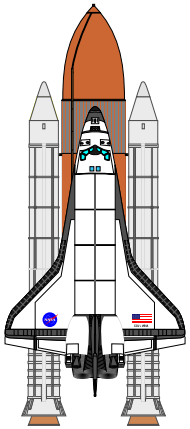
SuperGlue: IDL-Based, System-Level Fault Tolerance for Embedded Systems

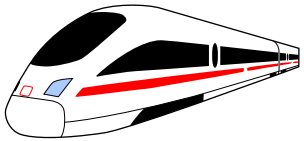
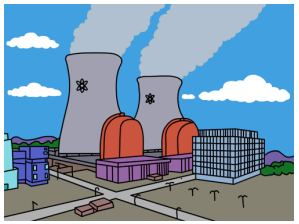
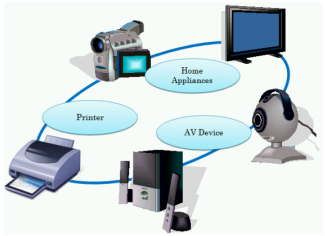
June 18, 2016

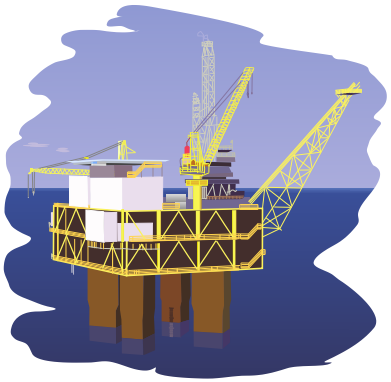
Computer Science Department
The George Washington University

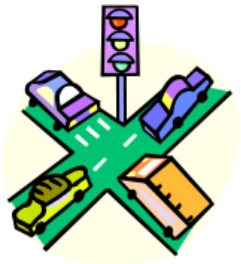
Real-Time and Embedded Systems



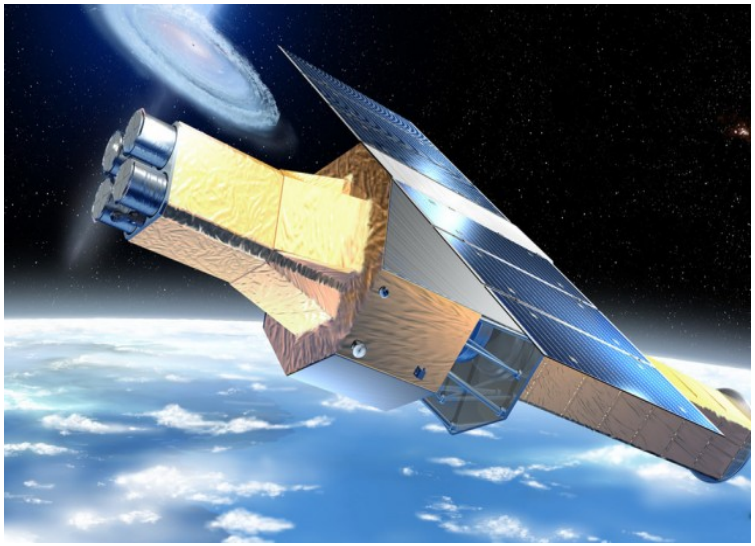





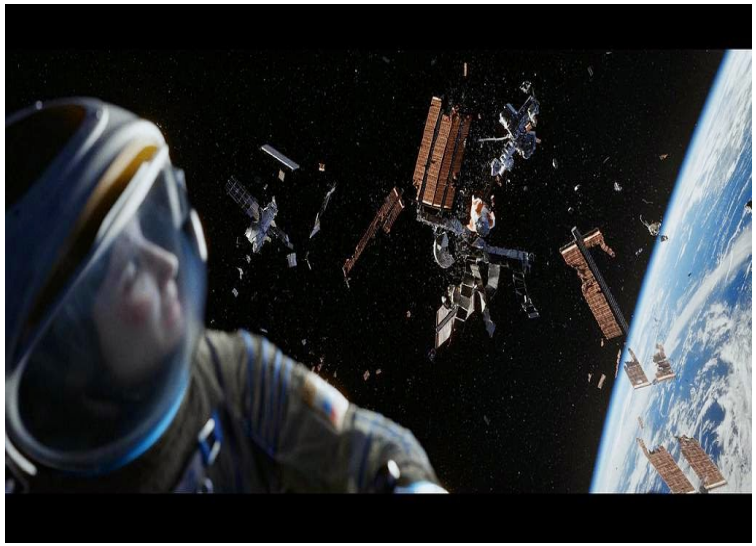

http://www.pcmag.com/encyclopedia/term/42554/embedded-system

Consequences of Embedded System Faults

Japanese X-ray astronomy satellite Hitomi lost in 2016



Fault

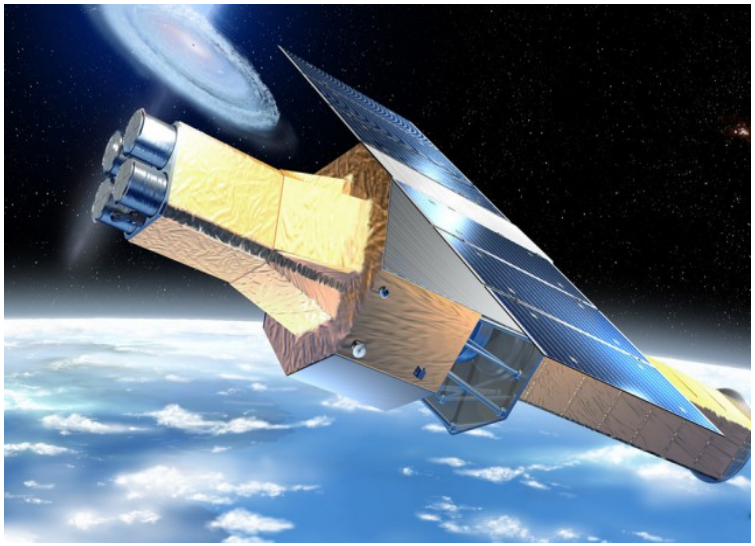


- Financial losses of $286 million USD
- "*It's a scientific tragedy*" - Richard Mushotzky, UMD

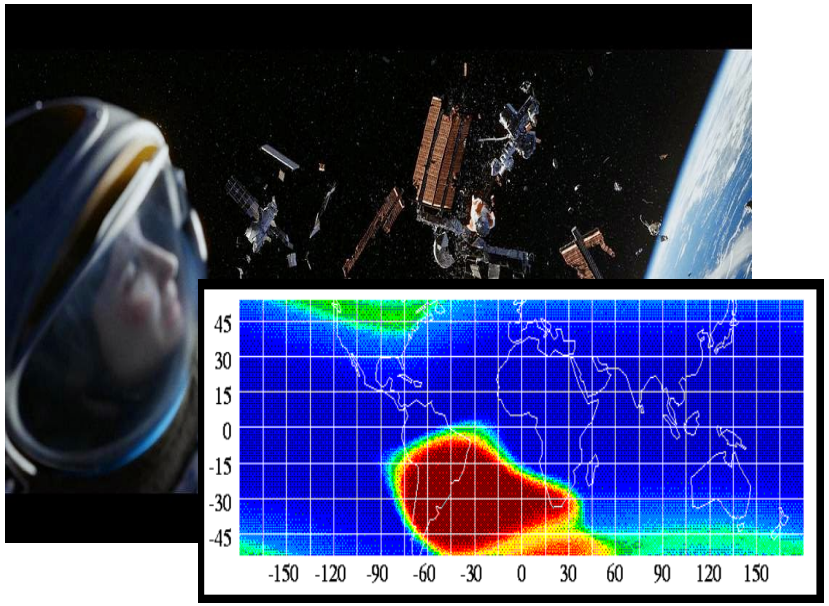
`http://www.nature.com/news/software-error-doomed-japanese-hitomi-spacecraft-1.19835`

Consequences of Embedded System Faults

Japanese X-ray astronomy satellite Hitomi lost in 2016



Fault



The South Atlantic Anomaly

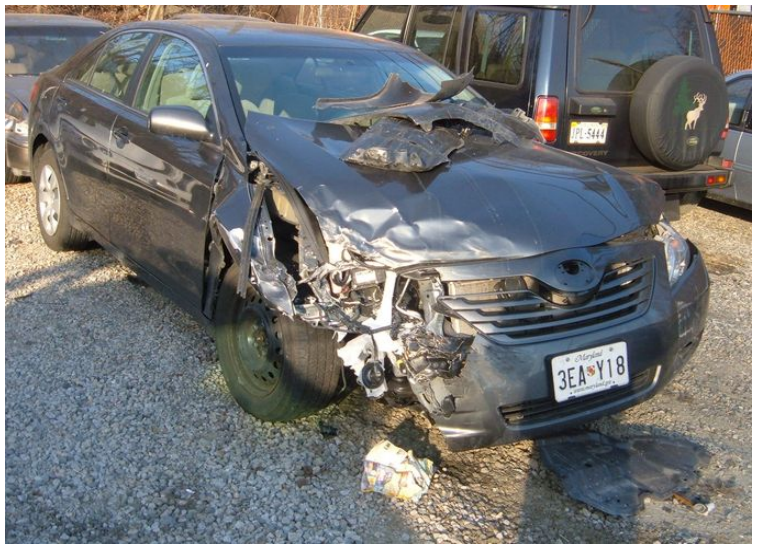
- Financial losses of $286 million USD
- "*It's a scientific tragedy*" - Richard Mushotzky, UMD

`http://www.nature.com/news/software-error-doomed-japanese-hitomi-spacecraft-1.19835`

Consequences of embedded system faults

Toyota Sudden Unintended Acceleration (SUA) in 2004 – 2010

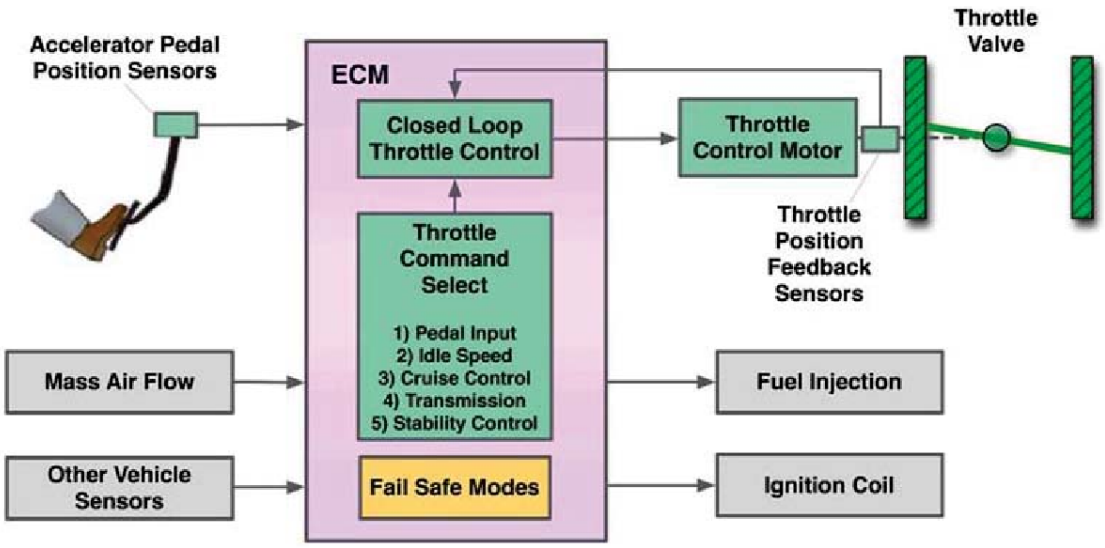
Fault

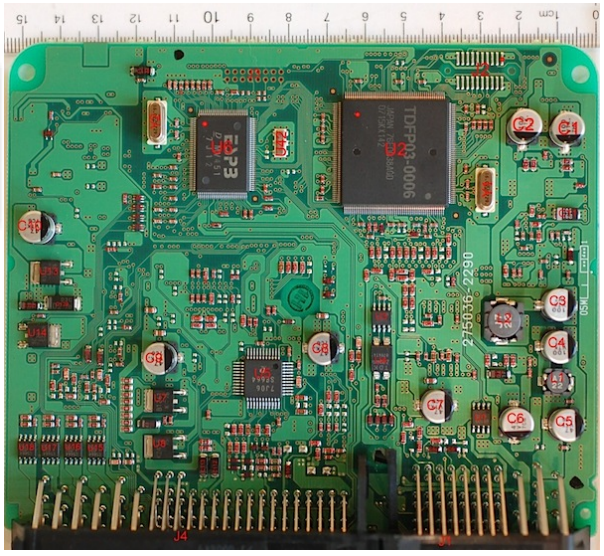


- 89 deaths as of May 2010 and nearly 400 U.S. lawsuits
- Recall 10+ million vehicles and pay $1.2 Billion USD fine

`http://www.cbsnews.com/news/toyota-unintended-acceleration-has-killed-89/`

Sudden Unintended Acceleration



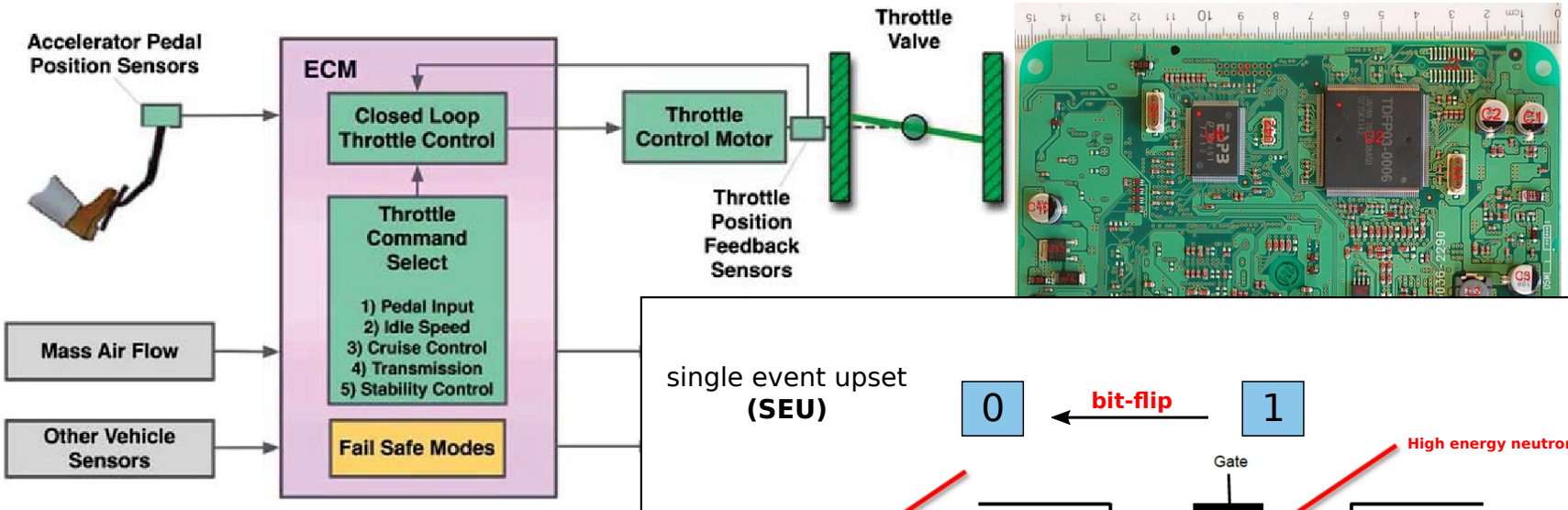


Uncontrolled acceleration in Toyota Camrys

- Electronic Throttle Control System (ETCS)
- OSEK OS, 24 tasks, 280K LOC of C
- bit flip in scheduler data-structures
  $\rightarrow$ reproducible 30-sec uncontrolled acceleration

"*a bit-flip there, will have the effect of killing one of the tasks*"
– Bookout v Toyota

Sudden Unintended Acceleration



Radiation strike causing transistor disruption [Gorini 2012]

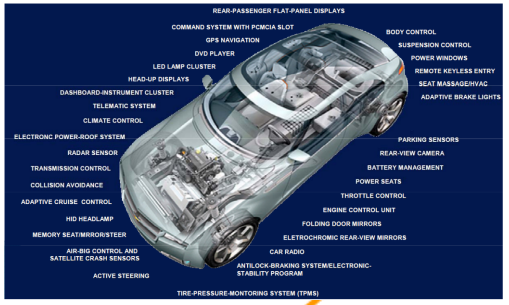
Uncontrolled acceleration in

- Electronic Throttle Con
- OSEK OS, 24 tasks, 28
- bit flip in scheduler data-structures
  $\rightarrow$ reproducible 30-sec uncontrolled acceleration

"*a bit-flip there, will have the effect of killing one of the tasks*"
– Bookout v Toyota

Embedded faults: Bad Now, Worse Tomorrow

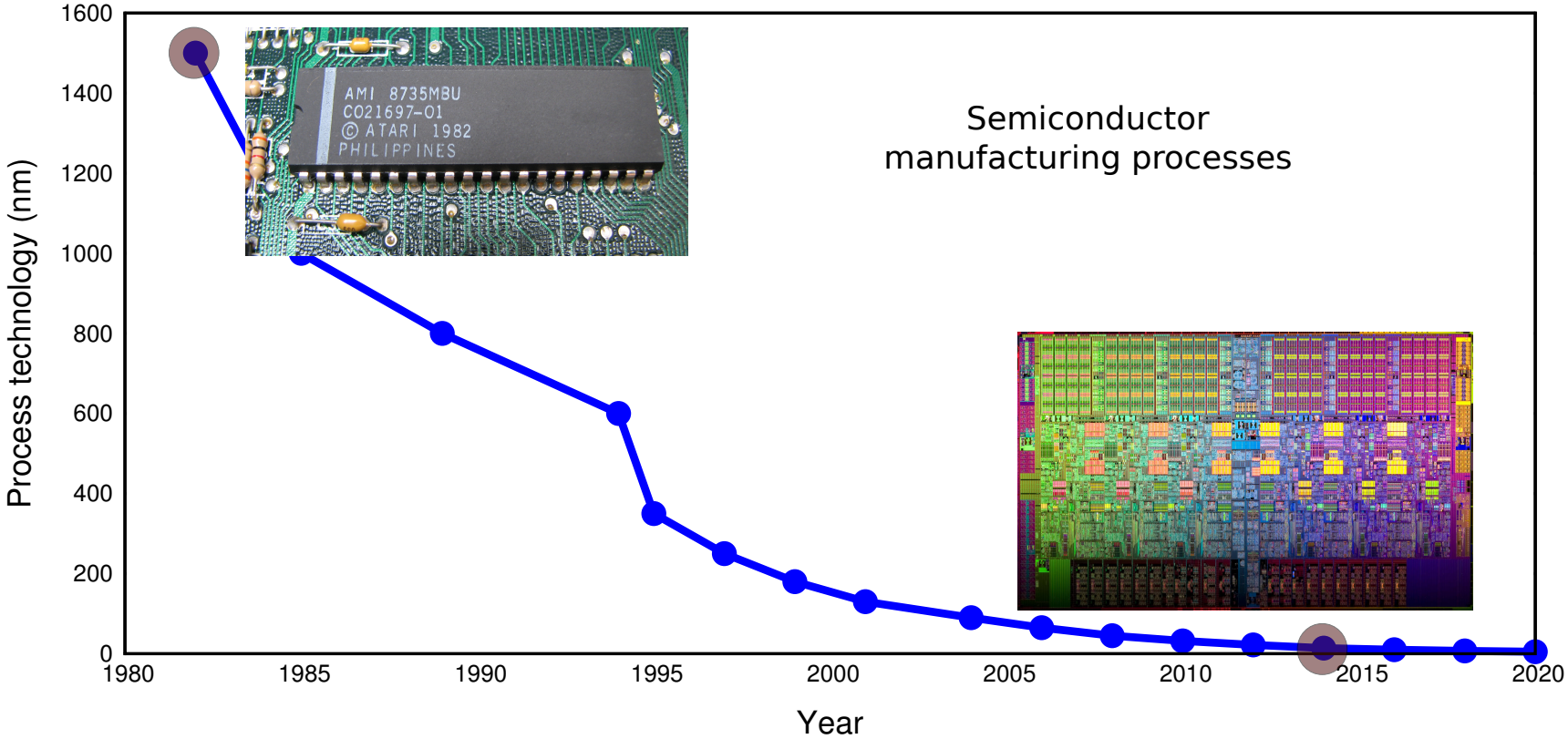
ABS: Antilock Brake System
ACC: Adaptive Cruise Control
BCM: Body Control Module
Cyl Deact: Cylinder Deactivation
ECS: Electronics, Controls, and Software
EGR: Exhaust Gas Recirculation.
GDI: Gas Direct Injection
OBD: Onboard Diagnostics
TCC: Torque Converter Clutch
PT: Powertrain
ESC: Electronic Stability Control



