



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

Department of Computer Science Institute of System Architecture, Operating Systems Group

REAL-TIME

MICHAEL ROITZSCH

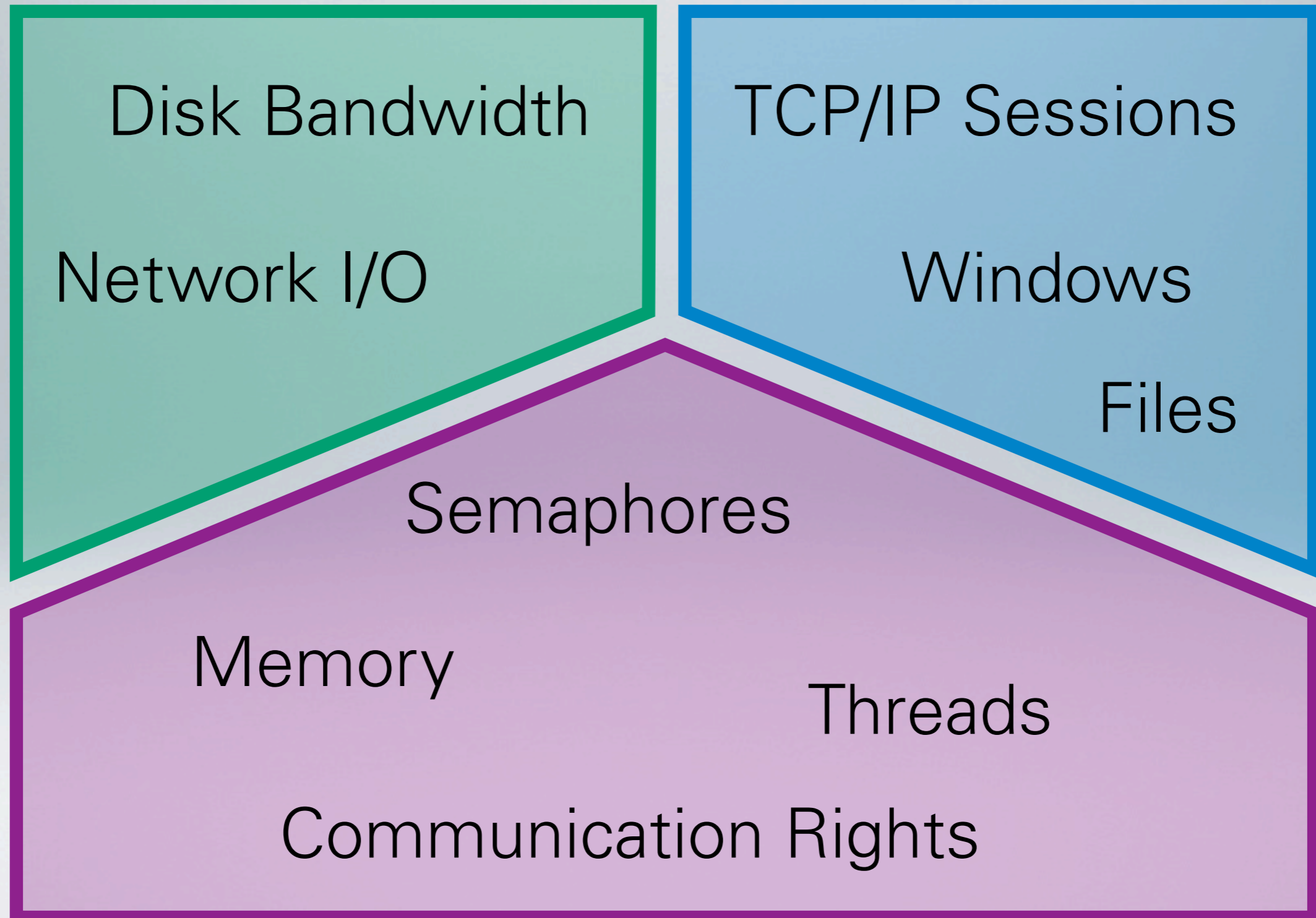
OVERVIEW

- talked about in-kernel building blocks:
 - threads
 - memory
 - IPC
- drivers will enable access to a wide range of non-kernel resources
- need to manage resources

Applications

System Services

Basic Abstractions



Memory

- discrete, limited
- hidden in the system
- managed by pager
- page-granular partitions
- all pages are of equal value
- active policy decisions,
passive enforcement
- hierarchical management

Time

- continuous, infinite
- user-perceivable
- managed by scheduler
- arbitrary granularity
- value depends on workload
- active policy decisions,
active enforcement
- Fiasco: flattened in-kernel
view

REAL-TIME

- a **real-time system** denotes a system, whose correctness depends on the timely delivery of results
- „it matters, **when** a result is produced“
- **real-time** denotes a predictable relation between system progress and wall-clock time

- real-time **is** about
 - predictability
 - guarantees
 - timeliness
 - responsiveness
- real-time **is not** about
 - being fast
 - live calculations

