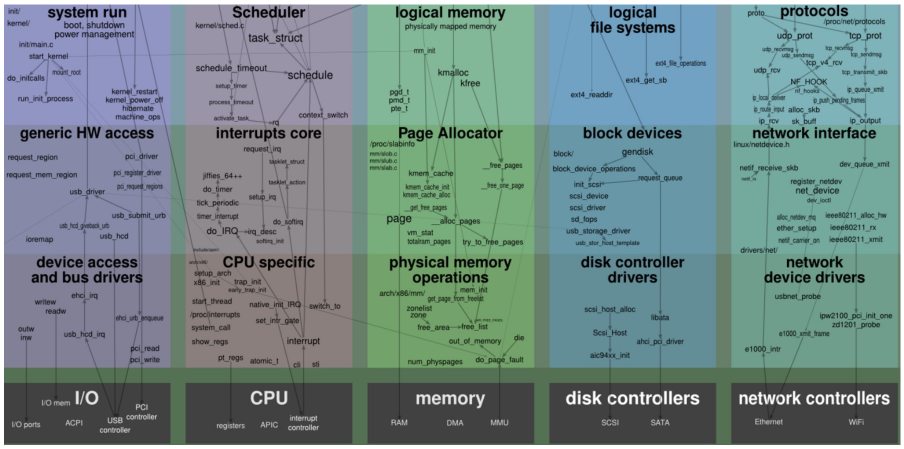
Our Contributions [PLDI'16]

- New techniques for certifying abstraction layers with multiple ***logical CPUs*** and devices.
- New techniques for building formal ***certified device hierarchies.***
- An abstraction-layer-based approach for reasoning about ***interrupts.***
- ***Case study***: interruptible mCertiKOS with device drivers.

10

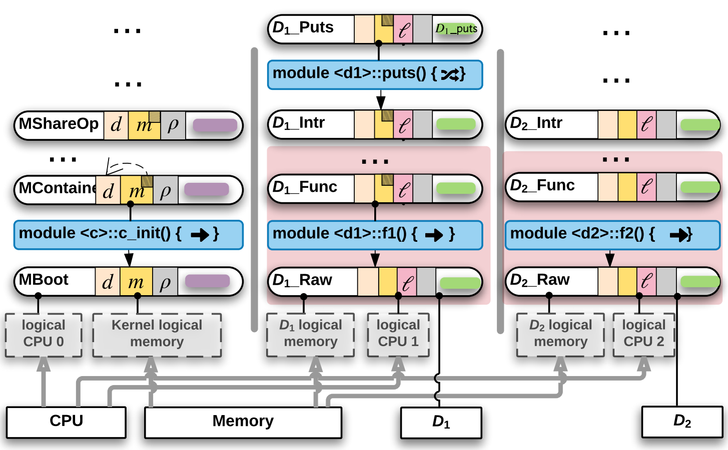
Linux Kernel Map

Kernel components are sorted into different stacks of abstraction layers based on their underlying hardware device.



11

New Machine Model



12

Our Contributions

- New techniques for certifying abstraction layers with multiple ***logical CPUs*** and devices.
- New techniques for building formal ***certified device hierarchies***.
- An abstraction-layer-based approach for reasoning about ***interrupts***.
- ***Case study***: interruptible mCertiKOS with device drivers.

13

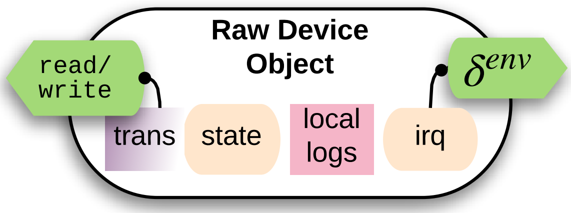
Hardware Device Model

- Devices are modeled as transition systems parameterized by all possible lists of external events.
- Example external events:
  - Recv (s: list char)
  - KeyPressed (c: Z)
- State: observable registers.
- Transition:
  - environmental transition: δ^{env}
  - I/O transition: δ^{CPU}