\+ more functionality
\− more complexity → dependability more challenging

https://www.src.org/calendar/e003506/03-sessionii-rajkumar.pdf

Embedded faults: Bad Now, Worse Tomorrow



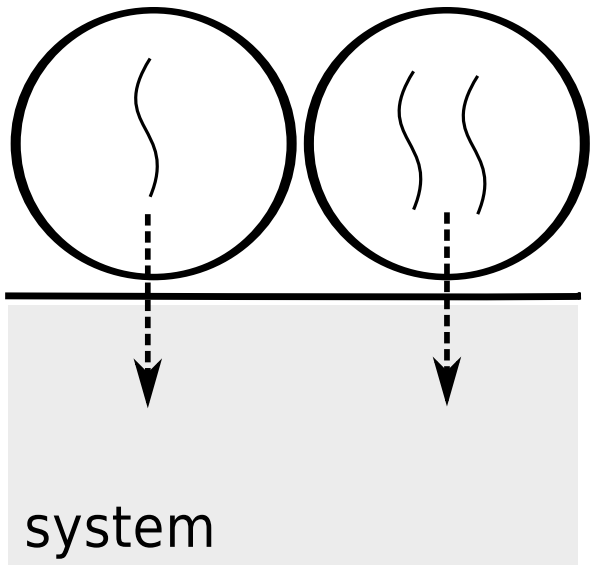
Decreasing process sizes $\rightarrow$ 5nm

- \+ faster
- \+ less power
- \+ smaller
- – increased vulnerability to HW transient faults

Application-Level Fault Tolerance

Application fault tolerance

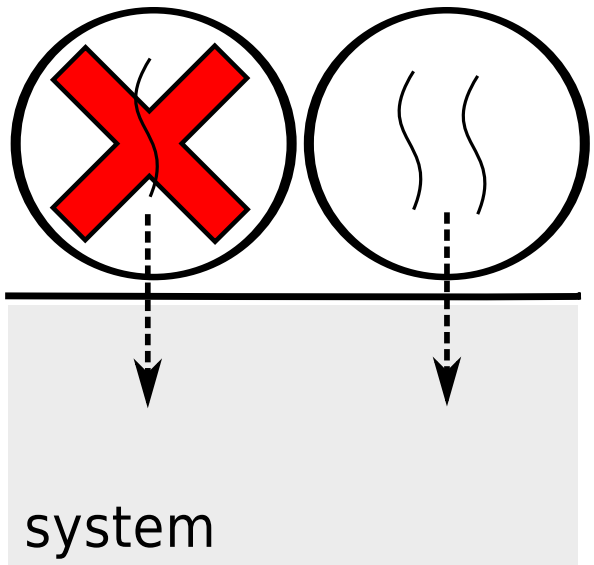
- example recovery techniques
  - recovery blocks
  - checkpointing
  - re-fork
- temporal redundancy
  - detect fault by job completion
  - replay execution from saved state
- re-execution impacts only lower-priority tasks



Application-Level Fault Tolerance

Application fault tolerance

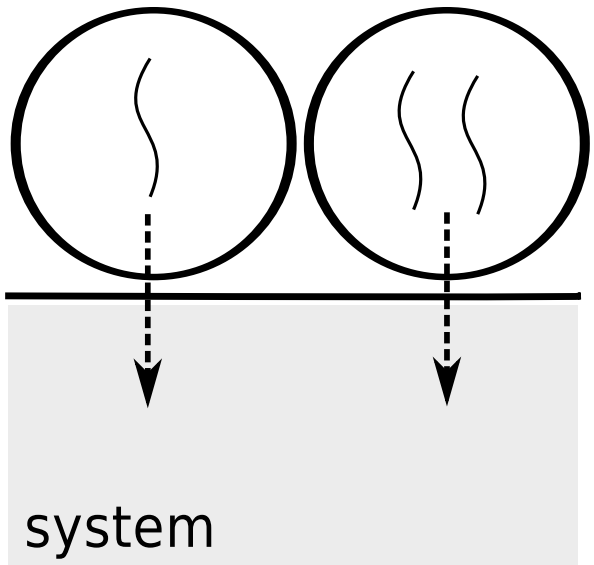
- example recovery techniques
  - recovery blocks
  - checkpointing
  - re-fork
- temporal redundancy
  - detect fault by job completion
  - replay execution from saved state
- re-execution impacts only lower-priority tasks



Application-Level Fault Tolerance

Application fault tolerance

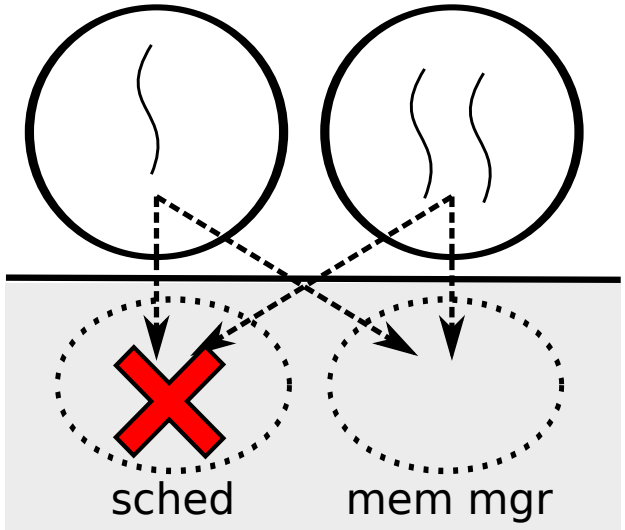
- example recovery techniques
  - recovery blocks
  - checkpointing
  - re-fork
- temporal redundancy
  - detect fault by job completion
  - replay execution from saved state
- re-execution impacts only lower-priority tasks



System-Level Fault Tolerance

System-level fault tolerance

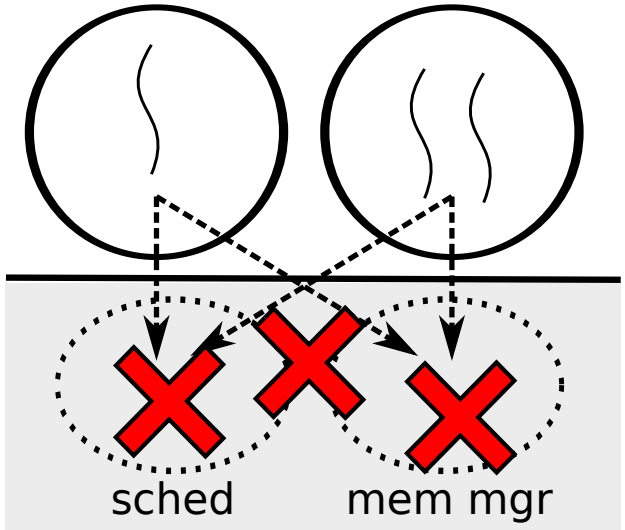
- failures in
  - scheduler
  - memory mapping manager
  - file-systems
  - synchronization manager
  - ...



System-Level Fault Tolerance

System-level fault tolerance

- failures in
  - scheduler
  - memory mapping manager
  - file-systems
  - synchronization manager
  - ...



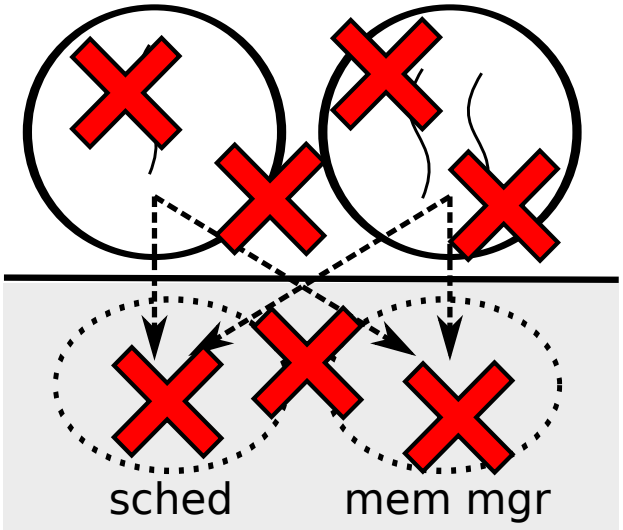
System components contain state for all tasks

- failure impacts memory of *all* tasks

System-Level Fault Tolerance

System-level fault tolerance

- failures in
  - scheduler
  - memory mapping manager
  - file-systems
  - synchronization manager
  - ...



System components contain state for all tasks

- failure impacts memory of *all* tasks

Recovery requires resources

- processing time...to recovery scheduler
- memory...to recover memory mapper

System-Level Fault Tolerance

System-level fault tolerance

- failures in
  - scheduler
  - memory mapping manager
  - file-systems
  - synchronization manager
  - ...

System components contain state for all tasks

- failure impacts memory of *all* tasks

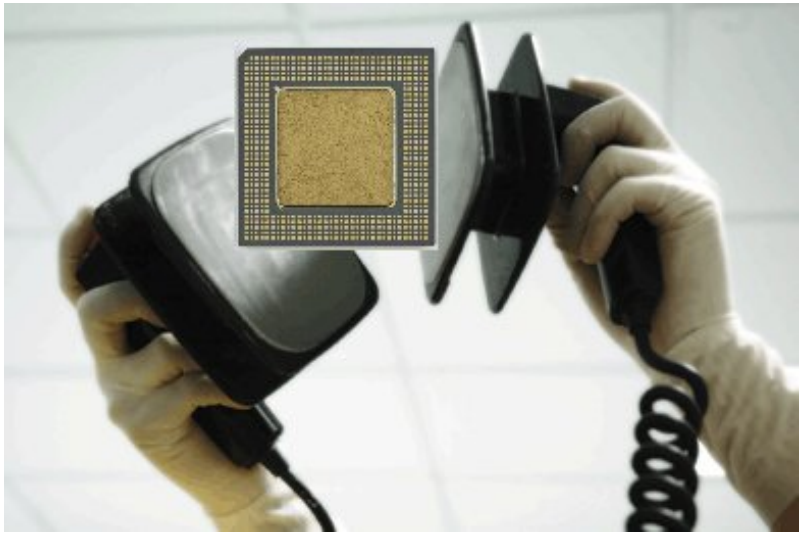
Recovery requires resources

- processing time...to recovery scheduler
- memory...to recover memory mapper

C^3: The Computational Crash Cart

C^3: Computational Crash Cart

- resuscitate system
- from system-level faults
- predictably



Main ideas

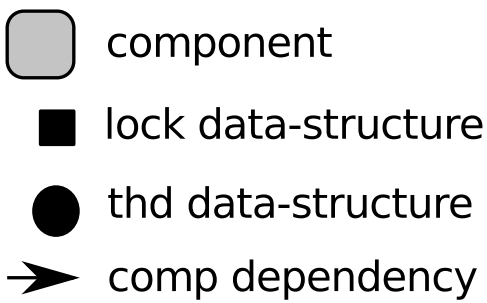
- pervasive fault isolation $\rightarrow$ restrict propagation
- efficient μ-reboot of individual components
- interface-driven, application-oblivious recovery
- on-demand recovery $\rightarrow$ bound recovery inversion

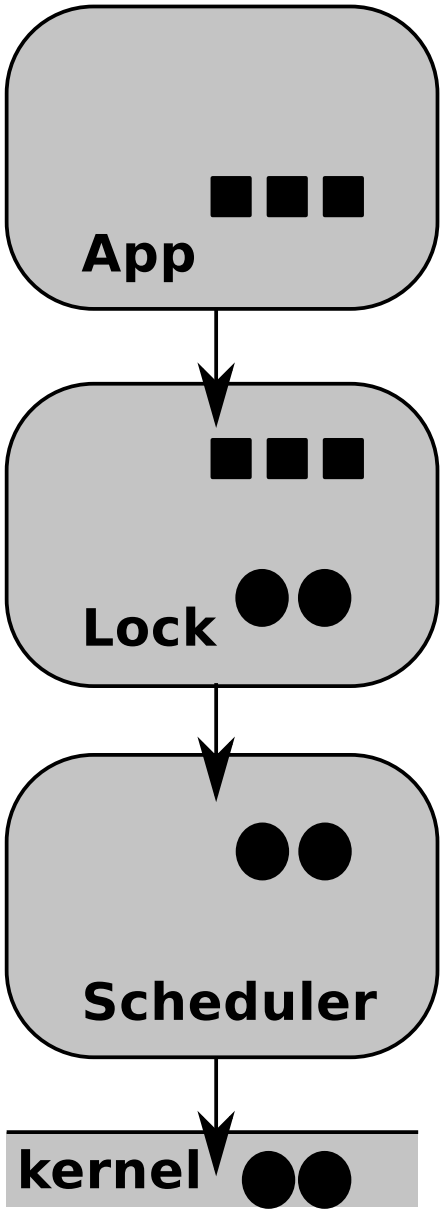
C^3 – an interface-driven, predictable system-level fault recovery mechanism