- engine control in a car
- break-by-wire
- avionics
- railway control

**focused
catastrophic failures**

- set-top box DVD player
- GSM-stack in your cell phone

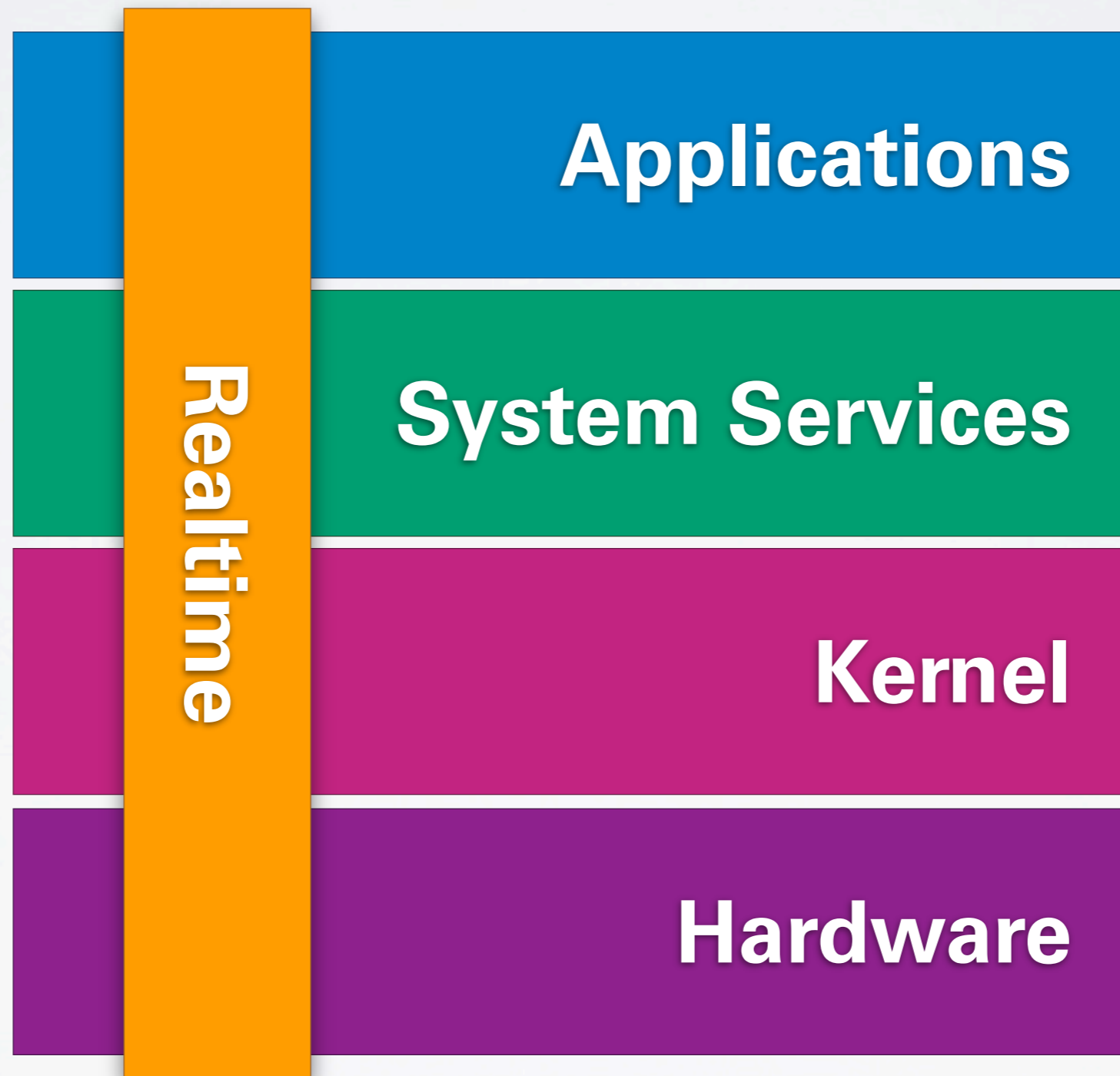
**benign failures
complex**

	hard real-time	firm real-time	soft real-time
missing some deadlines is tolerable	✘	✔	✔
a result delivered after its deadline is still useful	✘	✘	✔

- ① Predictability
- ② Guarantees
- ③ Enforcement

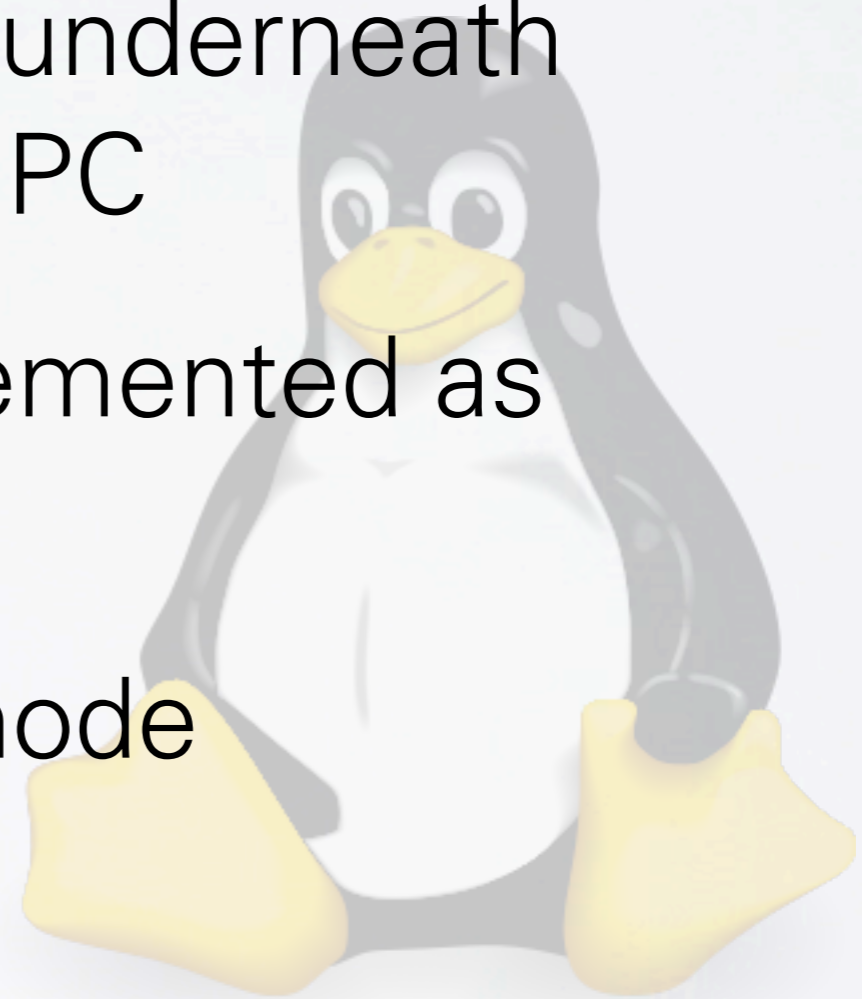
PREDICTABILITY

- gap between worst and average case
 - memory caches, disk caches, TLBs
- „smart“ hardware
 - system management mode
 - disk request reordering
- cross-talk from resource sharing
 - servers showing $O(n)$ behavior
 - SMP
- unpredictable external influences
 - interrupts