14

Raw Device Object

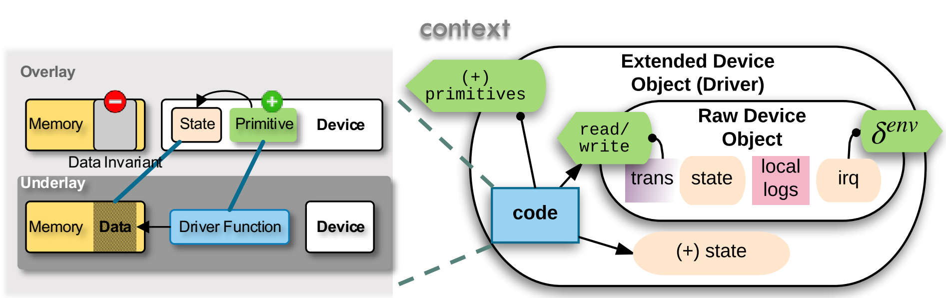
- Local log for the list of observed external events.
- Multiple local logs to handle disjoint set of external events asynchronously.
- Read/Write instructions: IN/OUT, memory mapped I/O, *etc*.



15

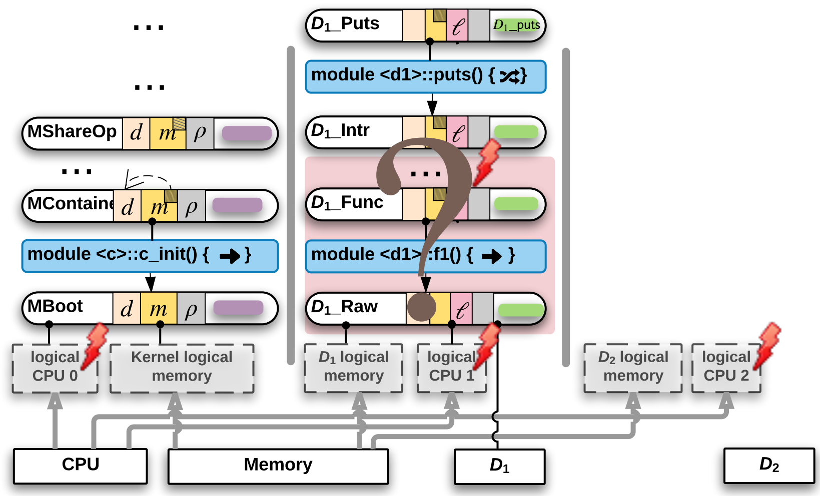
Extended Device Object

Driver as a logical device.



16

Recap: Machine Model



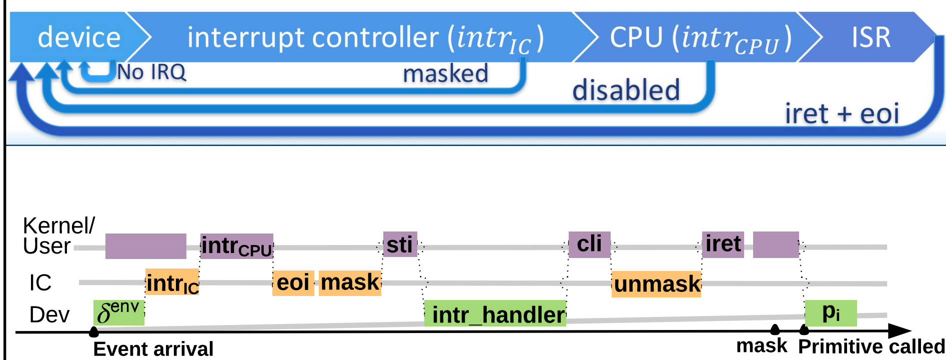
17

Our Contributions

- New techniques for certifying abstraction layers with multiple ***logical CPUs*** and devices.
- New techniques for building formal ***certified device hierarchies***.
- An abstraction-layer-based approach for reasoning about ***interrupts***.
- ***Case study***: interruptible mCertiKOS with device drivers.