J. Song, J. Wittrock, and G. Parmer, "Predictable, efficient system-level fault tolerance in C^3" in RTSS, 2013

C^3 System-Level Fault Recovery

Recovery sequence in C^3

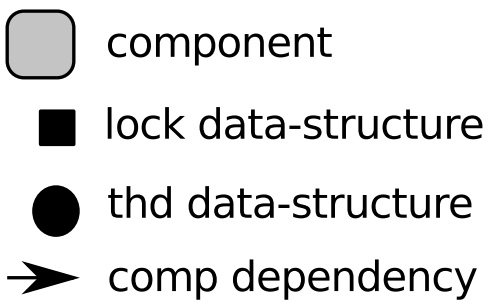


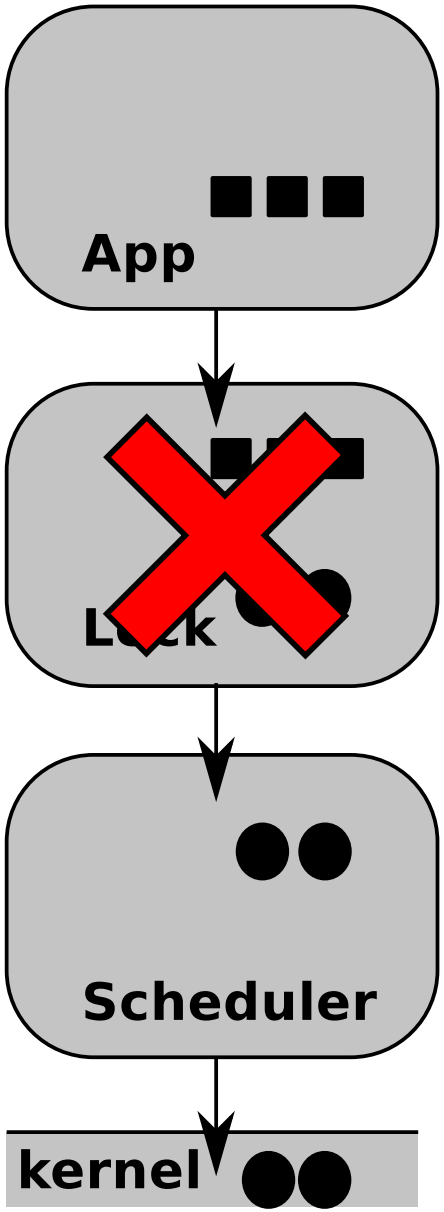


C^3 System-Level Fault Recovery

Recovery sequence in C^3

1. fault detection

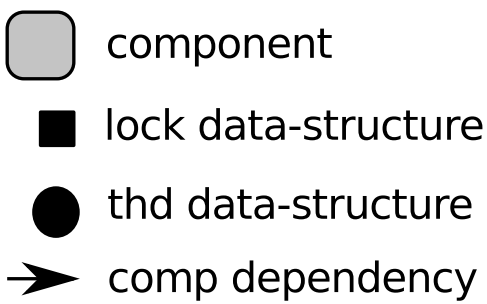


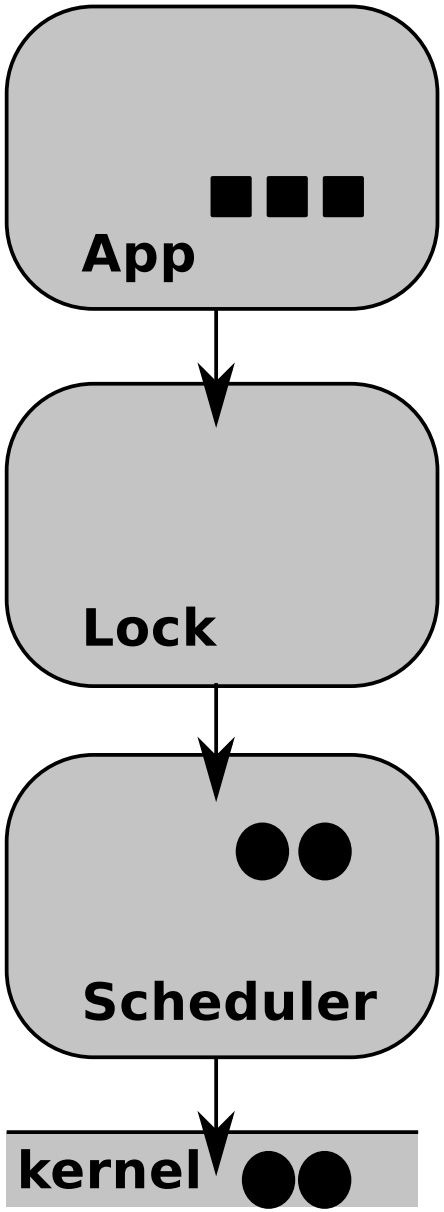


C^3 System-Level Fault Recovery

Recovery sequence in C^3

1. fault detection
2. safe-state recovery
   - μ-reboot

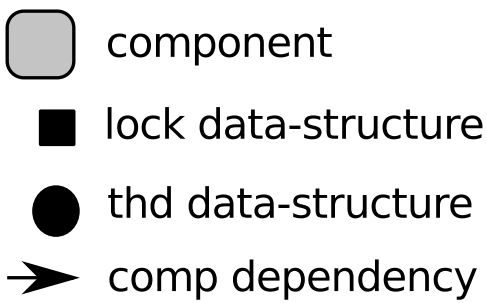


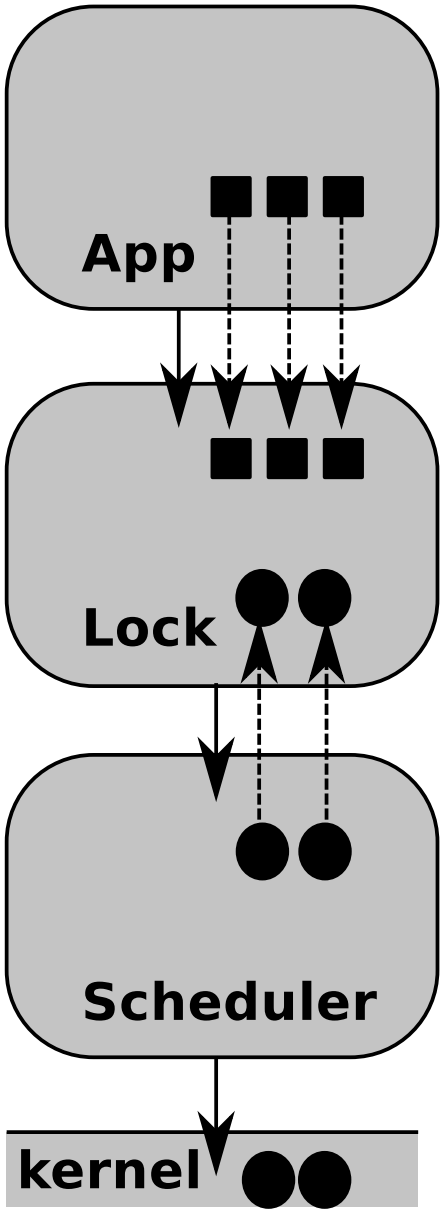


C^3 System-Level Fault Recovery

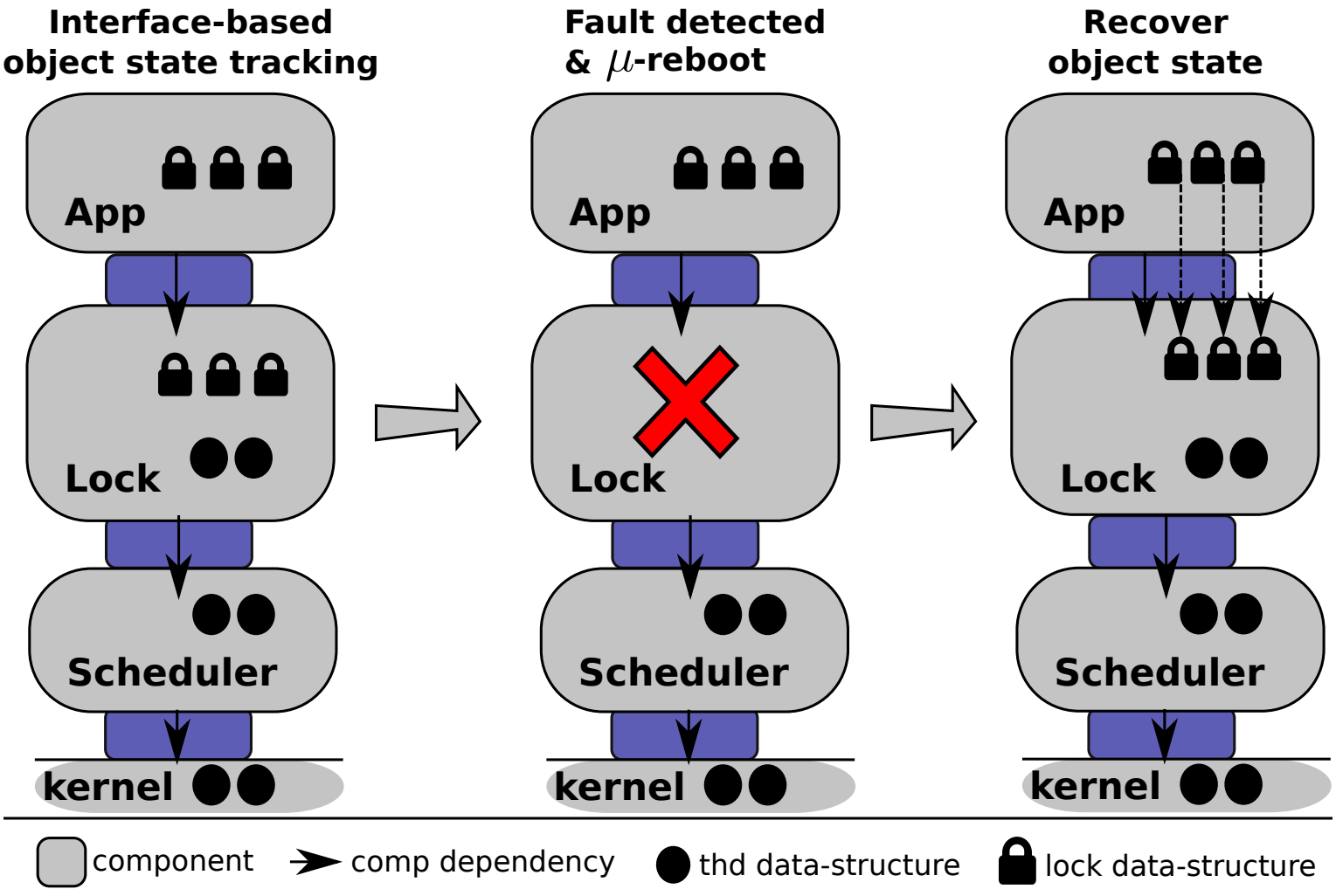
Recovery sequence in C^3

1. fault detection
2. safe-state recovery
   - μ-reboot
3. consistent-state recovery
   - object state recovery

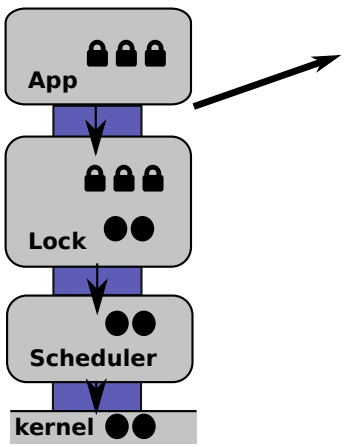




C^3 System-Level Fault Recovery

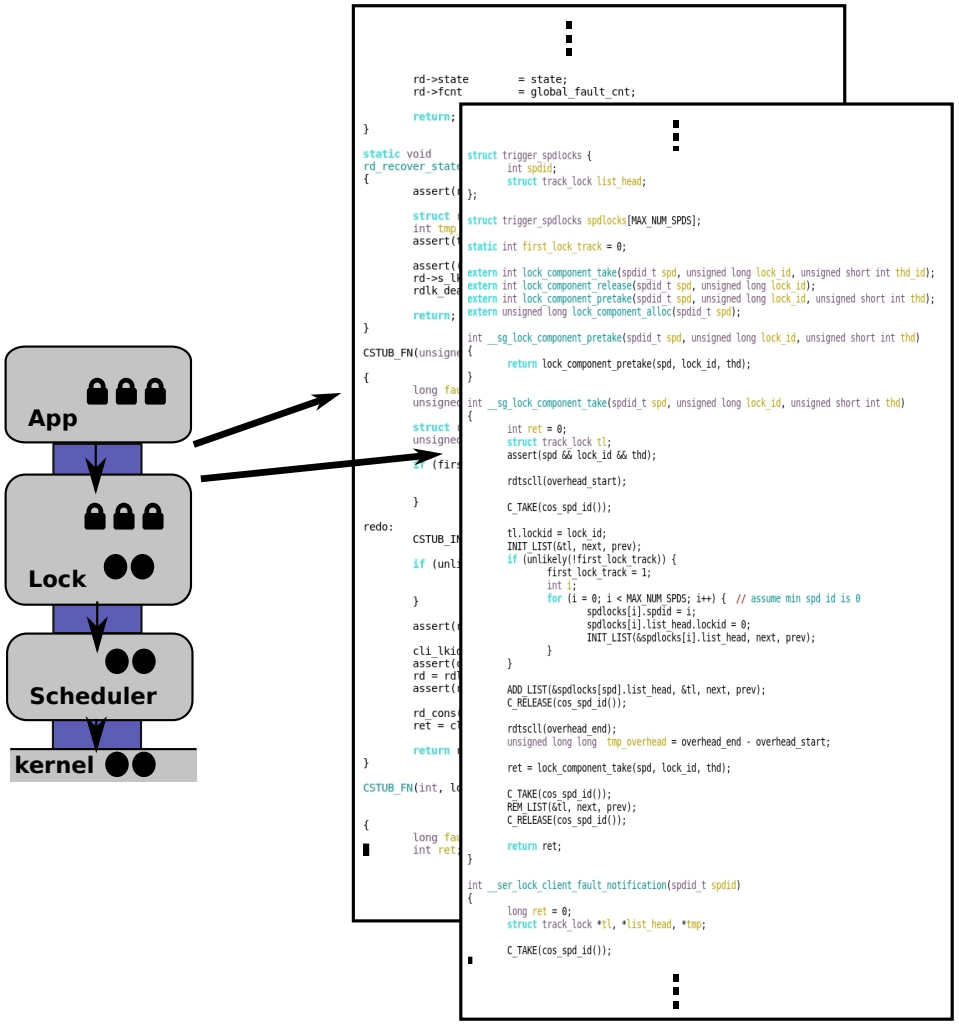


C³ Implementation – Writing Code Manually

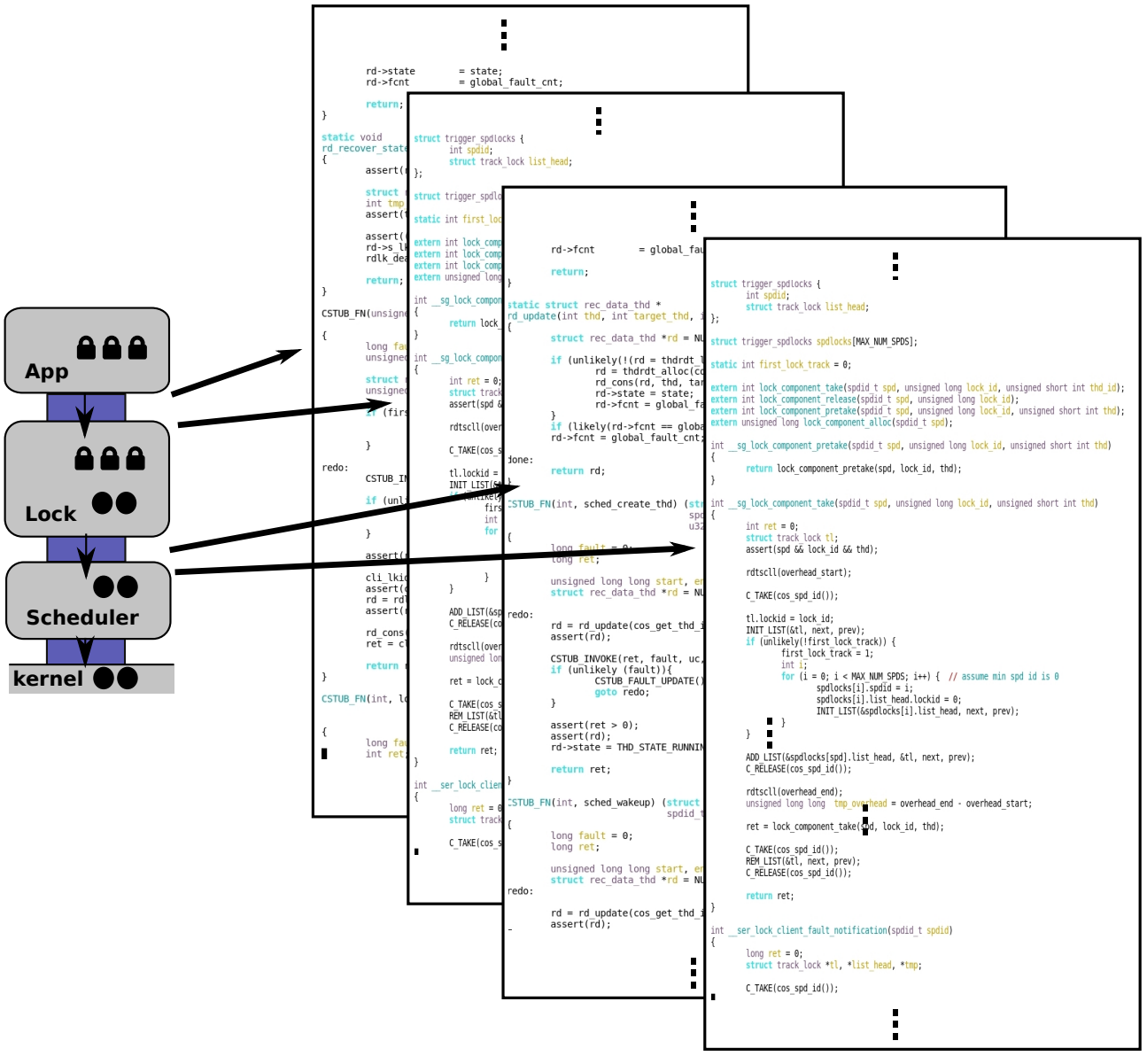



```
        rd->state        = state;
        rd->fcnt         = global_fault_cnt;

        return;
}

static void
rd_recover_state(struct rec_data_lk *rd)
{
        assert(rd && rd->c_lkid);

        struct rec_data_lk *tmp;
        int tmp_lkid = lock_component_alloc(cos_spd_id());
        assert(tmp_lkid);

        assert((tmp = rdlk_lookup(tmp_lkid)));
        rd->s_lkid = tmp->s_lkid;
        rdlk_dealloc(tmp_lkid);

        return;
}

CSTUB_FN(unsigned long, lock_component_alloc) (struct usr_inv_cap *uc,
                                               spdid_t spdid)
{
        long fault;
        unsigned long ret;

        struct rec_data_lk *rd = NULL;
        unsigned long ser_lkid, cli_lkid;

        if (first == 0) {
                cos_map_init_static(&uniq_lkids);
                first = 1;
        }

redo:
        CSTUB_INVOKE(ret, fault, uc, 1, spdid);

        if (unlikely (fault)){
                CSTUB_FAULT_UPDATE();
                goto redo;
        }

        assert(ret > 0);

        cli_lkid = rdlk_alloc();
        assert(cli_lkid >= 1);
        rd = rdlk_lookup(cli_lkid);
        assert(rd);

        rd_cons(rd, cos_spd_id(), cli_lkid, ret, LOCK_ALLOC);
        ret = cli_lkid;

        return ret;
}

CSTUB_FN(int, lock_component_pretake) (struct usr_inv_cap *uc,
                                       spdid_t spdid, unsigned long lock_id,
                                       unsigned short int thd)
{
        long fault = 0;
        int ret;
```

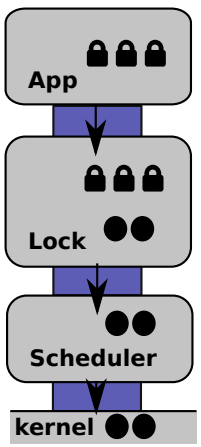
C^3 Implementation – Writing Code Manually

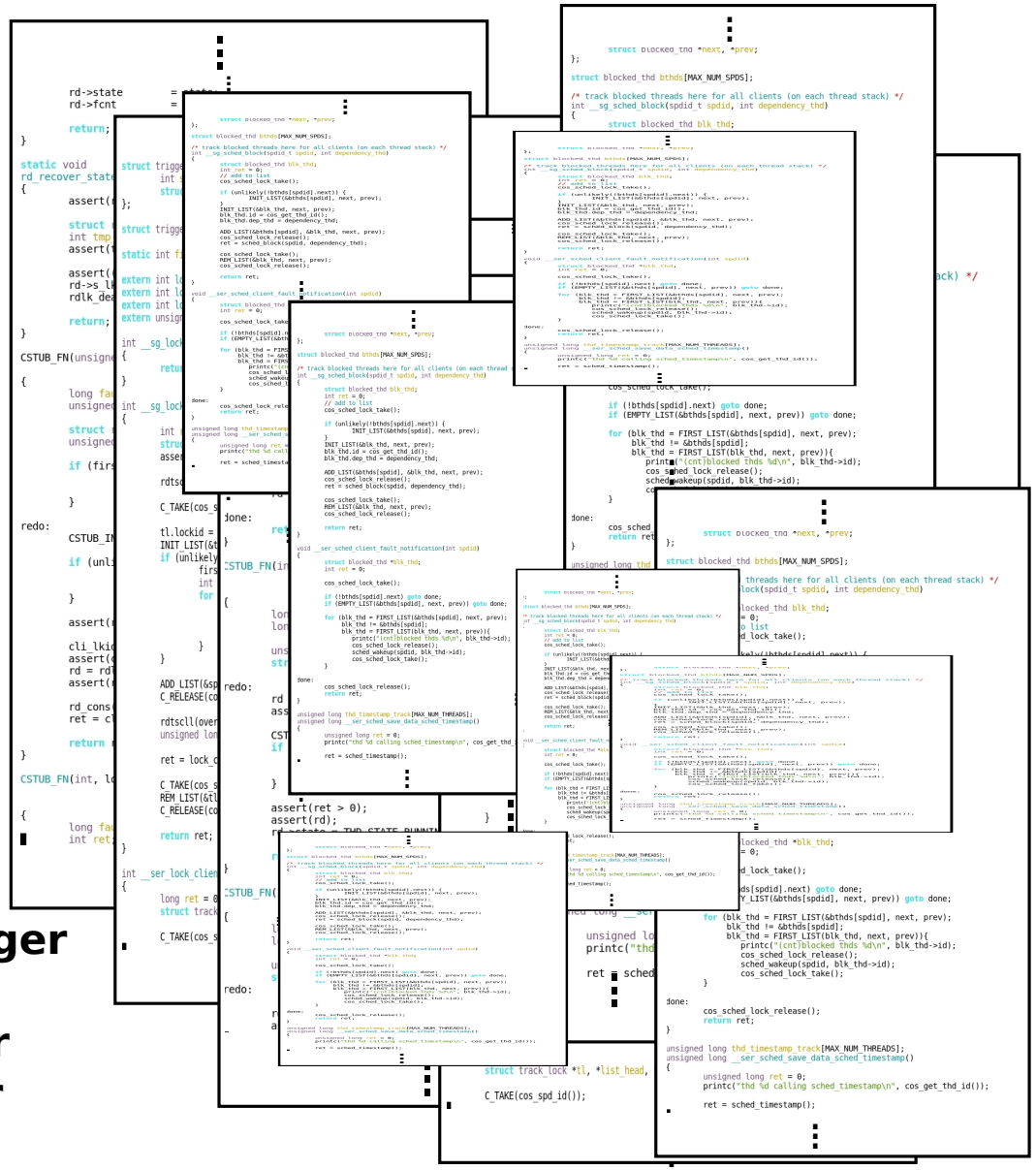


C³ Implementation – Writing Code Manually



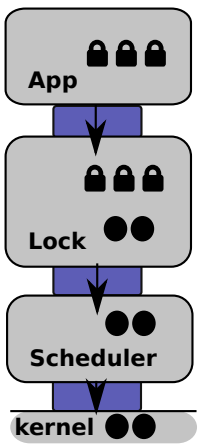
C³ Implementation – Writing Code Manually

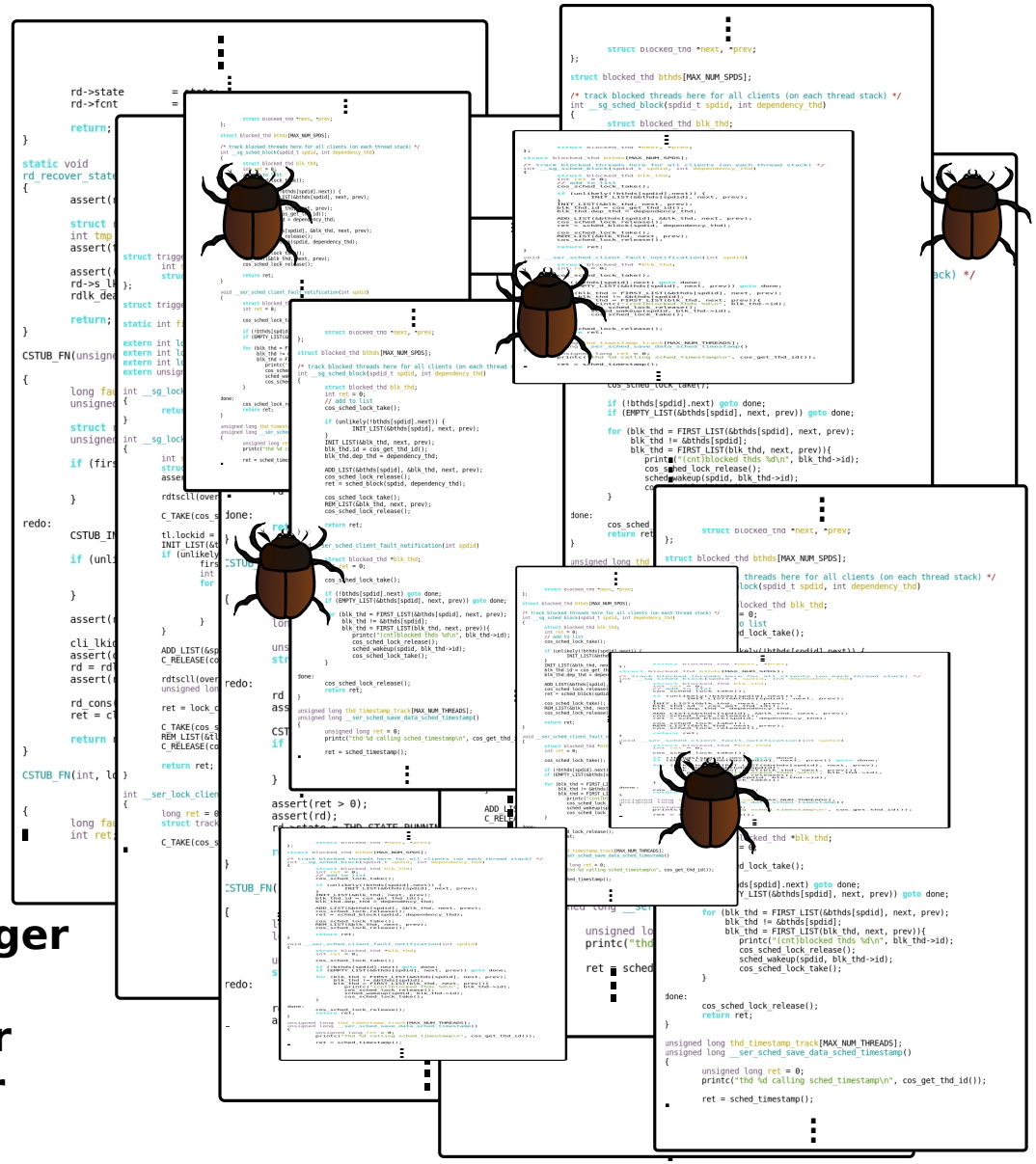




lock
scheduler
memory manager
file system
event manager
timer manager
...