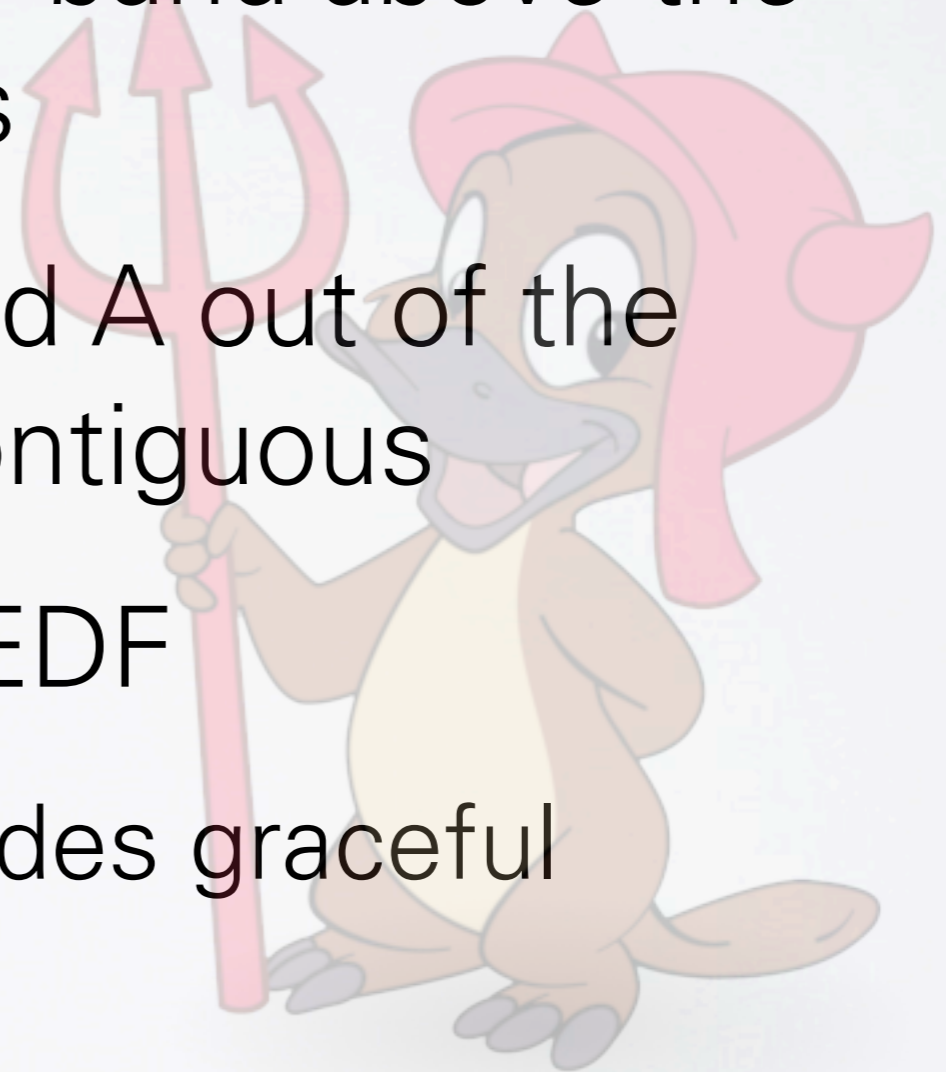


- small real-time executives tailor-made for specific applications
- fixed workload known a priori
- pre-calculated time-driven schedule
- used on small embedded controllers
- benign hardware

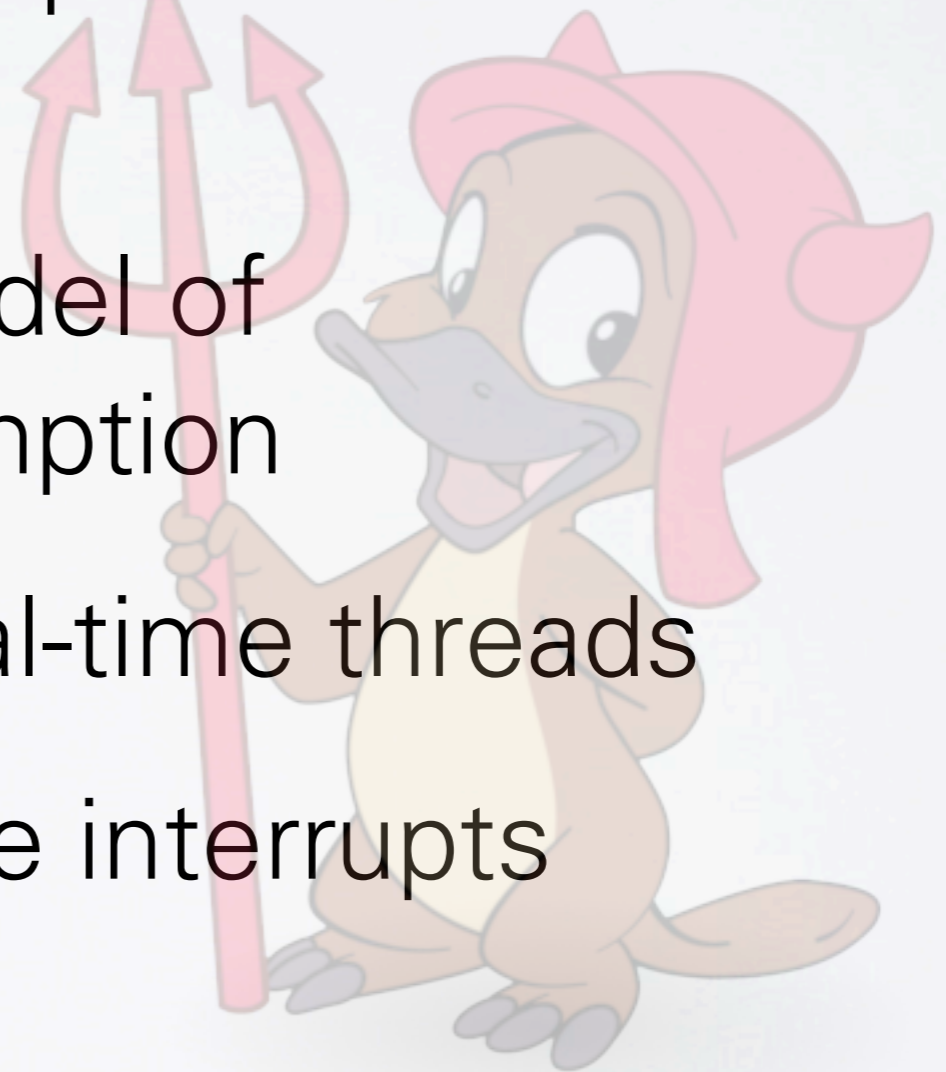
- full Linux kernel and real-time processes run side-by-side
- small real-time executive underneath supports scheduling and IPC
- real-time processes implemented as kernel modules
- all of this runs in kernel mode
- no isolation