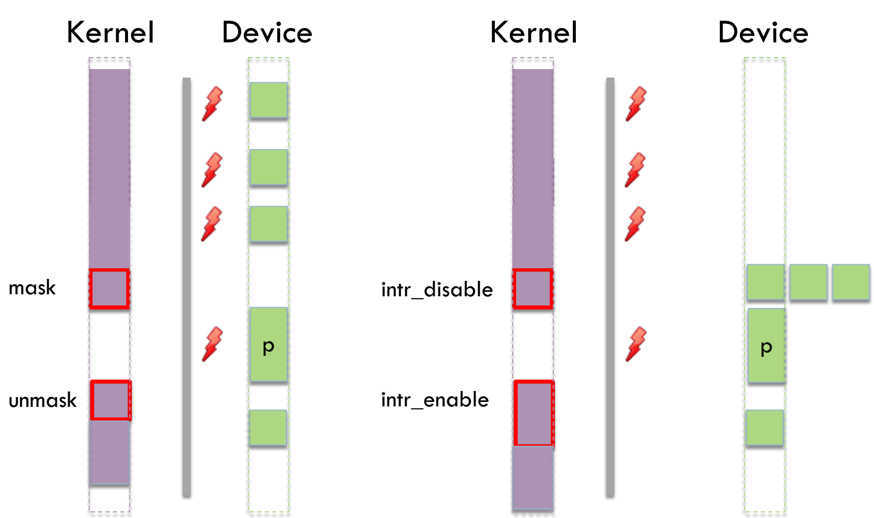
18

Interrupt Models



19

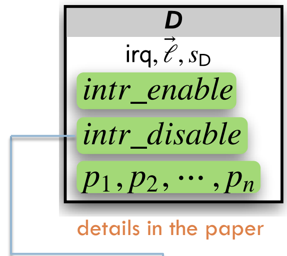
New HW Interrupt Model



20

Semantics of *intr_disable*

- Scans external events.
- Recursively performs the environmental transition.
- Synchronizes unhandled interrupts.



DisableNoIntr: Disable with no unhandled interrupt

$$\frac{(e, \ell_i') = \mathsf{next}(\ell^{env}, \ell_i) \qquad s_{\mathsf{tmp}} = \delta^{\mathsf{env}}(s, e) \qquad s_{\mathsf{tmp}}.irq = \mathsf{false} \qquad s' = s[\mathsf{iFlag} \leftarrow 0]}{\texttt{intr\_disable}(s, \ell_i, \ell^{env}) = (s', \ell_i)}$$

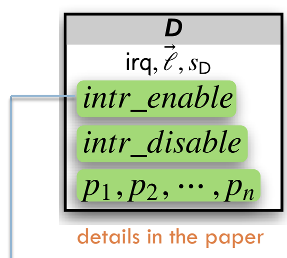
DisableIntr: Disable with unhandled interrupts

$$\frac{(e, \ell_i') = \mathsf{next}(\ell^{env}, \ell_i) \qquad s' = \delta^{\mathsf{env}}(s, e) \qquad s'.irq = \mathsf{true} \qquad (s'', \ell_i'') = \texttt{intr\_handler}(s', \ell_i', \ell^{env}) \qquad (s''', \ell_i''') = \texttt{intr\_disable}(s'', \ell_i'', \ell^{env})}{\texttt{intr\_disable}(s, \ell_i, \ell^{env}) = (s''', \ell_i''')}$$

21

Semantics of *intr_enable*

- Recursively discharges pending interrupts.
- Delayed interrupts that occur while the interrupt is disabled.



EnableNoIntr: Enable with no pending interrupt

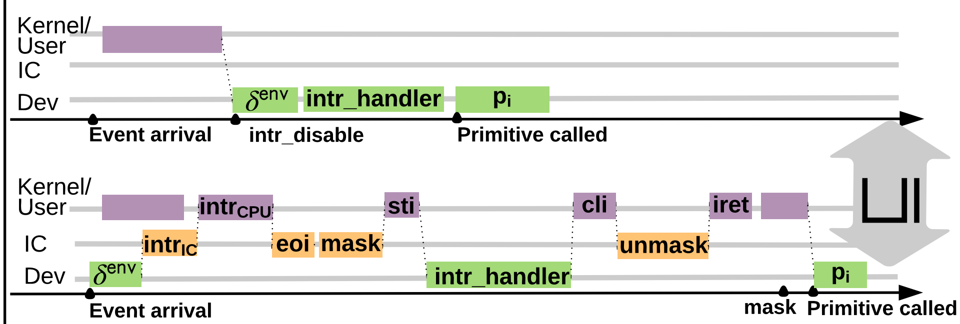
$$\frac{s.irq = \mathsf{false} \qquad s' = s[\mathsf{iFlag} \leftarrow 1]}{\texttt{intr\_enable}(s, \ell_i, \ell^{env}) = (s', \ell_i)}$$