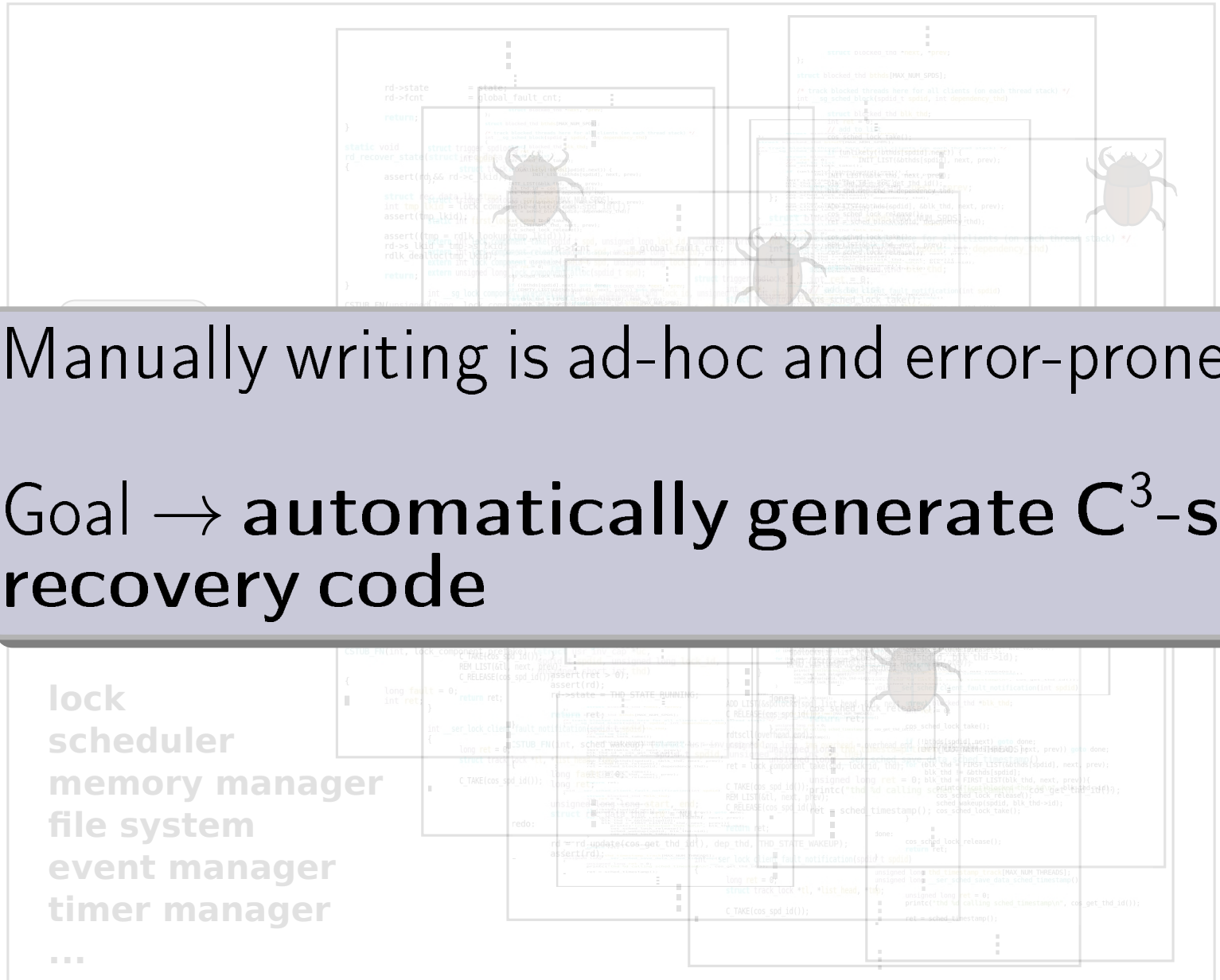
C³ Implementation – Writing Code Manually





lock
scheduler
memory manager
file system
event manager
timer manager
...

C^3 Implementation – Writing Code Manually



Manually writing is ad-hoc and error-prone

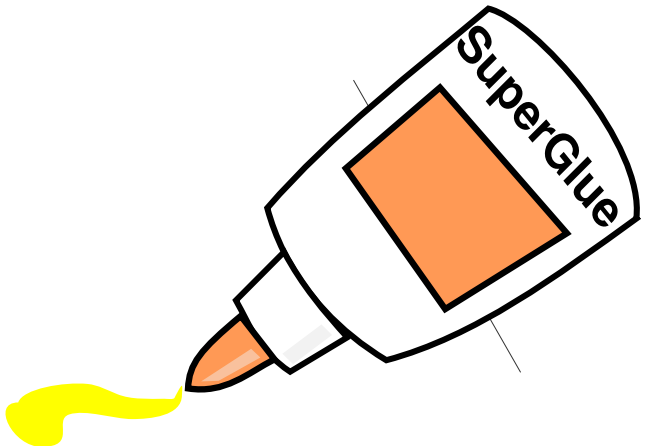
Goal $\rightarrow$ **automatically generate C^3-style recovery code**

lock
scheduler
memory manager
file system
event manager
timer manager
...

SuperGlue – IDL-based System-Level Fault Tolerance

SuperGlue

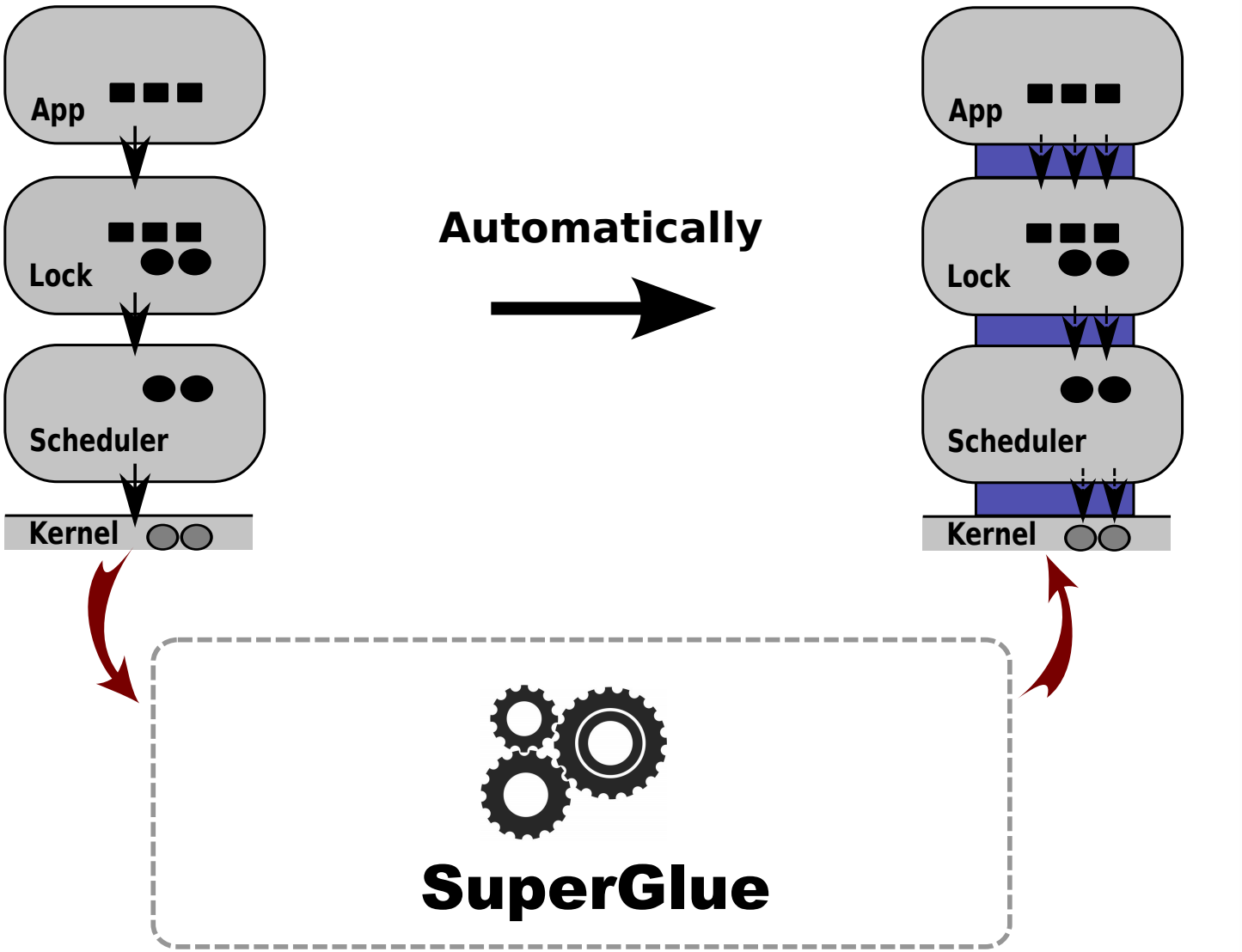
- automatically generate interface-driven recovery code
- for system-level services



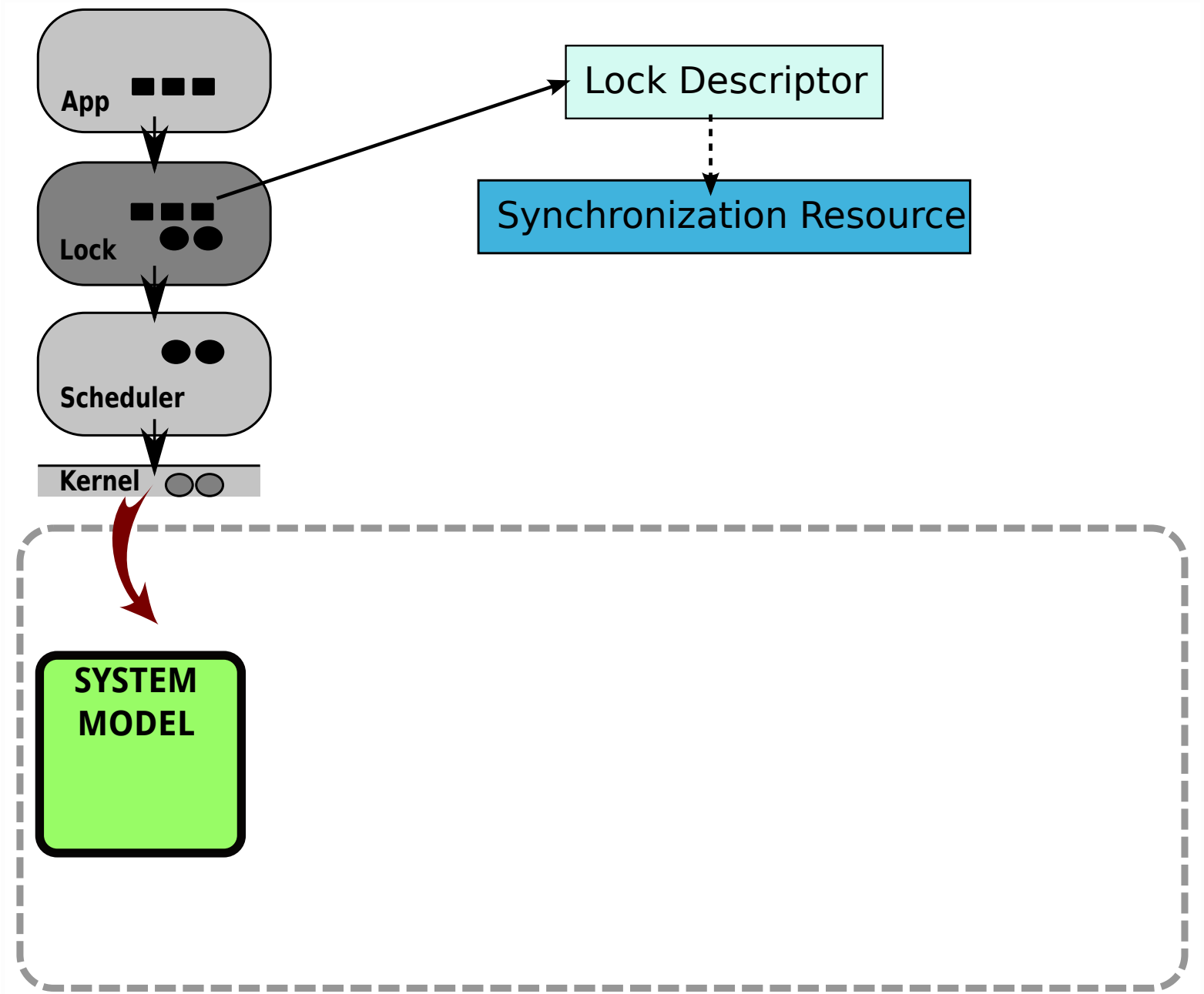
Main ideas

- A system model and interface description language (IDL)
- An IDL compiler synthesizes recovery code from the model

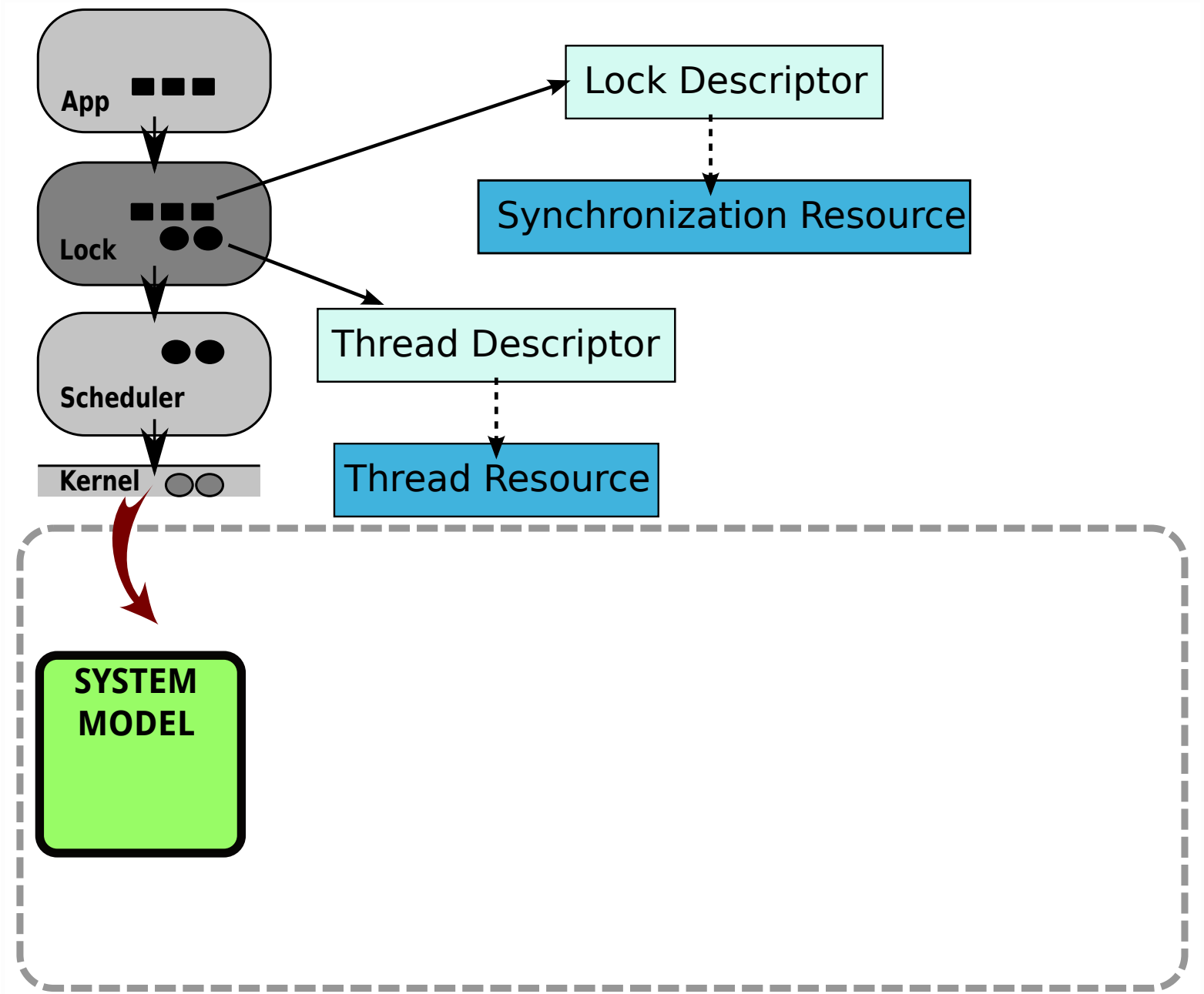
SuperGlue – IDL-based System-Level Fault Tolerance