- the kernel used in Mac OS X
- offers a real-time priority band above the priority of kernel threads
- you tell the kernel: I need A out of the next B cycles, can be contiguous
- scheduled using global EDF
 - bounded tardiness provides graceful overload behavior

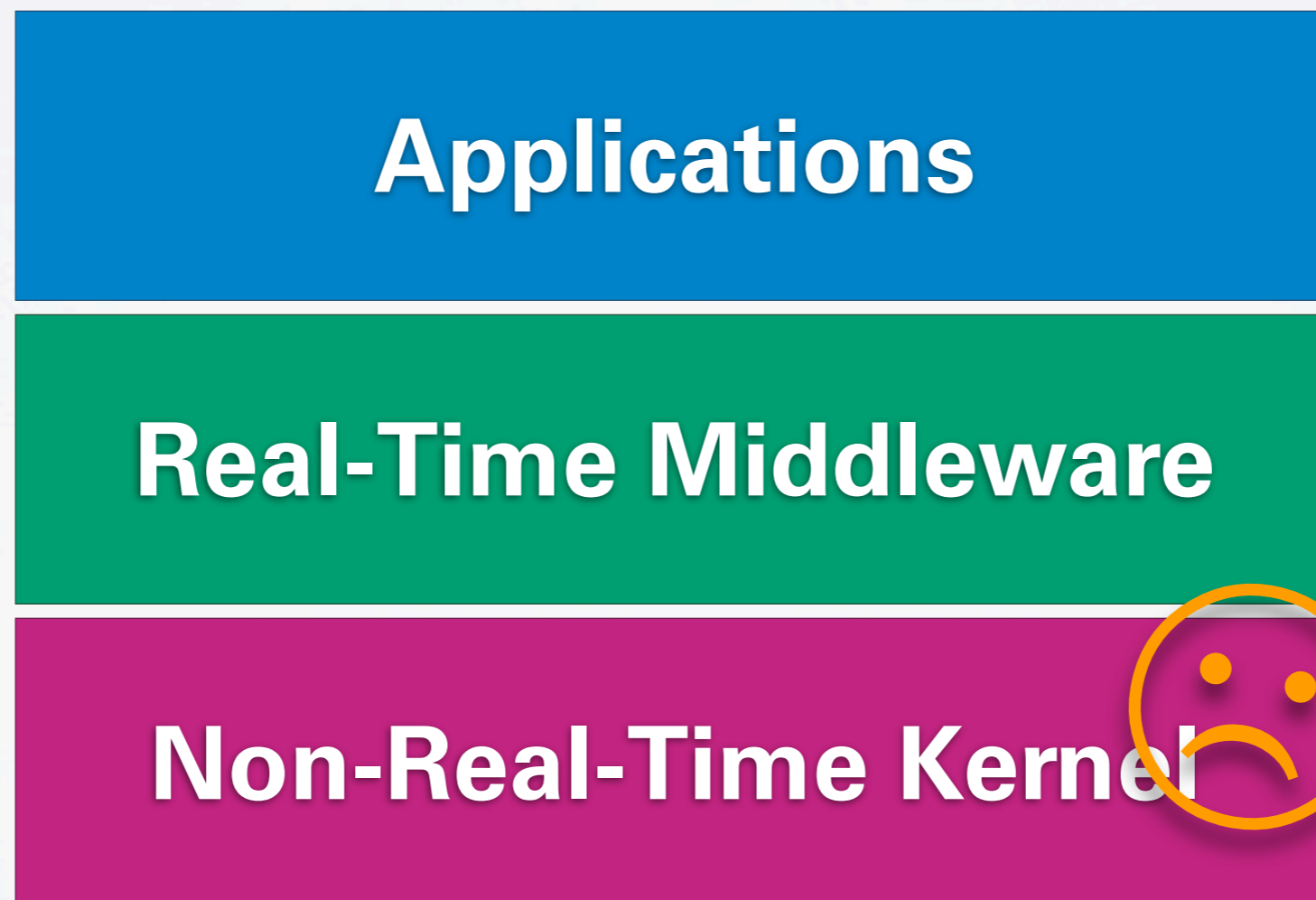


- normal users can create real-time threads
- threads exceeding their quantum will be demoted
- kernel uses a hybrid model of continuations and preemption
- full preemptibility for real-time threads
- all drivers need to handle interrupts correctly



- static thread priorities
- bounded interrupt latency
- fully preemptible in kernel mode
- lock-free synchronization
 - uses atomic operations
- wait-free synchronization
 - locking with helping instead of blocking