EnableIntr: Enable with pending interrupts

$$\frac{s.irq = \mathsf{true} \qquad (s', \ell_i') = \texttt{intr\_handler}(s, \ell_i, \ell^{env}) \qquad (s'', \ell_i'') = \texttt{intr\_enable}(s', \ell_i', \ell^{env})}{\texttt{intr\_enable}(s, \ell_i, \ell^{env}) = (s'', \ell_i'')}$$

22

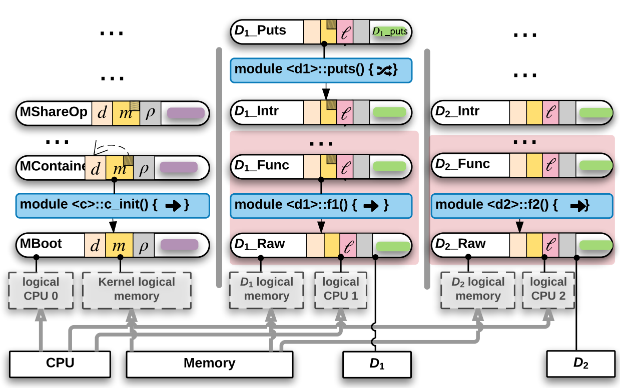
Refinement btw. The HW & Abstract Interrupt Model



23

Our Approach

- The driver code of each device runs on its own "logical CPU", operates its own internal states.
- Interruptible code can be naturally reasoned on top of the abstract interrupt model.



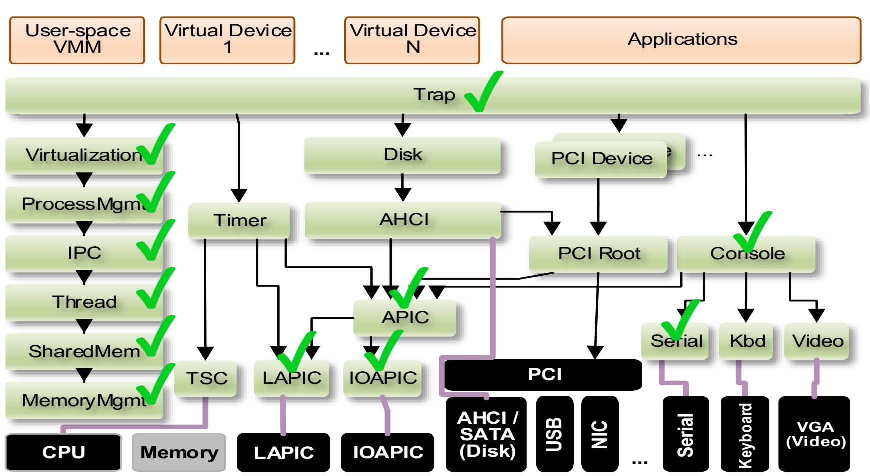
24

Our Contributions

- New techniques for certifying abstraction layers with multiple ***logical CPUs*** and devices.
- New techniques for building formal ***certified device hierarchies***.
- An abstraction-layer-based approach for reasoning about ***interrupts***.
- ***Case study***: interruptible mCertiKOS with device drivers.