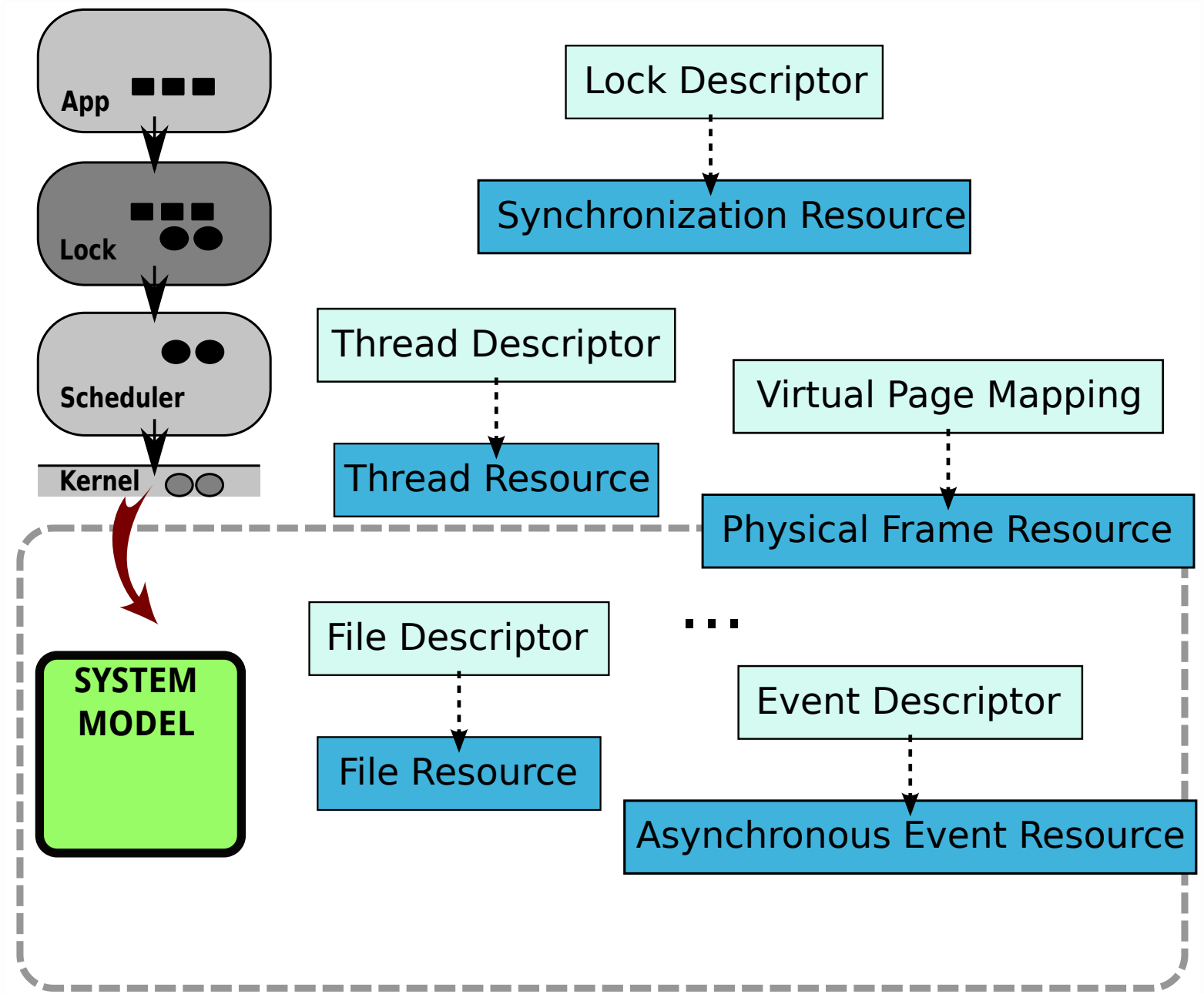
Descriptor-Resource Model



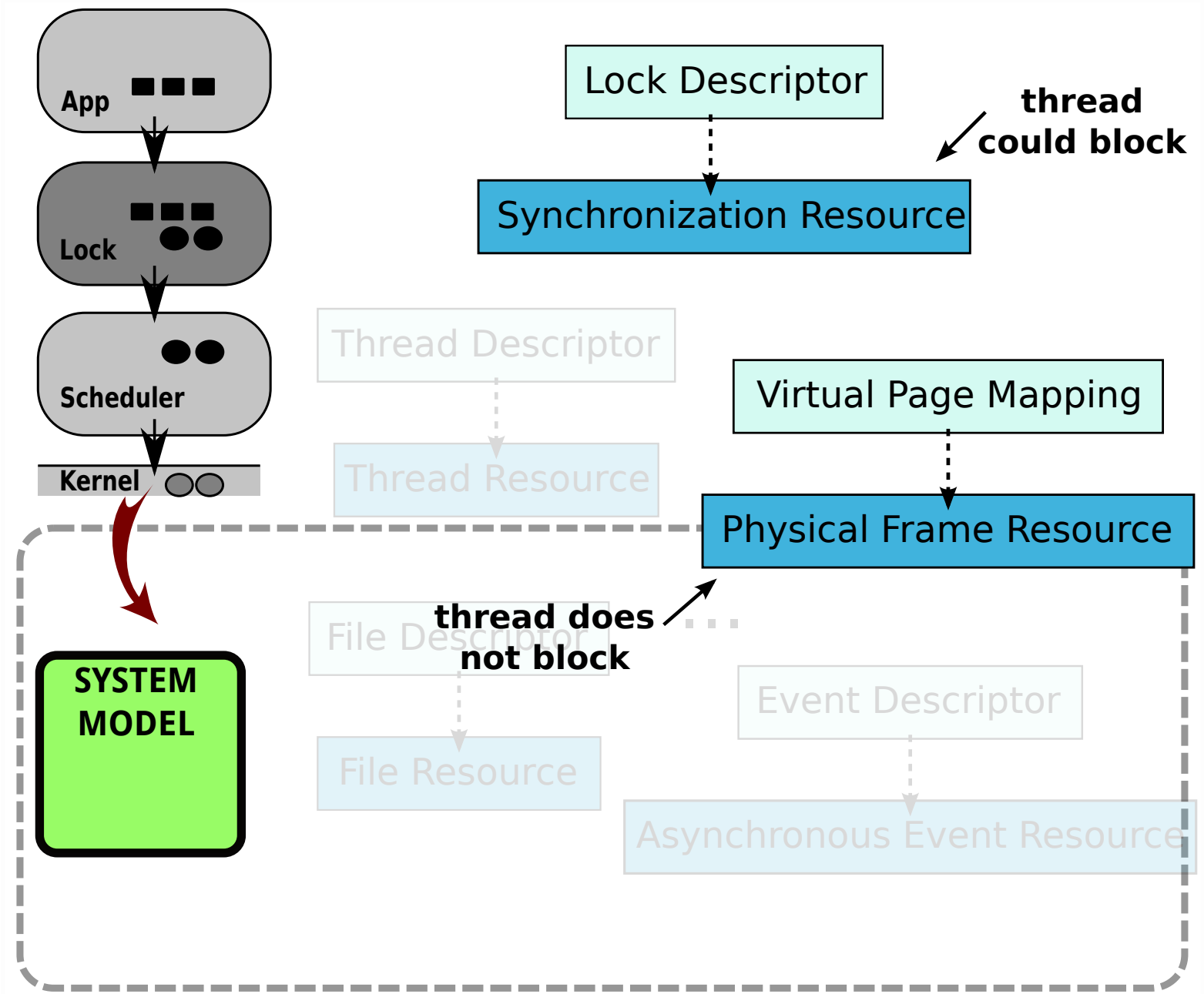
Descriptor-Resource Model



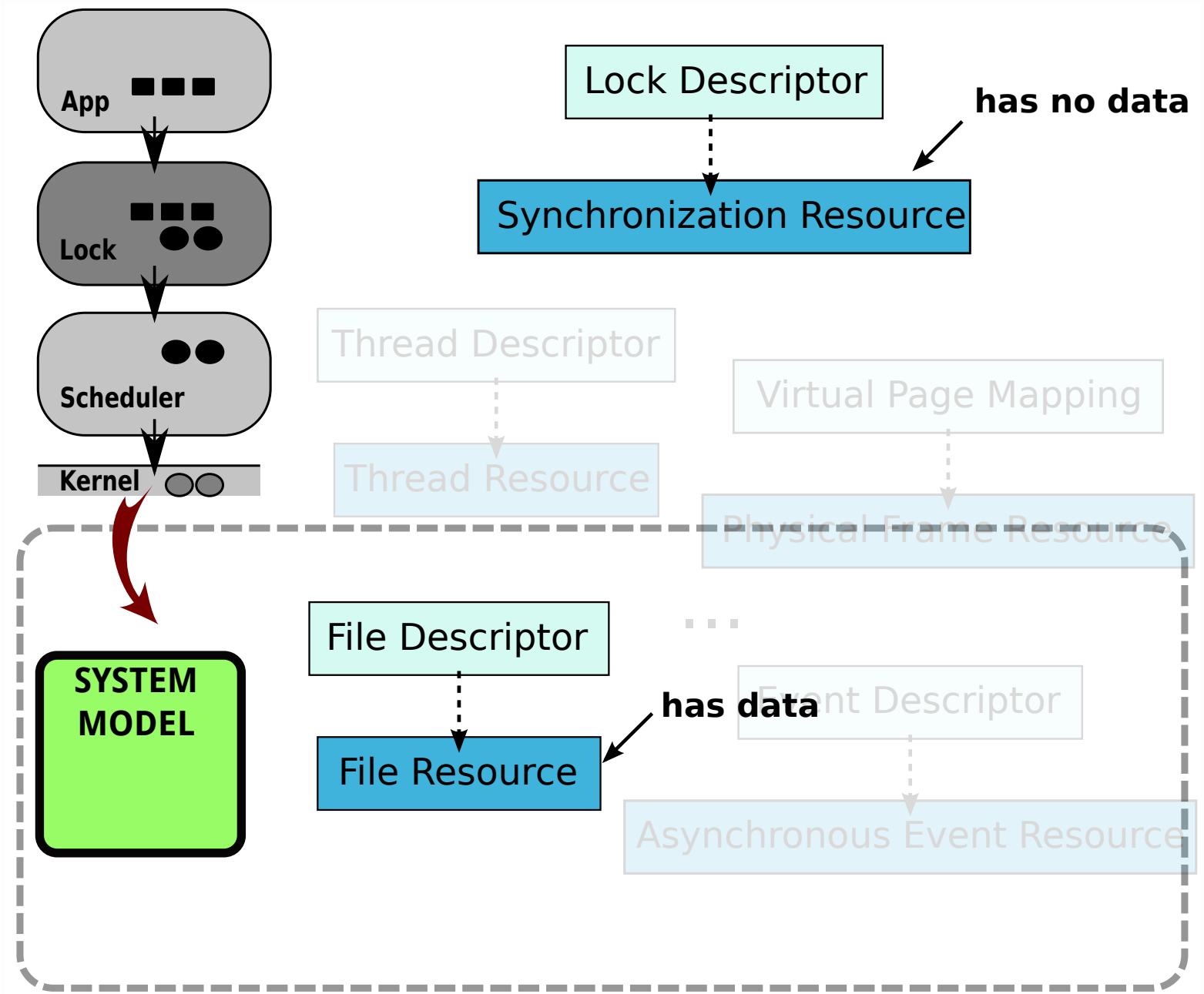
Descriptor-Resource Model



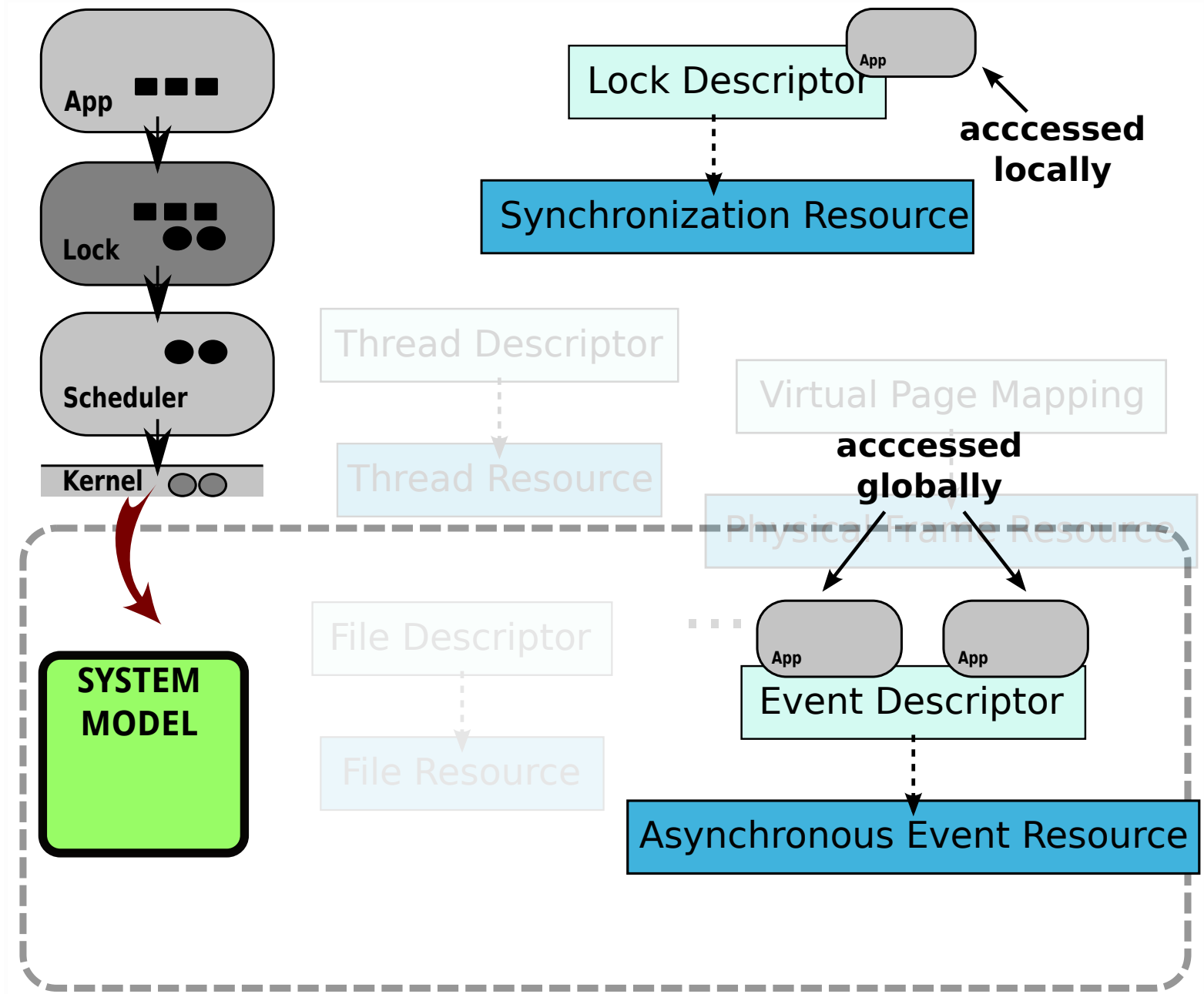
Descriptor-Resource Model



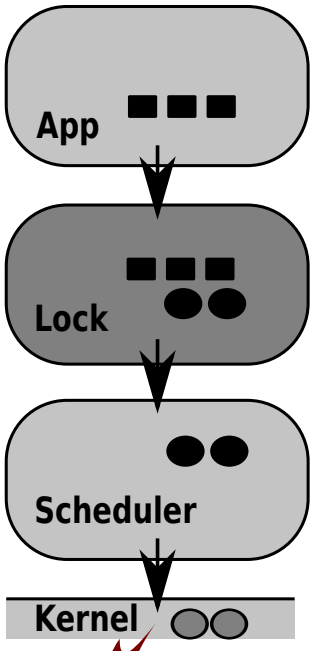
Descriptor-Resource Model



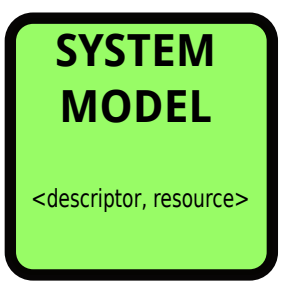
Descriptor-Resource Model



Descriptor-Resource Model



- **Resource**
  - e.g., synchronization, thread, file...
- **Descriptor**
  - e.g., lock, thread, file descriptor...

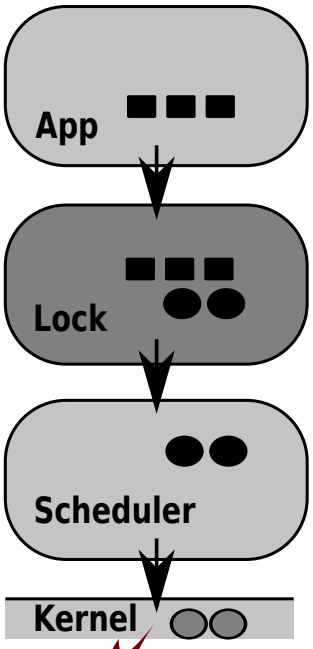


Descriptor-Resource Model --

$$(B_r, D_r, G_d, P_d, C_d, Y_d, D_d)$$

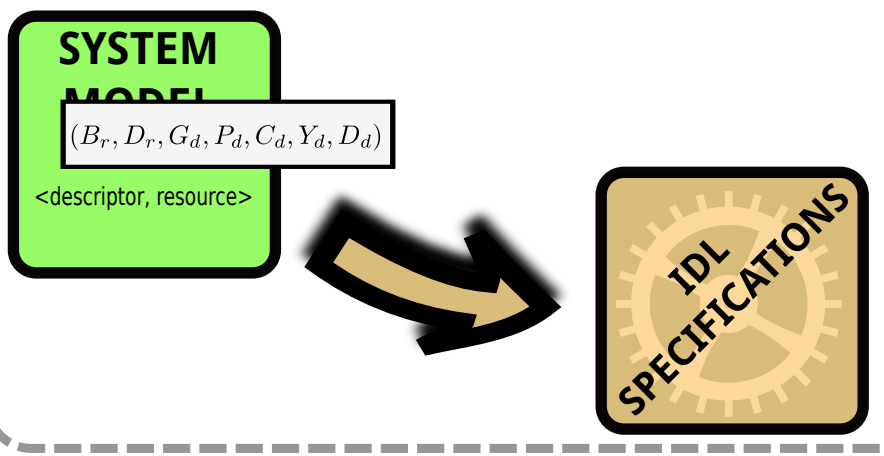
B_r : block when access resource
D_r : resource has meta data
G_d : can be accessed globally
P_d : has dependency descriptor
C_d : delete children when close
Y_d : need to delete tracking data
D_d : descriptor has meta data

Descriptor-Resource Model → IDL

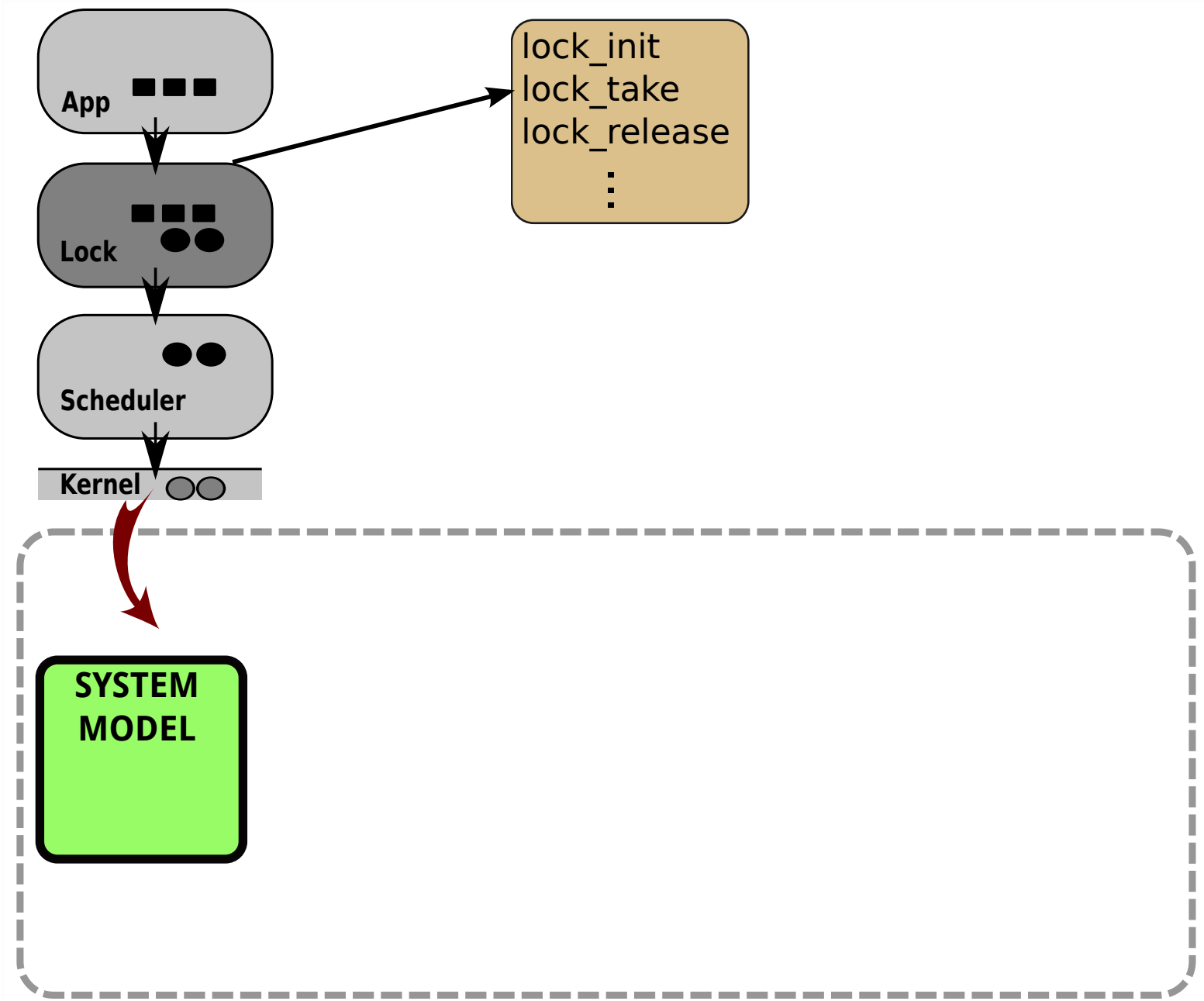



```
service_global_info = {
        service                 = lock,

        desc_close_self_only    = true,
        desc_dep_create_none    = true,
        desc_remove_tracking    = true,
        desc_global             = false,
        desc_block              = true,
        desc_has_data           = false,
        resc_has_data           = false,
};
```