- architecture for those afraid of touching the OS
- example: Real-Time Java

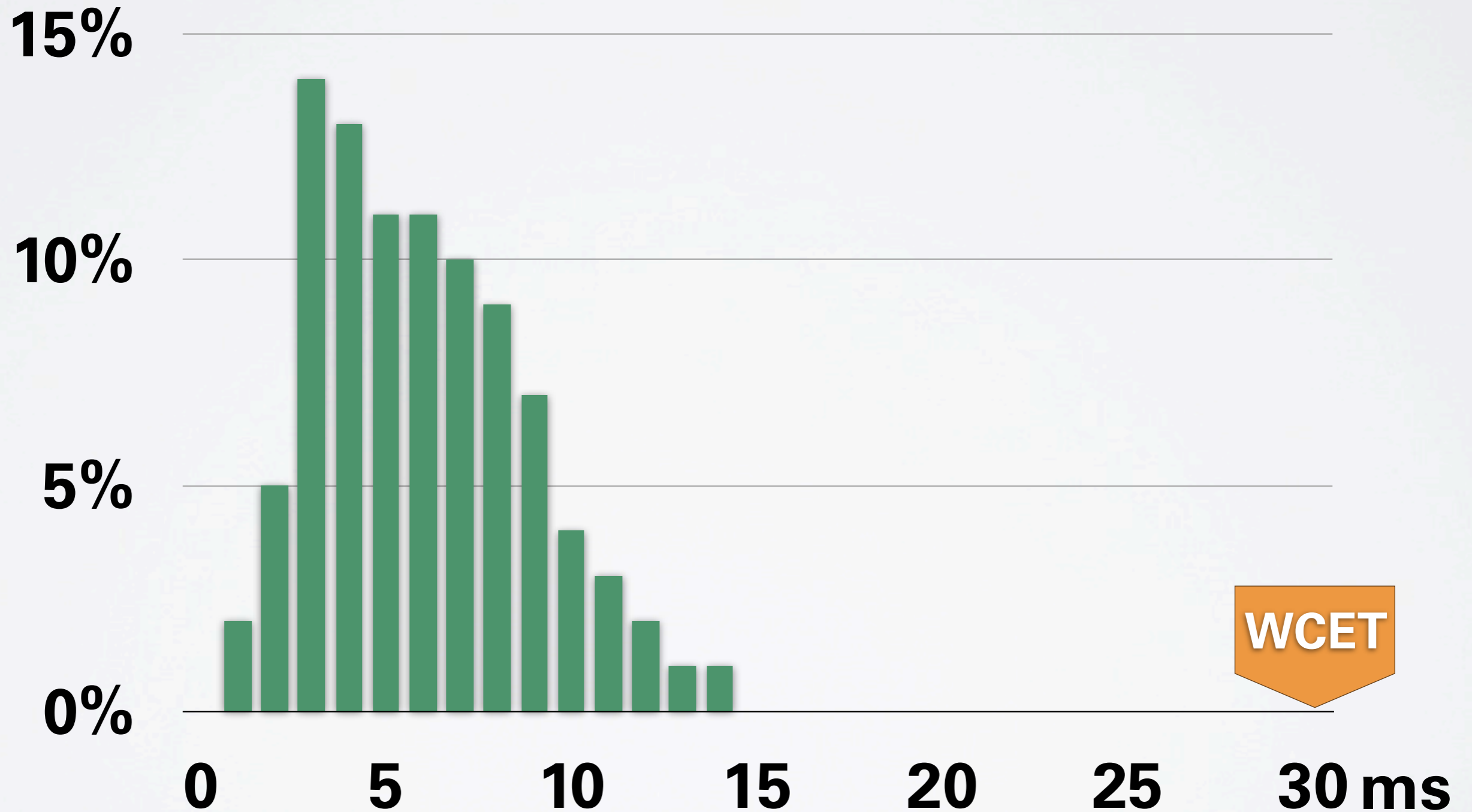


- a real-time kernel alone is not enough
- the microkernel solution:
 - real-time kernel enables temporal isolation
 - eliminates cross-talk and interrupt problems
 - user-level servers on top act as resource managers
 - implement real-time views on specific resources
- real-time is not only about CPU
- details in the resource management lecture