25

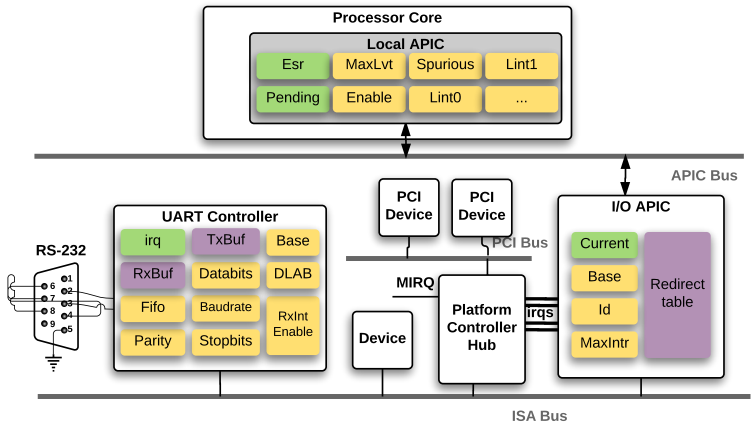
Interruptible mCertiKOS with Drivers



26

Case Study: Modeling HW Devices

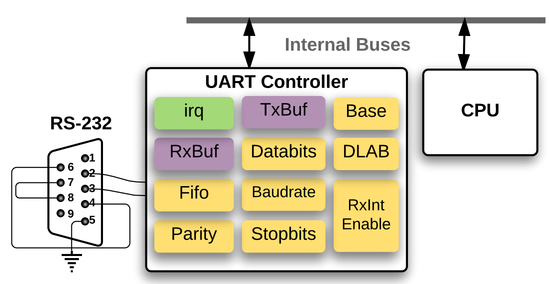
- Serial Port, I/O APIC, Local APIC, CPU interrupt handling.



27

Case Study: Serial Device

- States: see figure
- Transitions: serial_trans_env + serial_trans_IO
- Read/Write primitives: serial_read / serial_write



28

Serial Interrupt Handler

```
void serial_intr () {
  unsigned int hasMore;
  int t = 0;
  hasMore = serial_getc ();
  while (hasMore && t < CONSOLE_BUFFER_SIZE) {
    hasMore = serial_getc ();
    t++;
  }
}
```

```
unsigned int serial_getc () {
  unsigned int rv = 0;
  unsigned int rx;
  if (serial_exists()) {
    if (serial_read(COM1 + COM_LSR, BIT1) % 2 == 1)
    {
      rx = serial_read(COM1 + COM_RX, M_ALL);
      cons_buf_write(rx);
      rv = 1;
    }
  }
  return rv;
}
```

29

Serial Driver

```
void serial_puts(char * s, int len) {
  int i = 0;
  while (i < len && s[i] != 0) {
    serial_intr_disable ();
    serial_putc (s[i]);
    serial_intr_enable ();
    i++;
  }
}
```

```
void serial_putc (unsigned int c) {
  unsigned int lsr = 0, i;
  if ( serial_exists() ){
    for (i = 0; !lsr && i < 12800; i++) {
      lsr = serial_read(0x3FD) & 0x20;
      delay();
    }
    serial_write (0x3F8, c);
    ...
```

30

What We Have Proved

- Total functional correctness.
- Safety.
- *Contextual refinement* between the lowest and the top level abstract machine:
  $\forall P,\ [\![K \bowtie P]\!]_{\text{x86}} \sqsubseteq [\![P]\!]_{\text{mCertiKOS}}$
- Data invariants:
  - Console's circular buffer is always well-formed.
  - Interrupt controller states are always consistent.
- The framework also ensures that:
  - No code injection attacks, buffer overflow, integer overflow, null pointer access, *etc*.

31

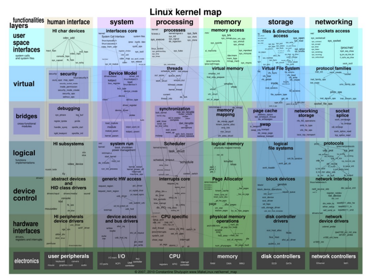
Size of TCB and Spec/Proof

- In the TCB
  - X86 hardware model
  - Hardware device/interrupt model (510 LOC)
  - System call specification (126 LOC)
  - Bootloader
  - Coq proof checker
  - Pretty-printing phase of the CompCert compiler
- Rest of the spec/proof (about 20k LOC)
  - Intermediate and auxiliary specifications and definitions
  - Coq proof scripts

32

Conclusion

- Compositional framework for building certified interruptible kernel with device drivers.
  - Certified abstraction layers with multiple logical CPUs.
  - An abstraction-layer-based approach for expressing interrupts.
- The first formally verified interruptible OS kernel with device drivers.
- Extensions:
  - Other drivers
  - Concurrency
  - Larger kernel



33