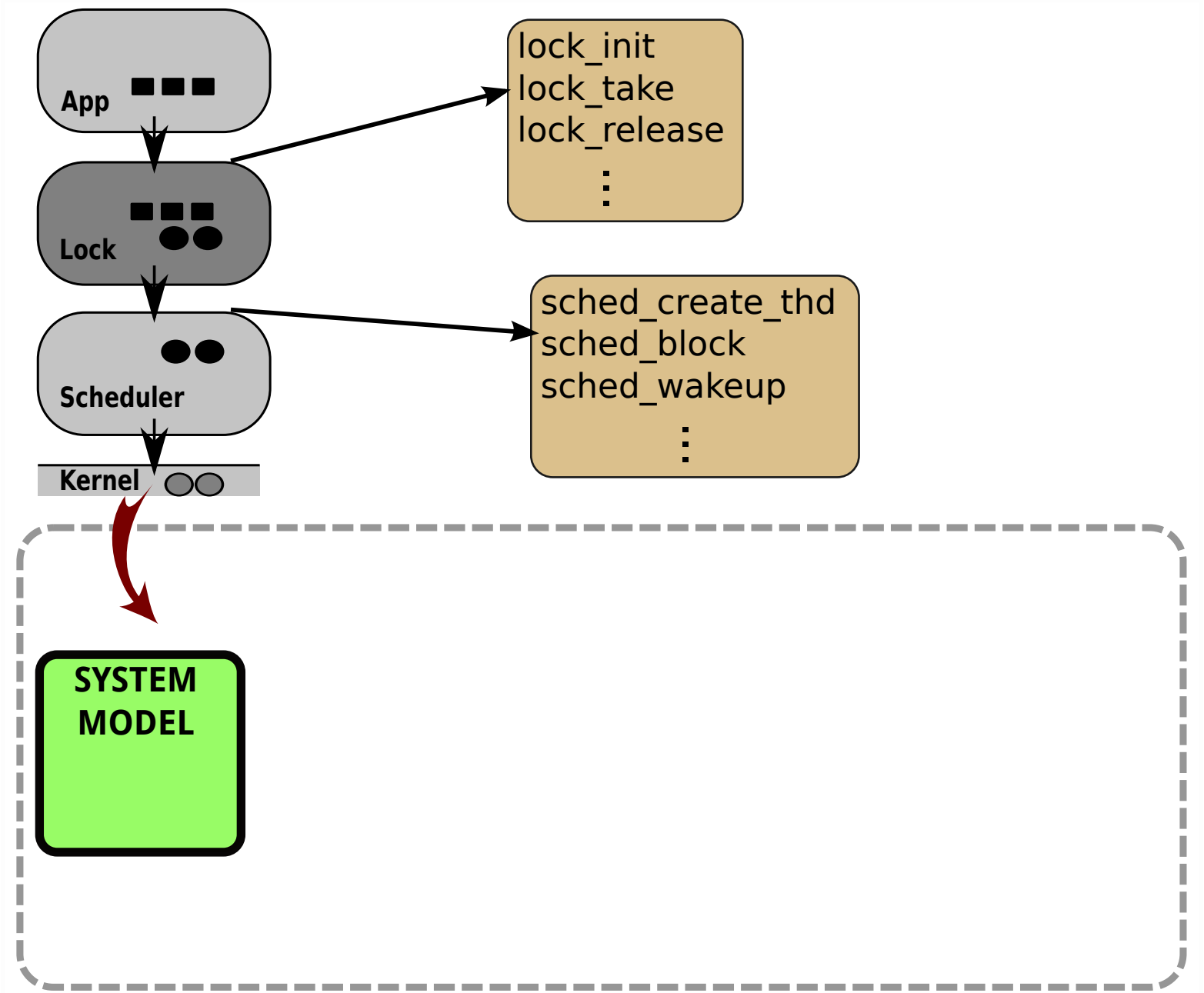
$$(B_r, D_r, G_d, P_d, C_d, Y_d, D_d)$$



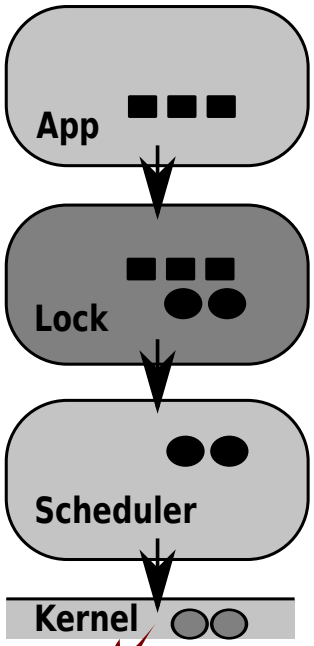
Component Interface Function



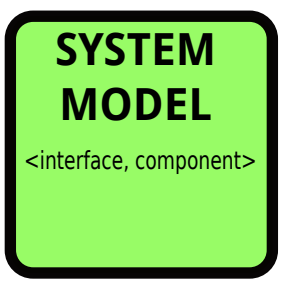
Component Interface Function



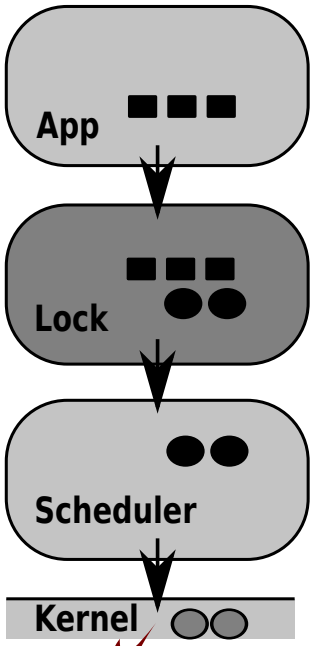
Component Interface Function

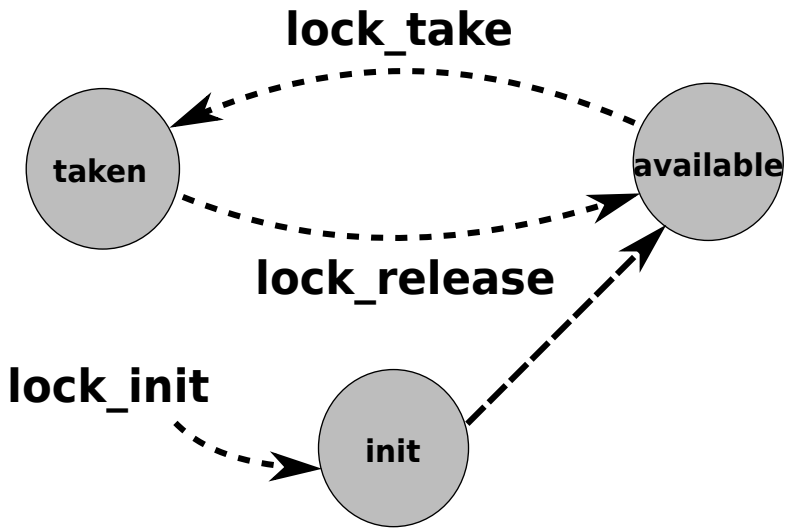


- **create descriptor** (e.g., lock_init)
- **terminate descriptor** (e.g., lock_free)
- **block invoking thread** (e.g., lock_take)
- **wake up thread** (e.g., lock_release)



Descriptor State Machine





lock descriptor state machine

SYSTEM MODEL

<interface, component>
<descriptor, resource>
<state machine>

Descriptor State Machine --

$$(I_d, S_d, s_0, s_f)$$

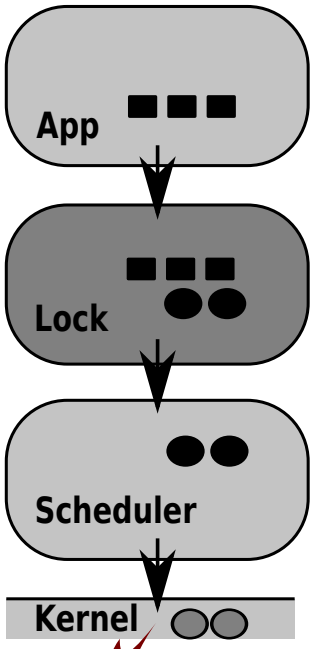
I_d : the set of interface functions
-- creation function
-- termination function
-- block function
-- wake up function

S_d : the set of descriptor states

s_0 : initial state

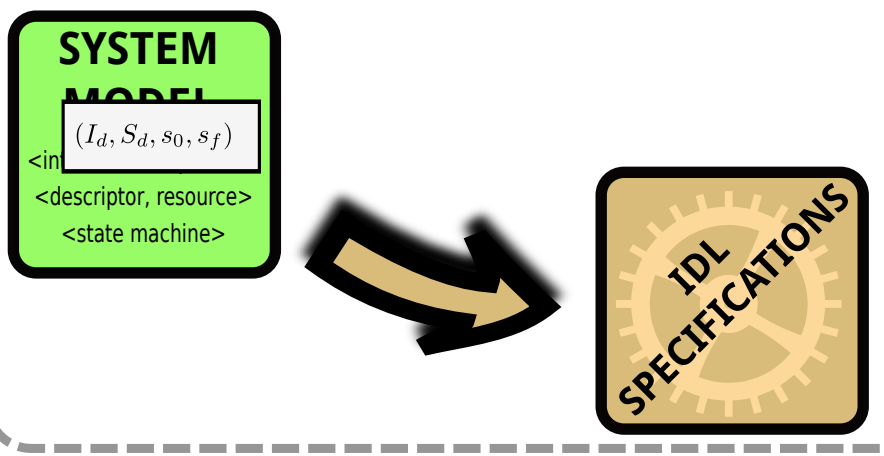
s_f : fault state

Descriptor State Machine → IDL

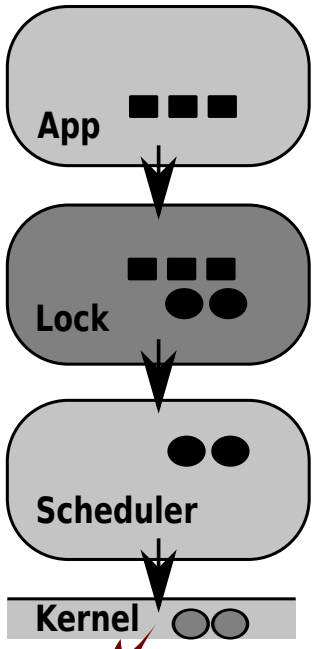



```
sm_creation(lock_component_alloc);
sm_transition(lock_component_alloc, lock_component_free);
sm_transition(lock_component_alloc, lock_component_pretake);
sm_transition(lock_component_pretake, lock_component_take);
sm_transition(lock_component_take, lock_component_release);
sm_transition(lock_component_release, lock_component_pretake);
sm_transition(lock_component_release, lock_component_free);
sm_terminal(lock_component_free);
sm_block(lock_component_take);
sm_wakeup(lock_component_release);
```

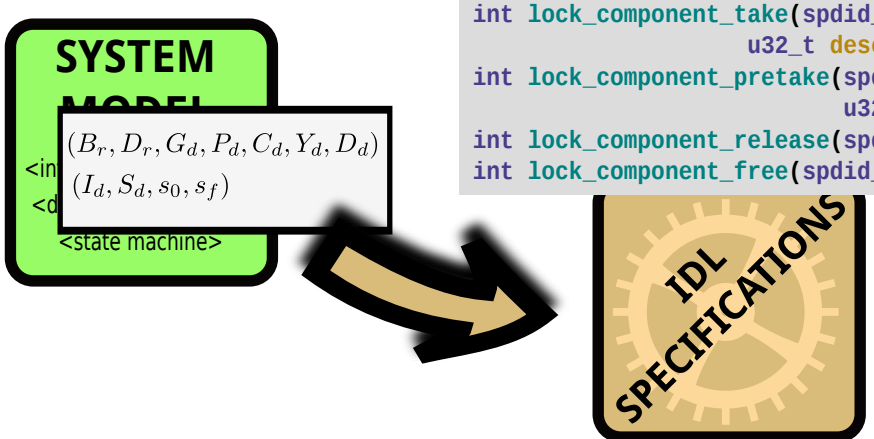
(I_d, S_d, s_0, s_f)



IDL-based Specification

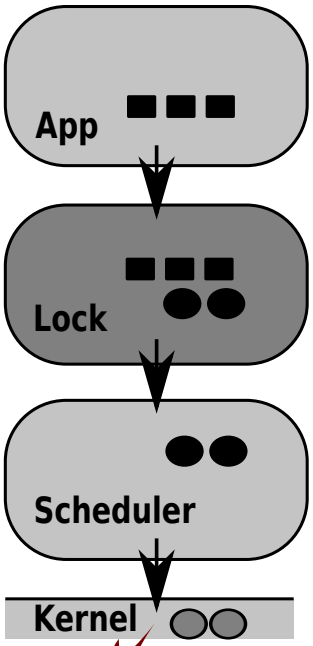



```
service_global_info = {
        service                 = lock,

        desc_close_self_only    = true,
        desc_dep_create_none    = true,
        desc_remove_tracking    = true,
        desc_global             = false,
        desc_block              = true,
        desc_has_data           = false,
        resc_has_data           = false,
};

sm_creation(lock_component_alloc);
sm_transition(lock_component_alloc, lock_component_free);
sm_transition(lock_component_alloc, lock_component_pretake);
sm_transition(lock_component_pretake, lock_component_take);
sm_transition(lock_component_take, lock_component_release);
sm_transition(lock_component_release, lock_component_pretake);
sm_transition(lock_component_release, lock_component_free);
sm_terminal(lock_component_free);
sm_block(lock_component_take);
sm_wakeup(lock_component_release);

desc_data_retval(ul_t, lock_id)
lock_component_alloc(spdid_t desc_data(spdid));
int lock_component_take(spdid_t spdid, ul_t desc(lock_id),
                        u32_t desc_data(thd_id));
int lock_component_pretake(spdid_t spdid, ul_t desc(lock_id),
                           u32_t thd_id);
int lock_component_release(spdid_t spdid, ul_t desc(lock_id));
int lock_component_free(spdid_t spdid, ul_t desc_terminate(lock_id));
```



Avg LOC = 37

Toward Generating Recovery Code

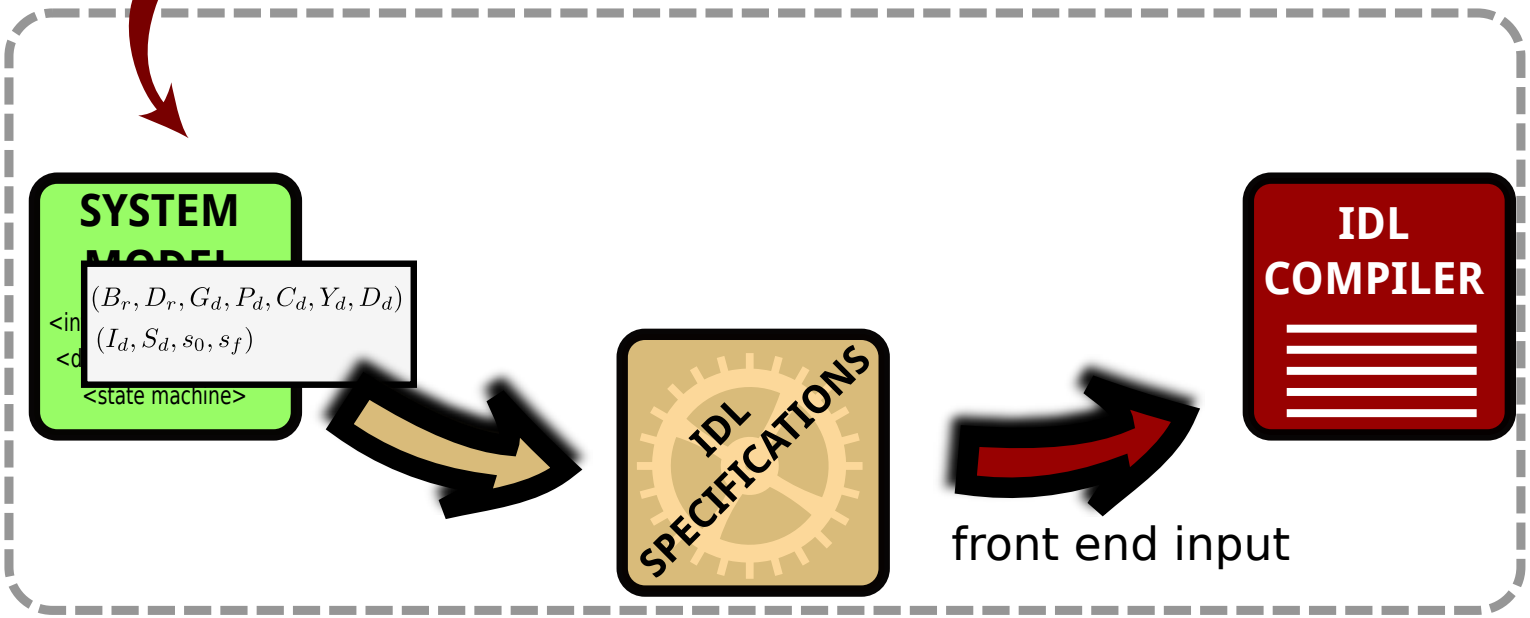


front end

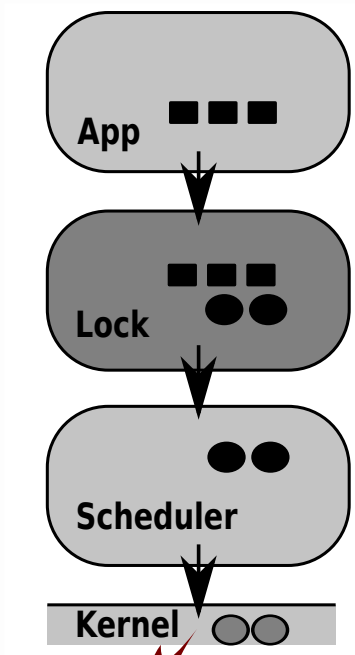
- parse IDL specification

intermediate representation

- encode the model information

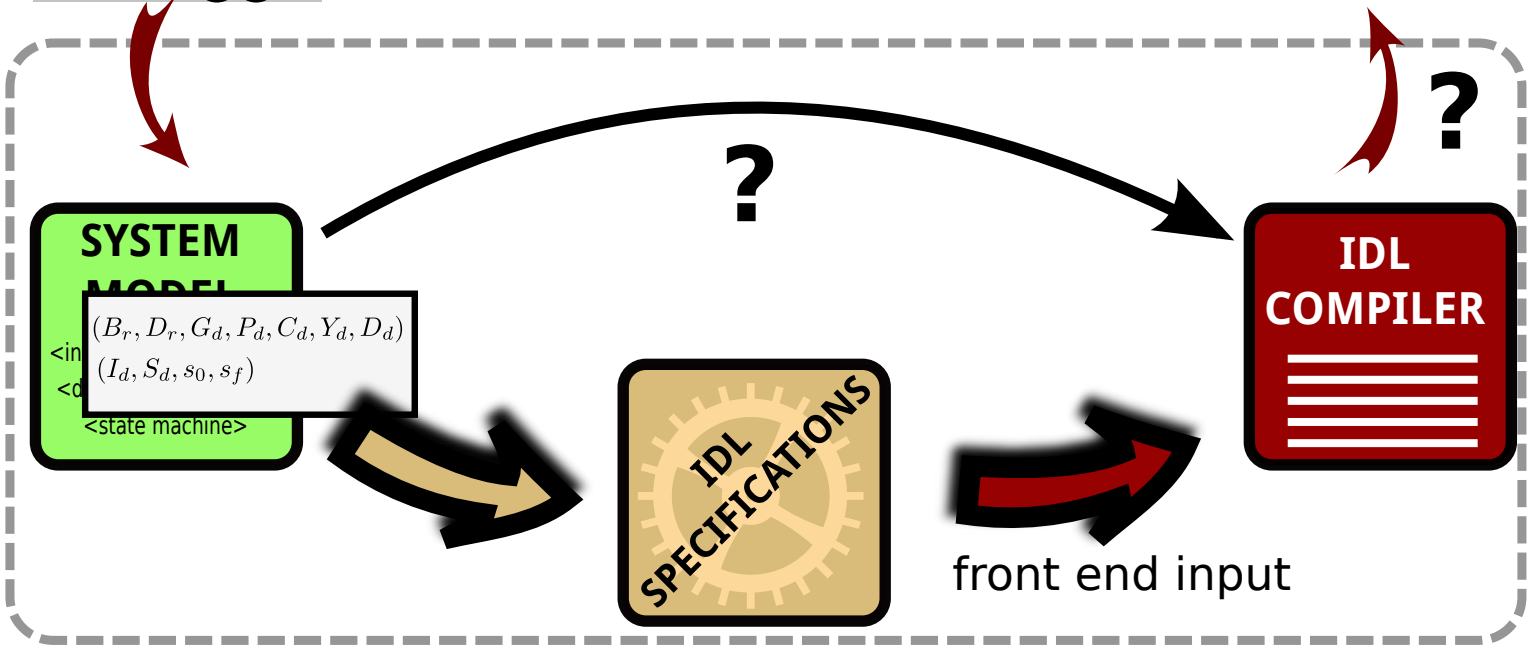


Toward Generating Recovery Code

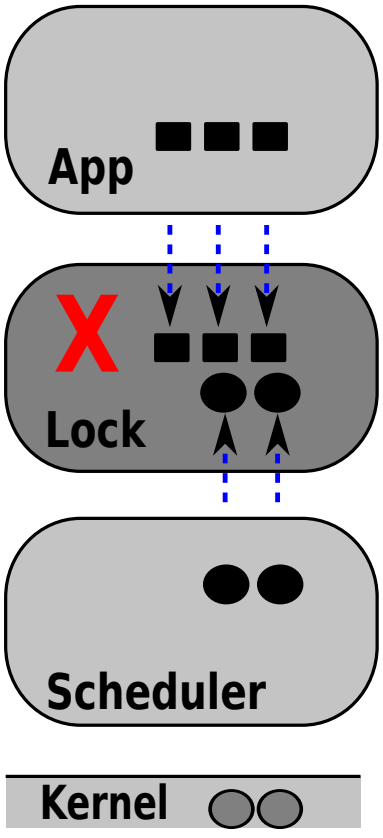


back end

- what recovery mechanism shoud be used?
- how to synthesize the code?