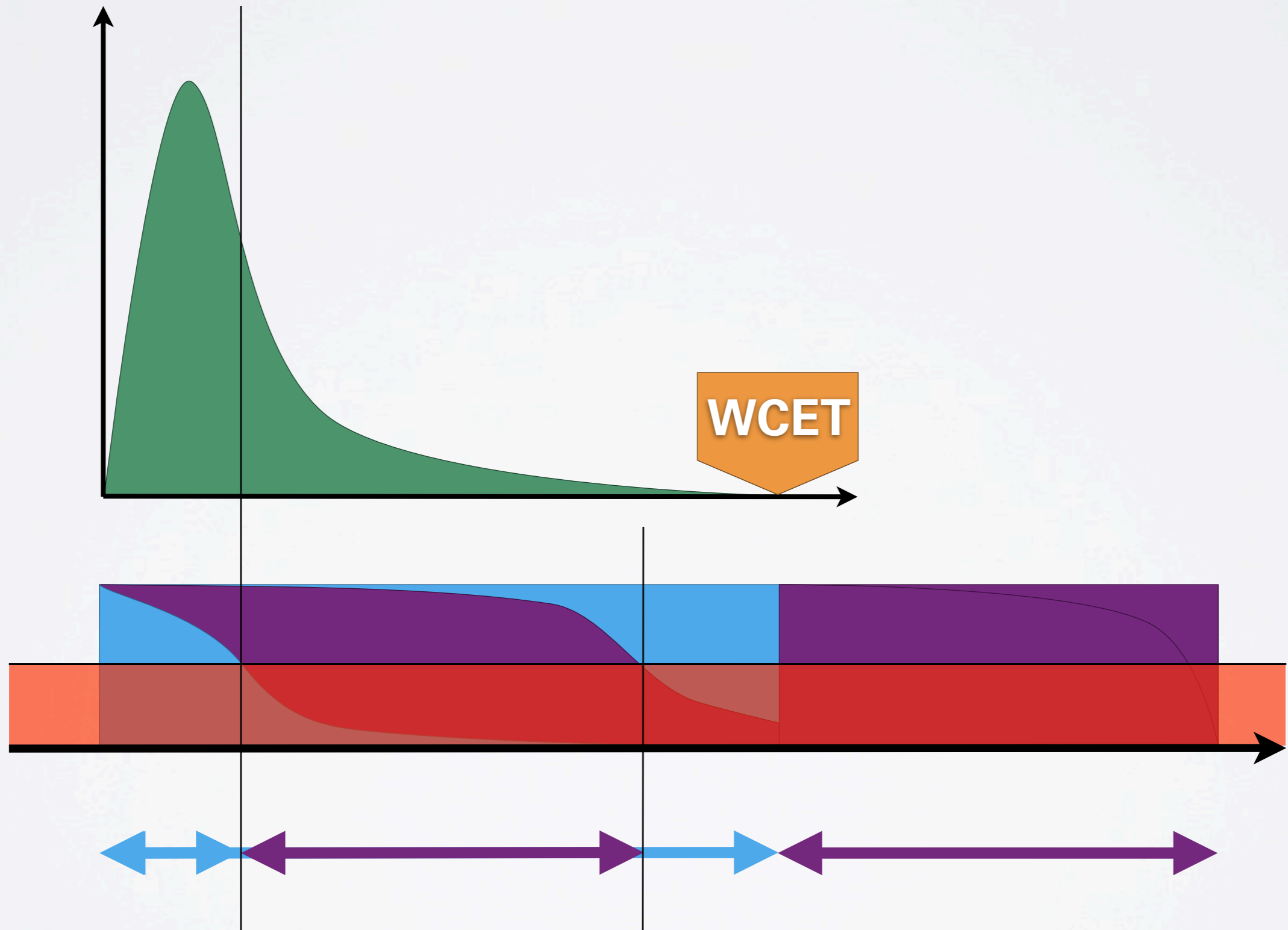
GUARANTEES

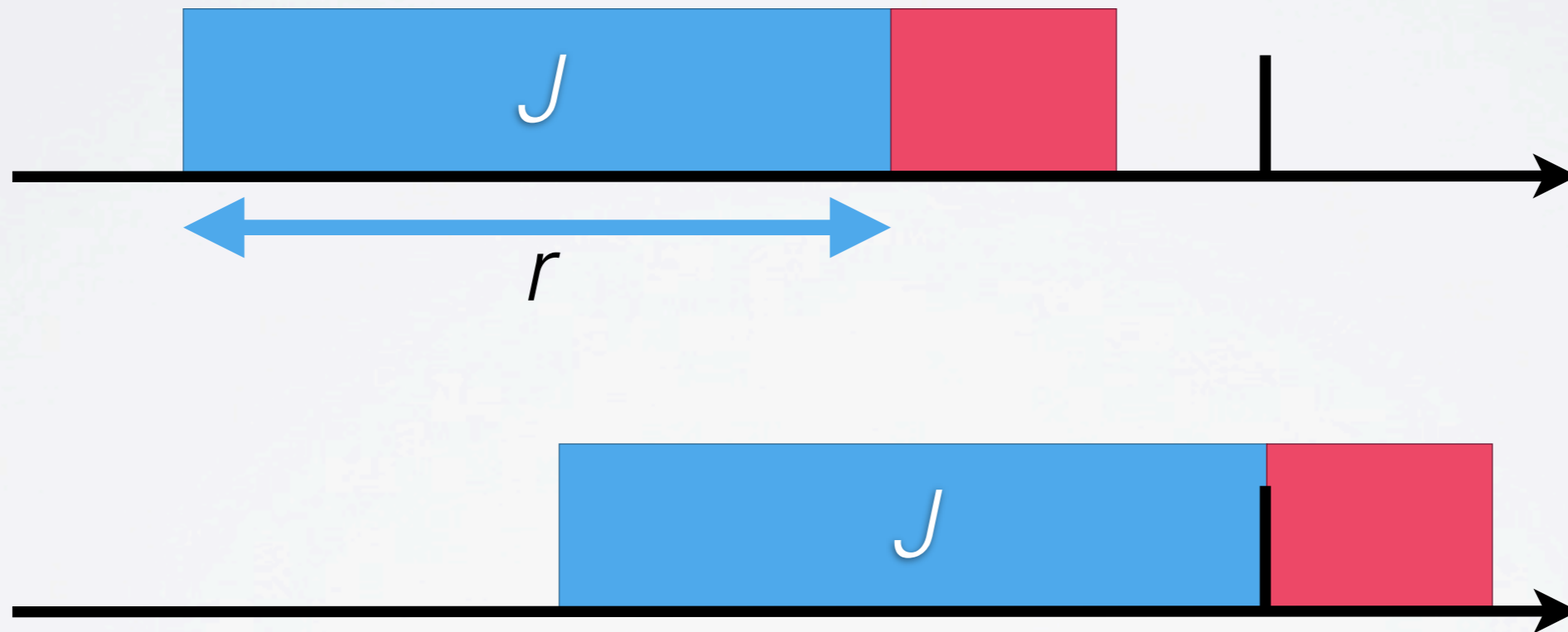
- worst case execution time (WCET) largely exceeds average case
- offering guarantees for the worst case will waste lots of resources
- missing some deadlines can be tolerated with the firm and soft real-time scheme

- desktop real-time
- there are no hard real-time applications on desktops
- there is a lot of firm and soft real-time
 - low-latency audio processing
 - smooth video playback
 - desktop effects
 - user interface responsiveness



- guarantees even slightly below 100% of WCET can dramatically reduce resource allocation
- unused reservations will be used by others at runtime
- use probabilistic planning to model the actual execution
- quality q : fraction of deadlines to be met