Interface-Driven Recovery Mechanisms



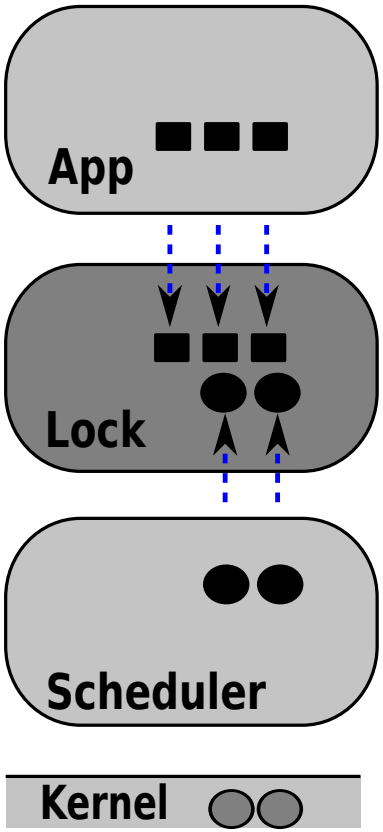
basic recovery

-- through component operation (**always**)

---

basic recovery

Interface-Driven Recovery Mechanisms



basic recovery

-- through component operation (**always**)

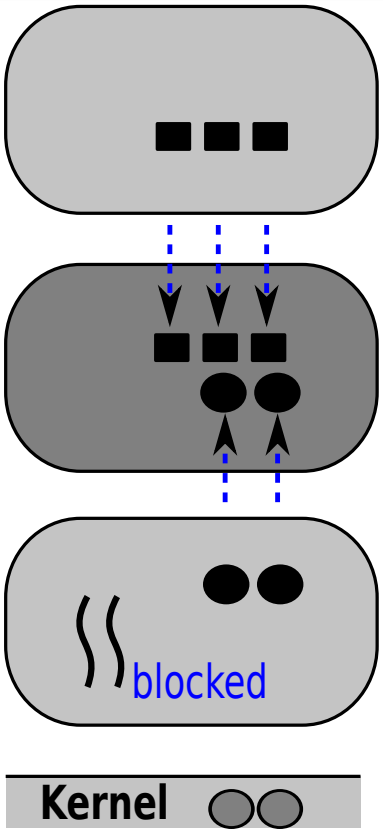
timing of recovery - on-demand

-- recover only when accessed (**always**)

---

basic recovery | timing of recovery

Interface-Driven Recovery Mechanisms



basic recovery

-- through component operation (**always**)

timing of recovery - on-demand

-- recover only when accessed (**always**)

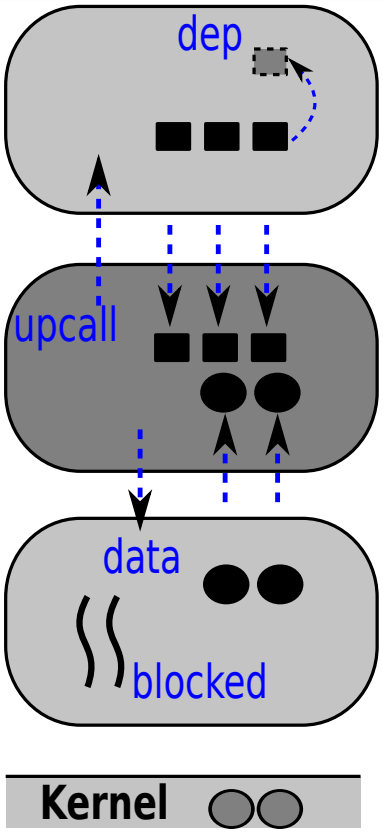
timing of recovery - eager

-- eagerly wake up blokcing threads (B_r)

---

basic recovery | timing of recovery

Interface-Driven Recovery Mechanisms



basic recovery

-- through component operation (**always**)

timing of recovery - on-demand

-- recover only when accessed (**always**)

timing of recovery - eager

-- eagerly wake up blokcing threads (B_r)

recovery with dependency

-- require to recontruct parent (P_d)

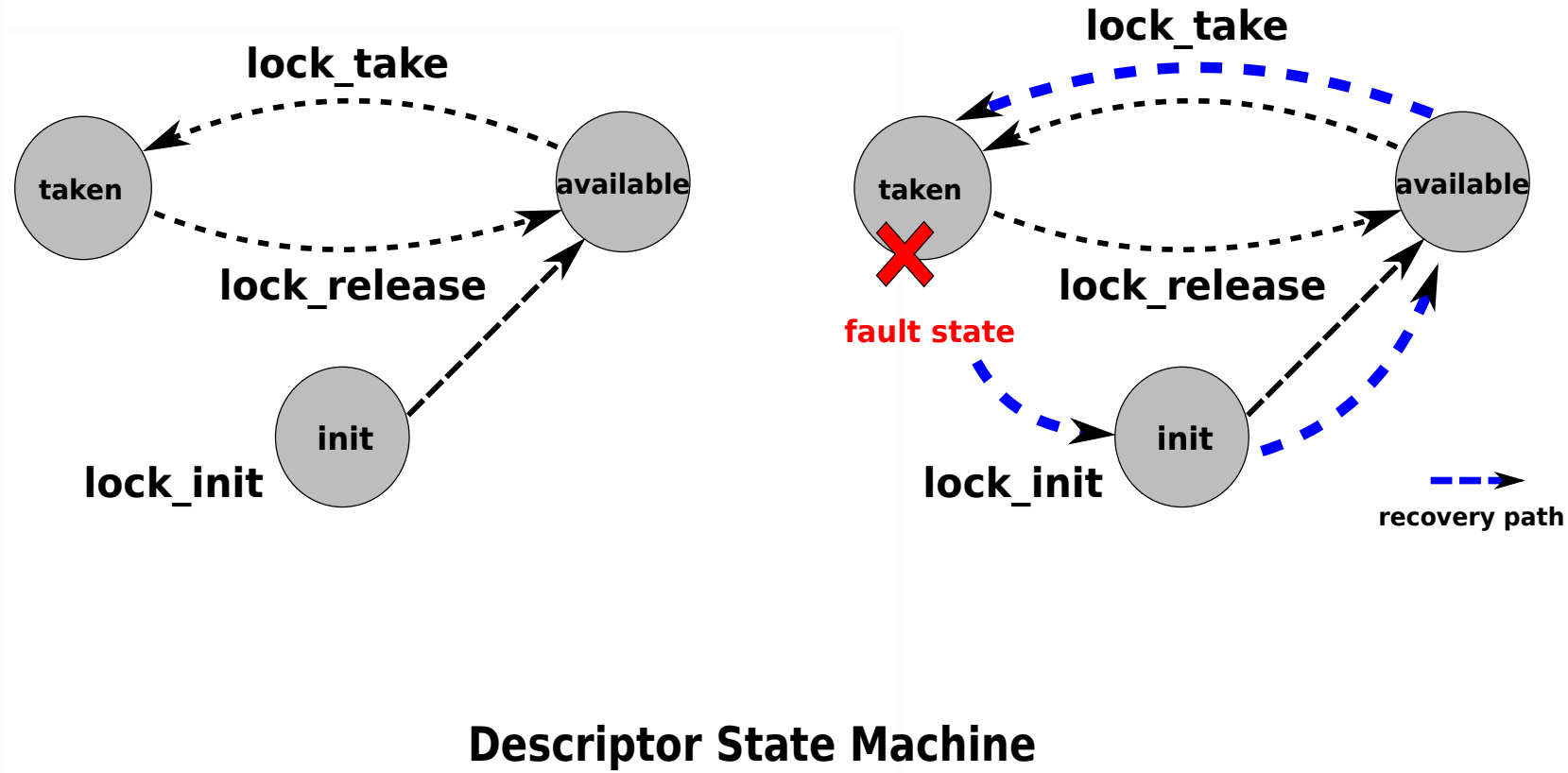
-- require to recontruct children (C_d)

⋮

---

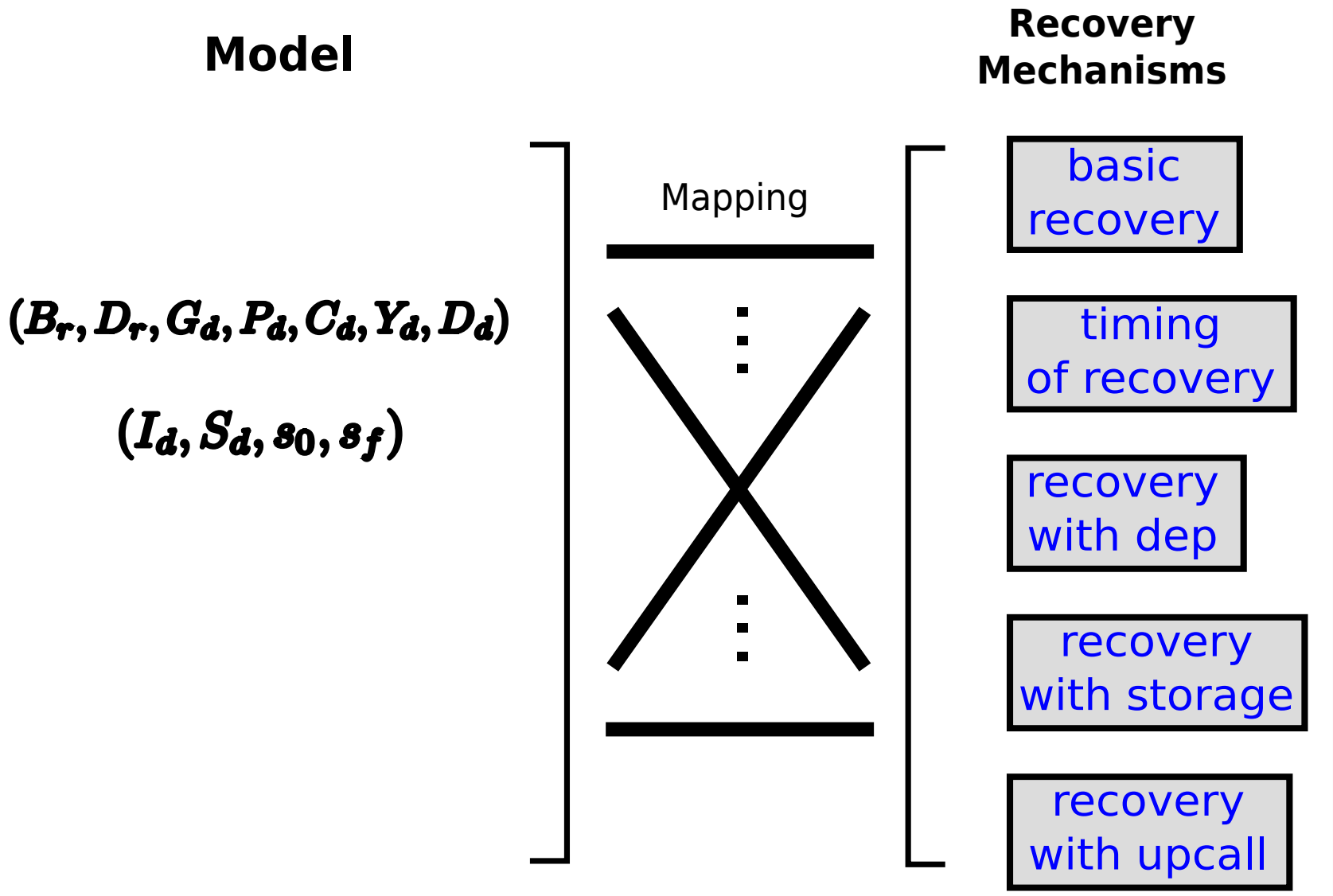
basic recovery	timing of recovery	recovery with dep	recovery with storage	recovery with upcall

Interface-Driven Recovery Mechanisms



Descriptor State Machine

Model → Recovery Mechanisms

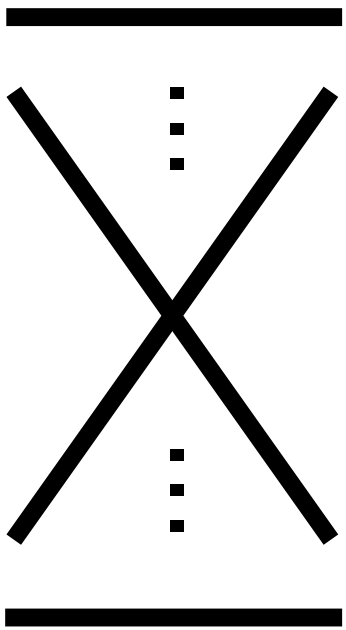


Synthesize the Recovery Code

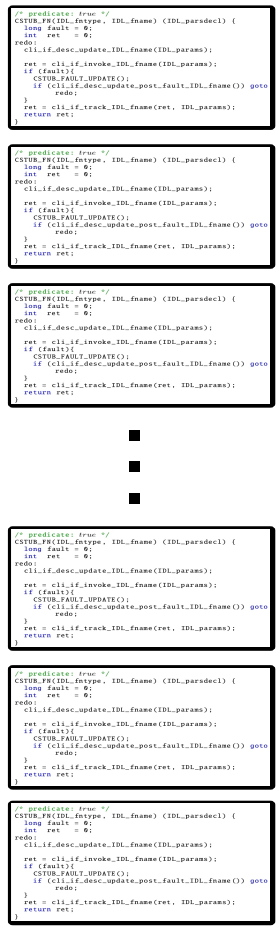
Predicates

$$\vdots$$
$$true$$
$$create\_fn \wedge \neg G_d$$
$$wakeup\_fn \wedge B_r$$
$$block\_fn \wedge B_r$$
$$terminate\_fn \wedge C_d$$
$$G_d \wedge P_d \wedge C_d \wedge D_d$$
$$D_r \wedge \neg P_d \wedge Y_d$$
$$\vdots$$

Mapping



Templates



Synthesize the Recovery Code

Predicates **Templates**

$true$

$create\_fn \wedge$

$wakeup\_fn \wedge$

$block\_fn \wedge$

$terminate\_$

$G_d \wedge P_d \wedge C$

$D_r \wedge \neg P_d \wedge$

Mapping

```
/* predicate: true */
CSTUB_FN(IDL_fntype, IDL_fname) (IDL_parsdecl) {
  long fault = 0;
  int  ret   = 0;
redo:
  cli_if_desc_update_IDL_fname(IDL_params);

  ret = cli_if_invoke_IDL_fname(IDL_params);
  if (fault){
    CSTUB_FAULT_UPDATE();
    if (cli_if_desc_update_post_fault_IDL_fname()) goto
        redo;
  }
  ret = cli_if_track_IDL_fname(ret, IDL_params);
  return ret;
}
```

Synthesize the Recovery Code

Predicates **Templates**

$true$

$create\_fn \wedge$

$wakeup\_fn \wedge$

$block\_fn \wedge E$

$terminate\_f$

$G_d \wedge P_d \wedge C$

$D_r \wedge \neg P_d \wedge$

Mapping

```
/* predicate: true */
CSTUB_FN(IDL_fntype, IDL_fname) (IDL_parsdecl) {
  long fault = 0;
  int  ret   = 0;
redo:
  cli_if_desc_update_IDL_fname(IDL_params);

  ret = cli_if_invoke_IDL_fname(IDL_params);
  if (fault){
    CSTUB_FAULT_UPDATE();
    if (cli_if_desc_update_post_fault_IDL_fname()) goto
        redo;
  }
  ret = cli_if_track_IDL_fname(ret, IDL_params);
  return ret;
}
```

Synthesize the Recovery Code

Predicates **Templates**

$\vdots$

$true$

$create\_fn \wedge \neg G_d$

$wakeup\_fn \wedge$

$block\_fn \wedge$

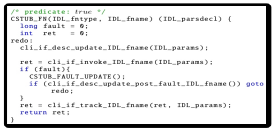
$terminate$

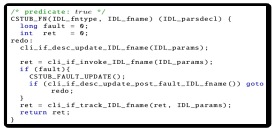
$G_d \wedge P_d \wedge$

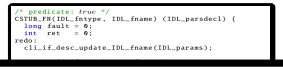
$D_r \wedge \neg P_d$

Mapping

$\vdots$

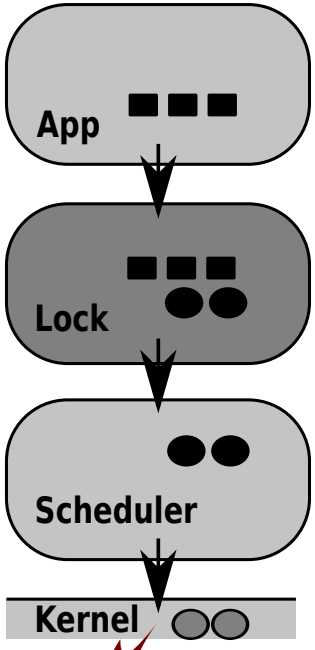






```
/* predicate: f ∈ I_{d_r}^{create} ∧ ¬G_{d_r} */
static inline int cli_if_track_IDL_fname(int ret,
                                         IDL_parsdecl) {
  if (ret == -EINVAL) return ret;

  struct desc_track *desc = call_desc_alloc();
  if (!desc) return -ENOMEM;
  call_desc_track(desc, ret, IDL_params);

  return desc->IDL_id;
}
```

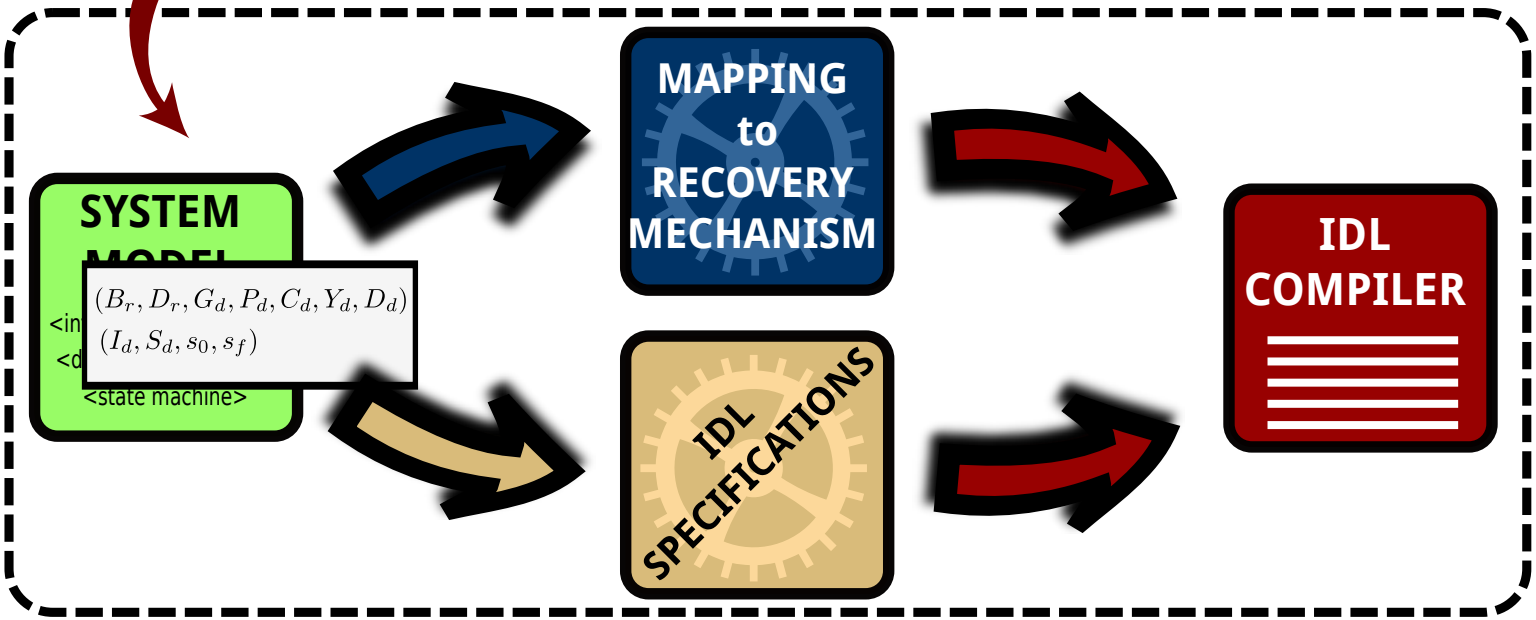
Generate Recovery Code



front end
- parse IDL-based specification

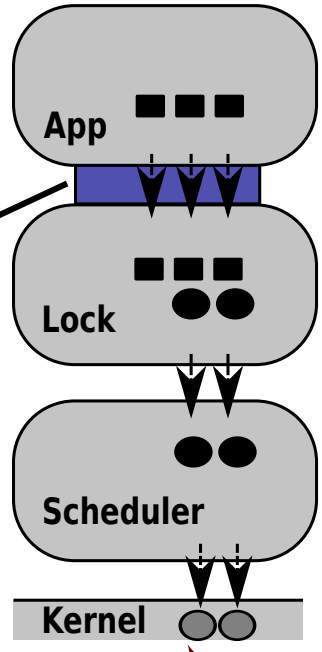
intermediate representation
- encode the model information

back end
- evaluate predicates
- generate the code from templates

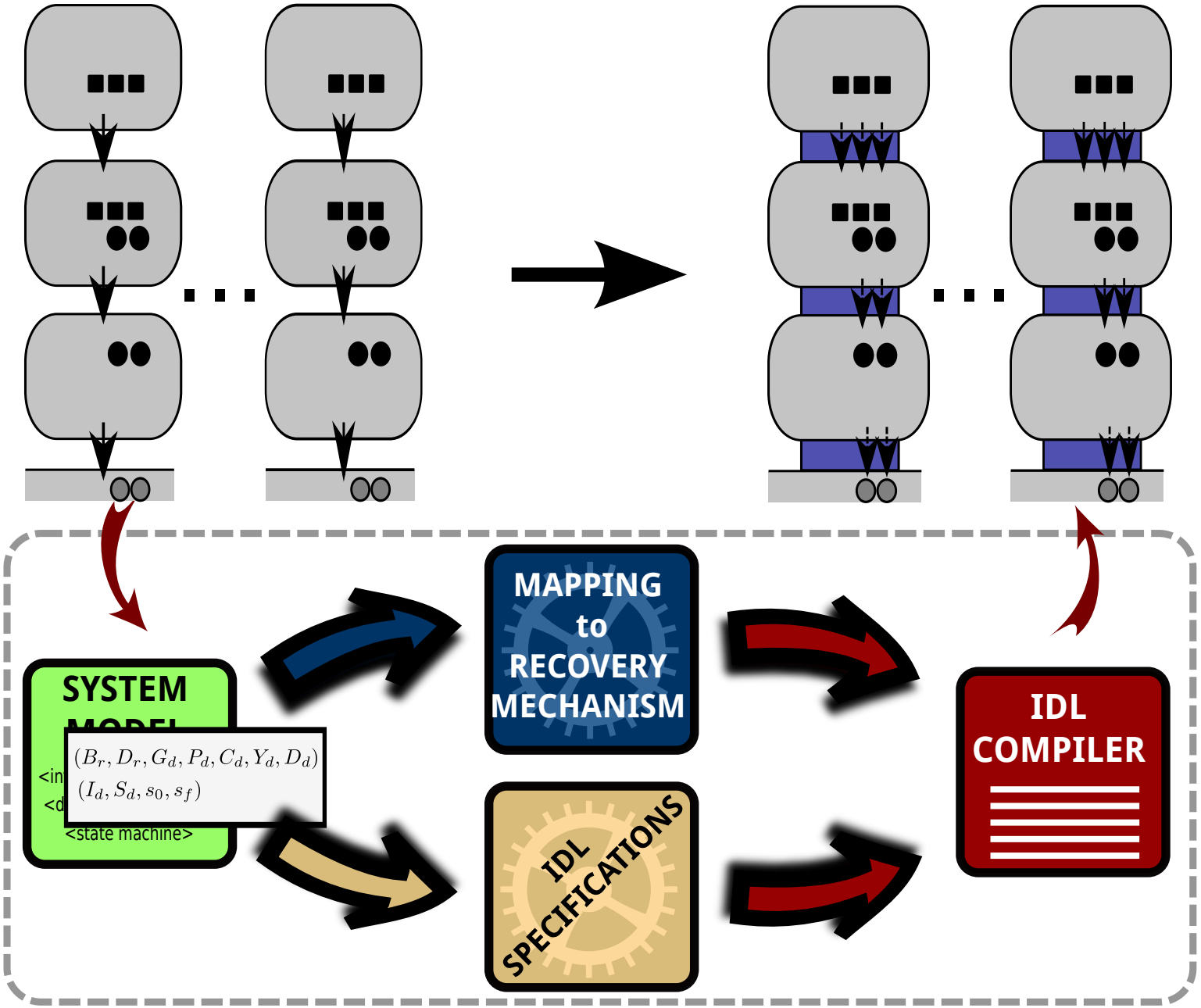


Generate Recovery Code

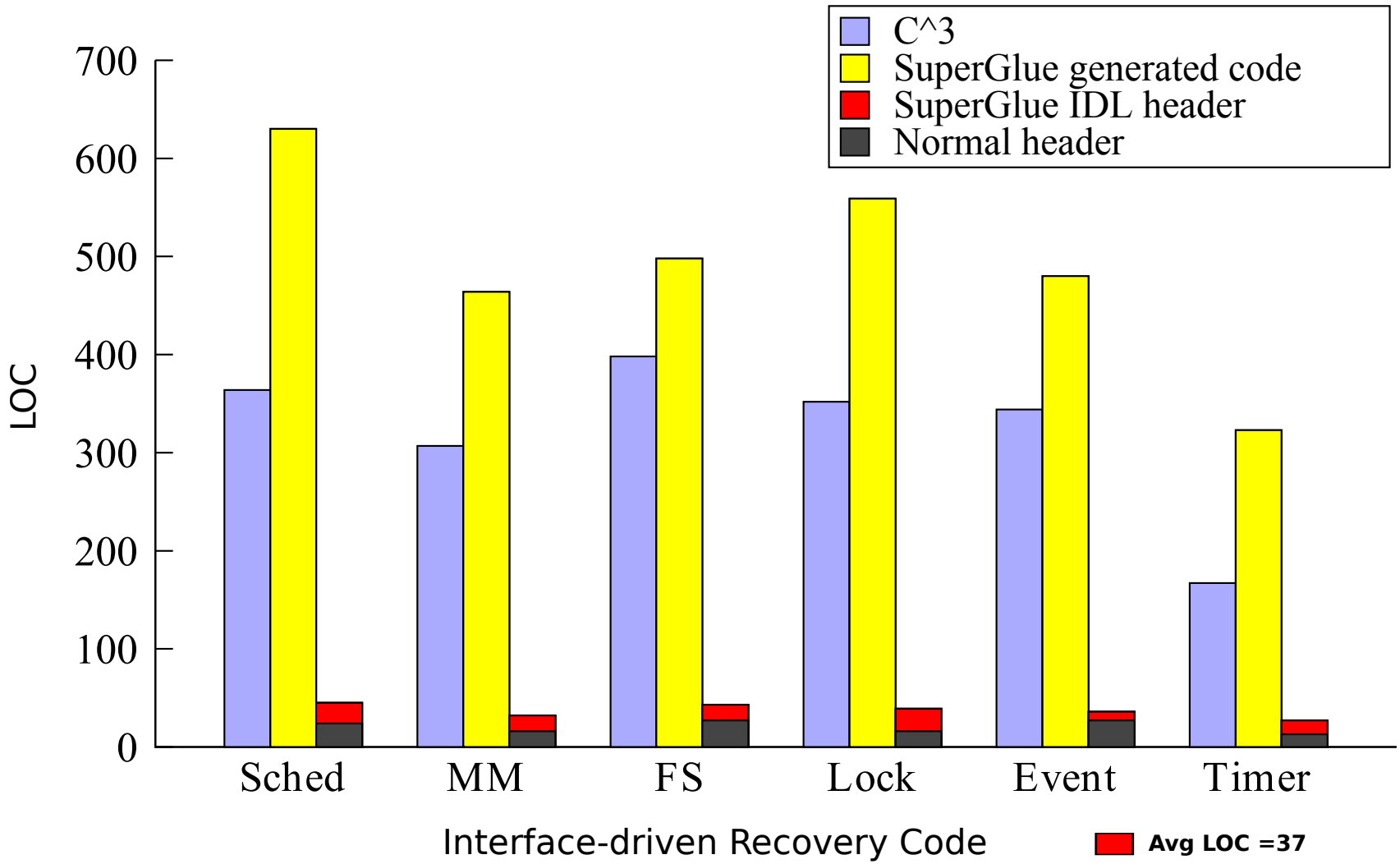
```
        desc->server_lock_id = new_desc->server_lock_id;
        desc->state = state_lock_component_alloc;
        desc->fault_cnt = global_fault_cnt;

        call_desc_dealloc(new_desc);
        block_cli_if_recover_data(desc);
}

static inline struct desc_track *call_desc_update(ul_t id, int next_state)
{
        struct desc_track *desc = NULL;
        unsigned int from_state = 0;
        unsigned int to_state = 0;

        if (id == 0)
                return NULL;

        desc = call_desc_lookup(id);
        if (unlikely(!desc))
                goto done;

        desc->next_state = next_state;

        if (likely(desc->fault_cnt == global_fault_cnt))
                goto done;
        desc->fault_cnt = global_fault_cnt;

done:
        return desc;
}

CSTUB_FN(int, lock_component_take)(struct usr_inv_cap * uc, spdid_t spdid,
                                   ul_t lock_id, u32_t thd_id) {
        long fault = 0;
        int ret = 0;

        call_map_init();

redo:
        block_cli_if_desc_update_lock_component_take(spdid, lock_id, thd_id);

        ret =
            block_cli_if_invoke_lock_component_take(spdid, lock_id, thd_id, ret,
                                                    &fault, uc);
        if (unlikely(fault)) {

                CSTUB_FAULT_UPDATE();
                if (block_cli_if_desc_update_post_fault_lock_component_take()) {
                        goto redo;
                }
```



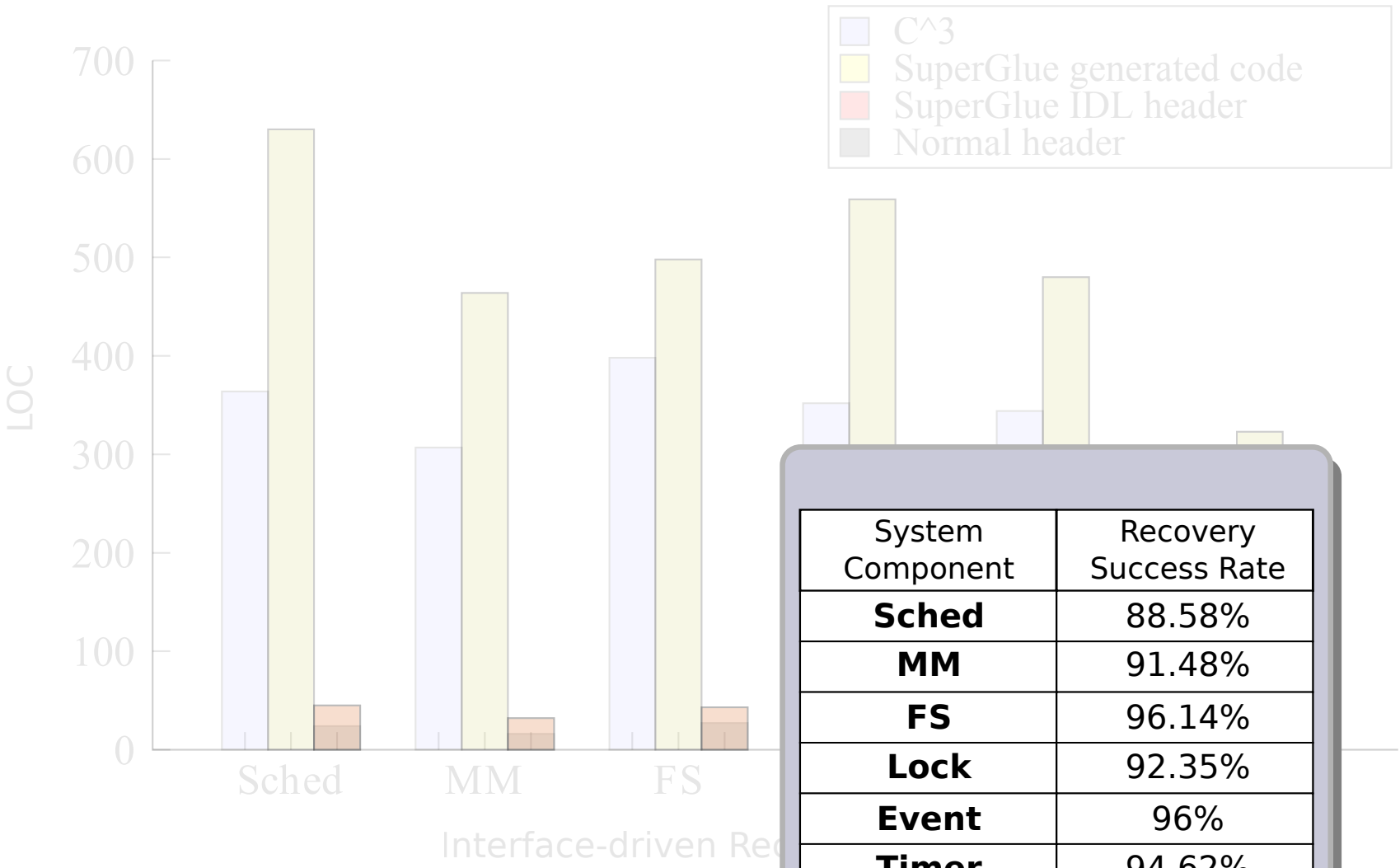
Generate Recovery Code



Code Generation Result



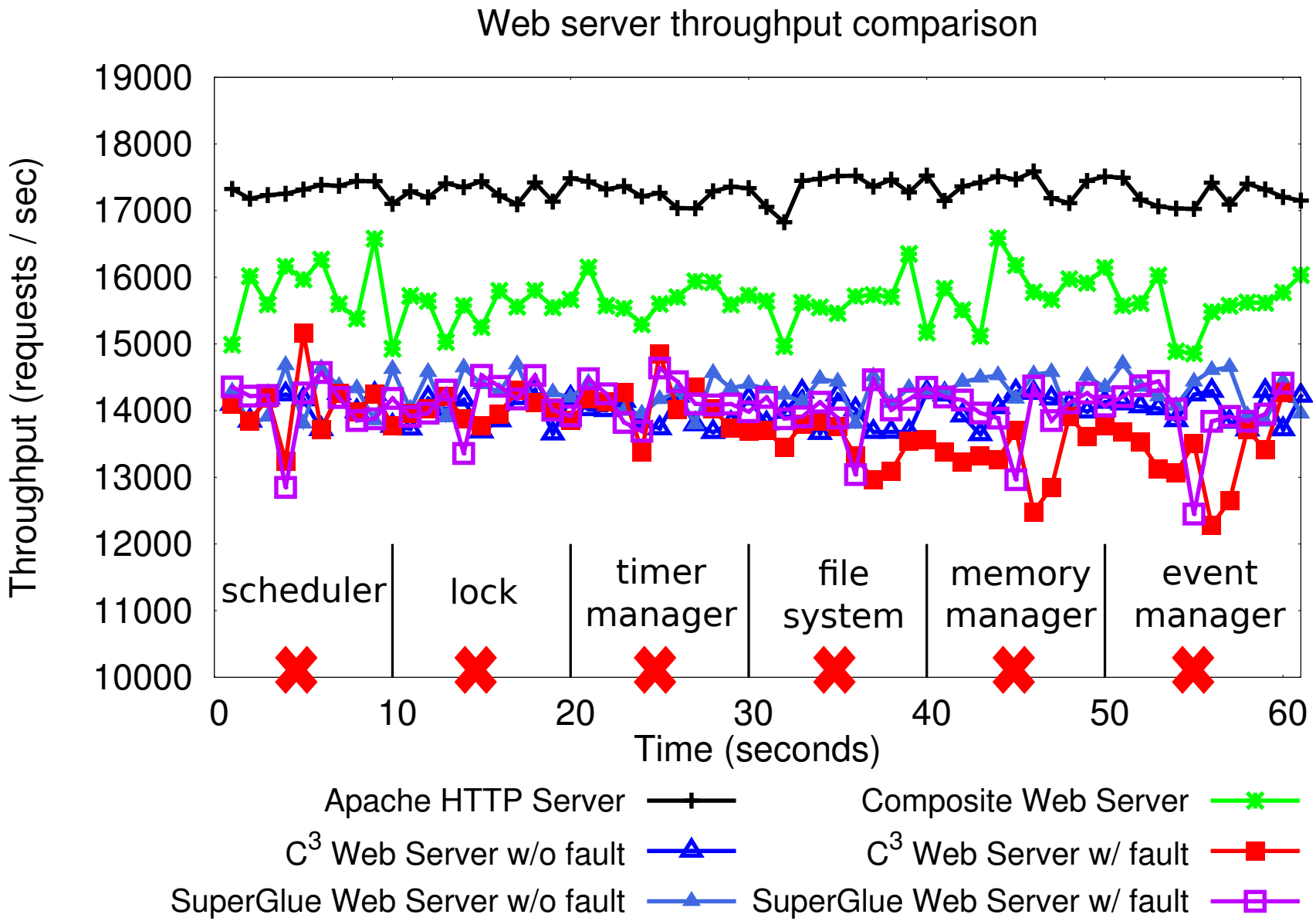
Fault Injection Result



System Component	Recovery Success Rate
Sched	88.58%
MM	91.48%
FS	96.14%
Lock	92.35%
Event	96%
Timer	94.62%

SWIFI-based Fault Injection Campaign

Web Server Evaluation with Injected Faults



Conclusion on SuperGlue

SuperGlue – code generation for system-level fault tolerance

- descriptor-resource model and descriptor state machine
- IDL-based declarative specifications
- compiler for synthesizing C^3-style recovery code

Conclusion

Thanks

? || /* */

composite.seas.gwu.edu