$\mathbf{P}(J \text{ does not run longer than } r \wedge$
 $J \text{ is completed until its relative deadline}) \geq q$

$$r'_i = \min(r \in \mathbb{R} \mid \frac{1}{m_i} \sum_{k=1}^{m_i} \mathbf{P}(X_i + k \cdot Y_i \leq r) \geq q_i)$$

$$r_i = \max(r'_i, w_i) \quad i = 1, \dots, n$$

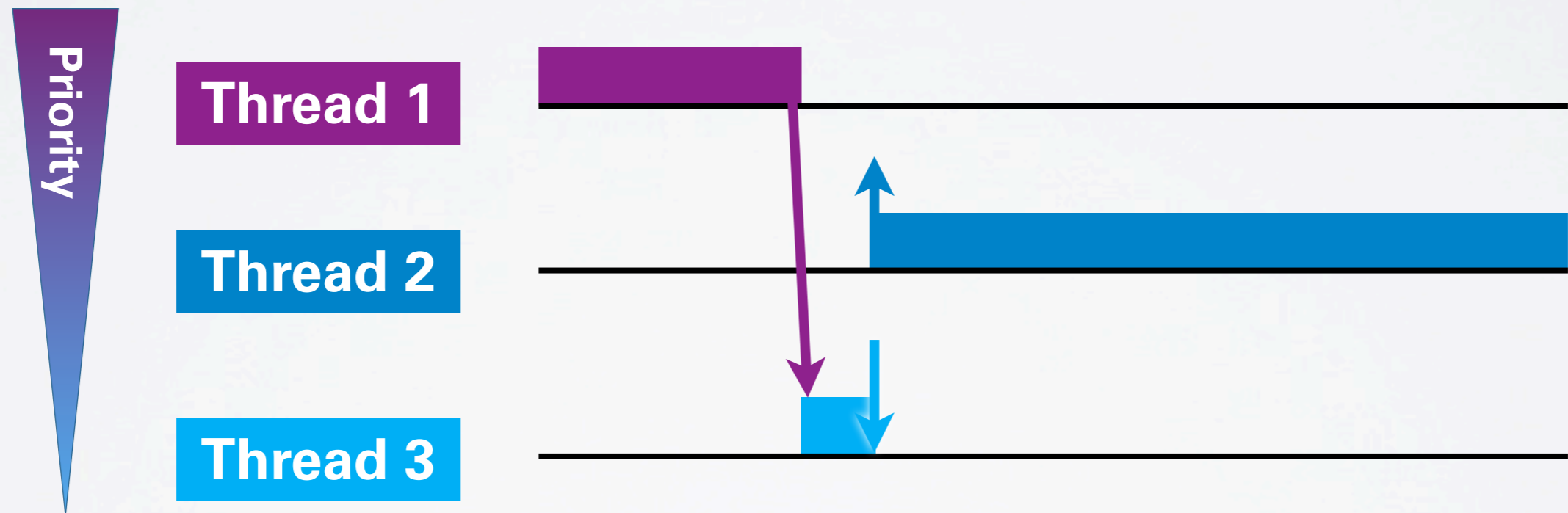
- to fully understand this:
see real-time systems lecture
- good for microkernel: reservation can be
calculated by a userland service
- kernel only needs to support static priorities

- **scheduling = admission + dispatch**
- admission
 - evaluates the requests from clients, which follow some task model
 - calculates task parameters for dispatcher
 - verifies the feasibility of the guarantees
 - can reject requests
- dispatch
 - executes and enforces the schedule

ENFORCEMENT

- executes schedule
- enforces task parameters by preemption
 - e.g. on deadline overrun
- picks the next thread
 - static priorities (e.g. RMS, DMS)
 - dynamic priorities (e.g. EDF)
- seems simple...

- high priority thread calls low priority service with a medium priority thread interfering:



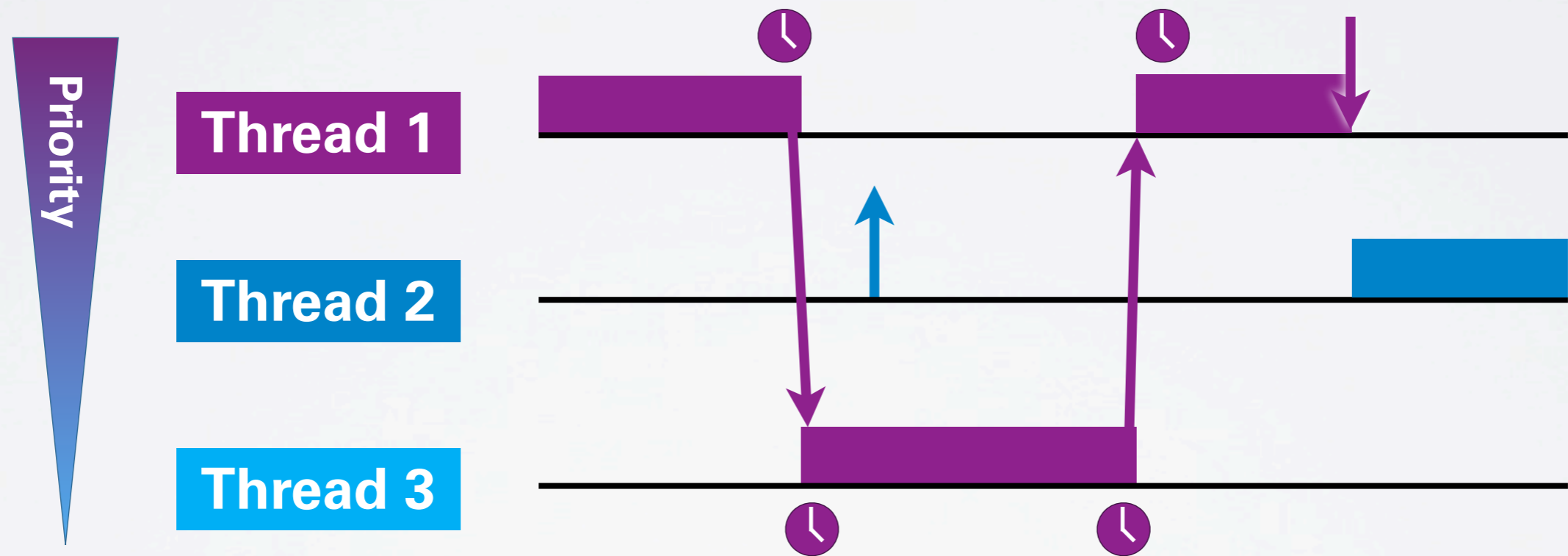
1 waits for 3 ✓

3 waits for 2 ✓

= 1 waits for 2 ✗

Priority Inversion

- priority inheritance, priority ceiling
- nice mechanism for this in Fiasco, NOVA: timeslice donation
- implemented by splitting thread control block
 - **execution context:** holds CPU state
 - **scheduling context:** time and priority
- on IPC-caused thread switch, only the execution context is switched



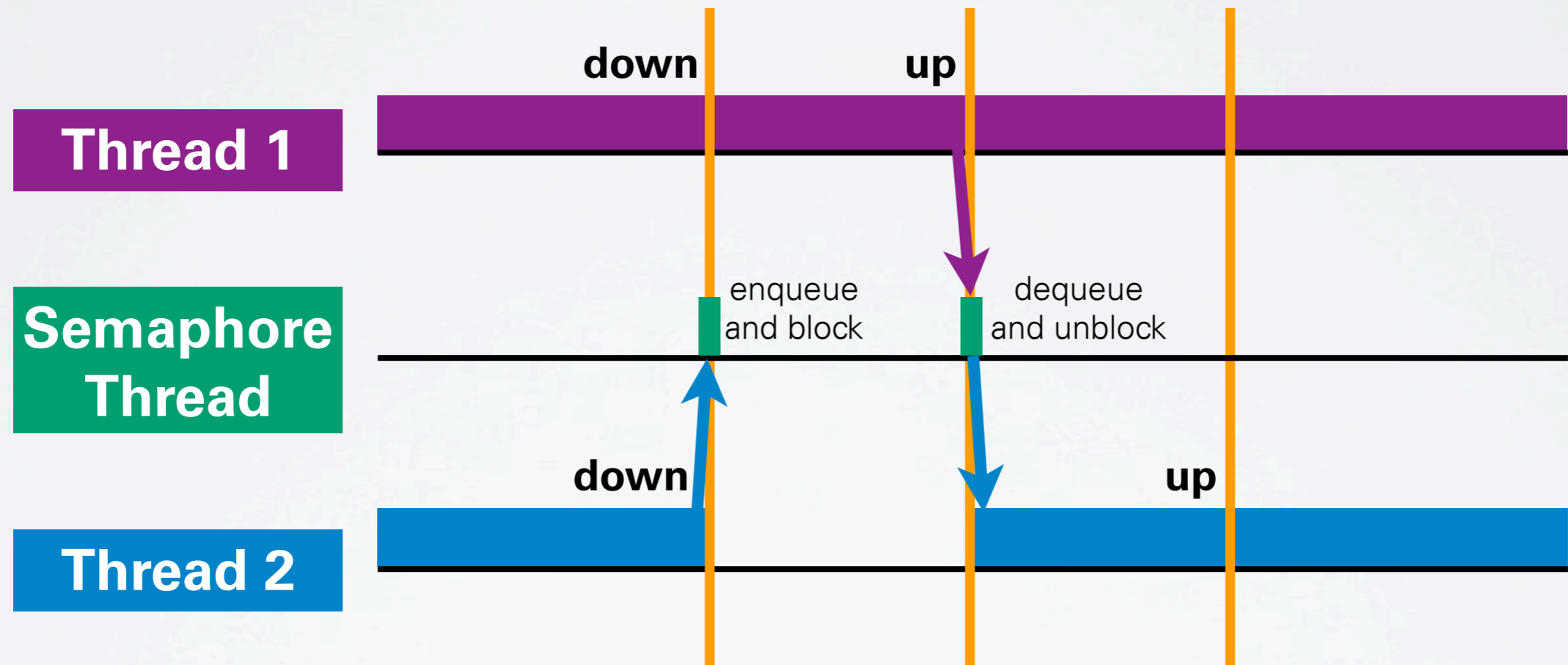
- IPC receiver runs on the sender's scheduling context
- priority inversion problem solved with priority inheritance

- servers run on their clients time slice
 - when the server executes on behalf of a client, the client pays with its own time
- this allows for servers with no scheduling context
 - server has no time or priority on its own
 - can only execute on client's time
 - relieves dispatcher from dealing with servers

- servers could be malicious, so you need timeouts to get your time back
- now, malicious clients can call the server with a very short timeout
 - server's time will be withdrawn while the server is in the middle of request handling
- servers need to do session cleanup
 - on whose time?
 - open research problem

- timeslice donation does not work in multiprocessor case:
 - server is running on a different CPU
 - cross-CPU donation would be needed
- you cannot donate time between CPUs
 - think of real-time guarantees
 - if one CPU is 100% loaded, donating time to it will overload it
- servers need to migrate to client CPU?

OPTIMIZATION



- IPC only in the contention case
- optimized for low contention
- bad for producer-consumer (high contention)

- consumer
 - request wakeup: set flag, enqueue
 - block
- producer
 - check for wakeup: test flag, dequeue
 - send wakeup
- very short critical sections
- semaphores too expensive (2 IPCs)

- allow threads to have short periods where they are never preempted
 - like a very low cost global system lock
 - like a userland-way of disabling interrupts
- **delayed preemption**
- threads can set „don't preempt me“ flag in UTCB
 - very low cost

- unbounded delay
 - if preemption would occur, the kernel honors the delayed preemption flag only for a **fixed maximum delay**
- delay affects all threads
 - problematic for real-time guarantees
 - affected threads can be limited to those within a **priority band**
 - other threads can still preempt any time

- system must be designed carefully
- threads outside the affected priority band must not interfere with protected data structures
- more problems
 - unaffected threads might donate time to a thread within the priority band
 - does not work cross-CPU

- managing time is necessary
 - we interact with the system based on time
- real-time is a cross-cutting concern
- „hard real-time is hard, soft real-time is even harder“ (E. Douglas Jensen)
- priority inheritance by timeslice donation
- delayed preemption
- next week: drivers