TECHNISCHE UNIVERSITÄT DRESDEN

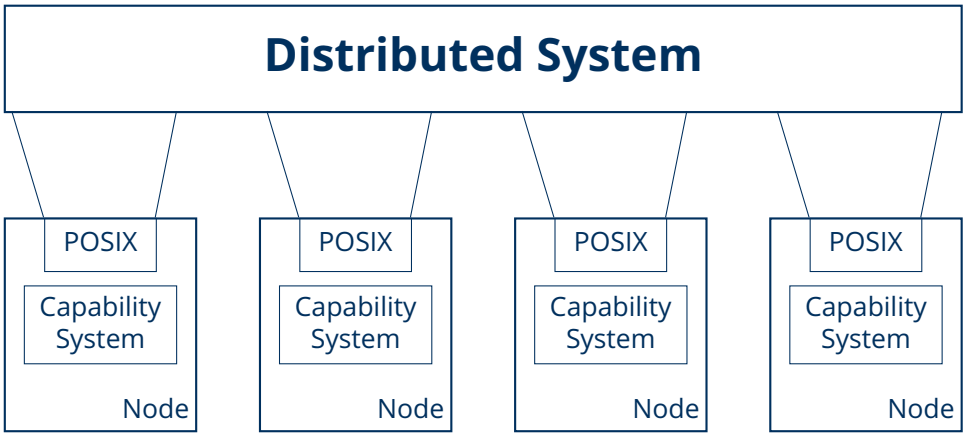
SemperOS: A Distributed Capability System

Matthias Hille[†] Nils Asmussen[†] [*] Pramod Bhatotia[‡] Hermann Härtig[†] [*]

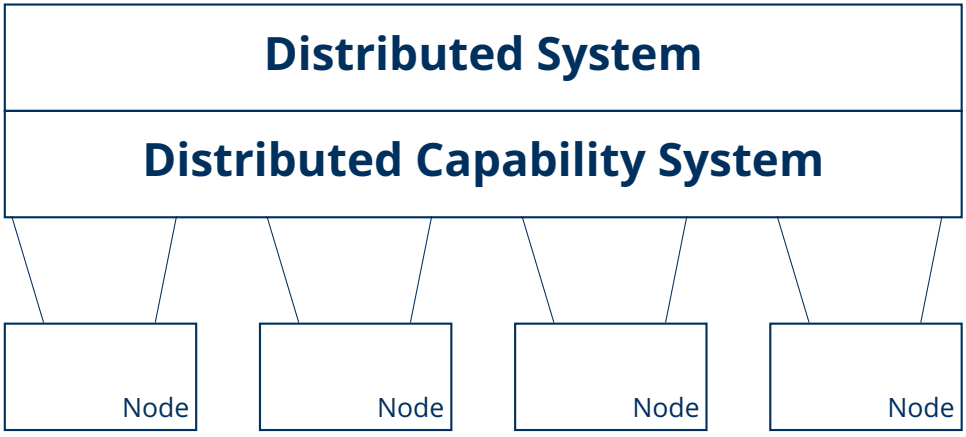
[†]Technische Universität Dresden [‡]The University of Edinburgh [*] Barkhausen Institut

USENIX ATC '19
July 11, 2019
Renton, WA, USA

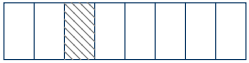
Motivation


Distributed System
POSIX
Capability System
Node
POSIX
Capability System
Node
POSIX
Capability System
Node
POSIX
Capability System
Node

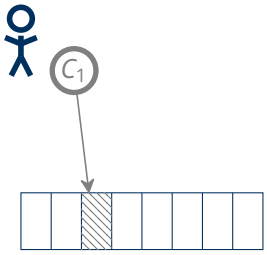
Motivation


Distributed System
Distributed Capability System
Node
Node
Node
Node

Capability Systems

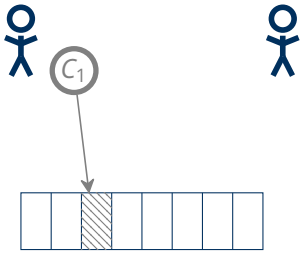


Capability Systems



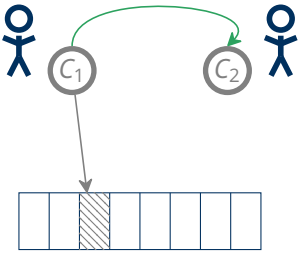
Capability Tree

Capability Systems

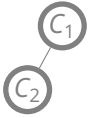


Capability Tree

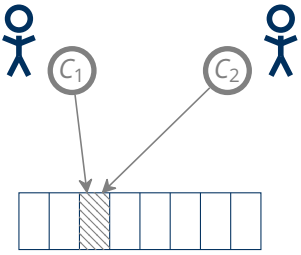
Capability Systems



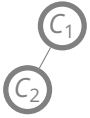
Capability Tree



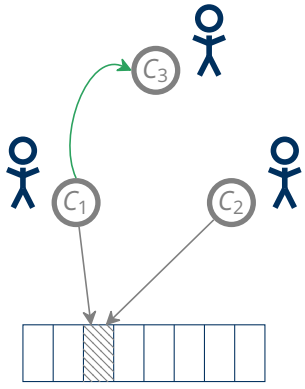
Capability Systems



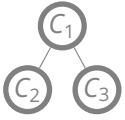
Capability Tree



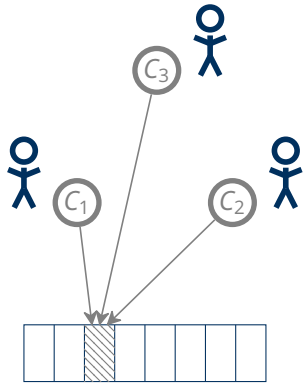
Capability Systems



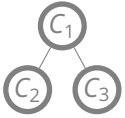
Capability Tree



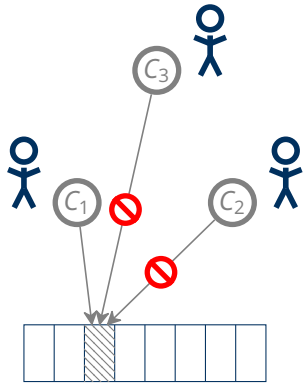
Capability Systems



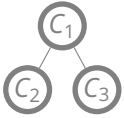
Capability Tree



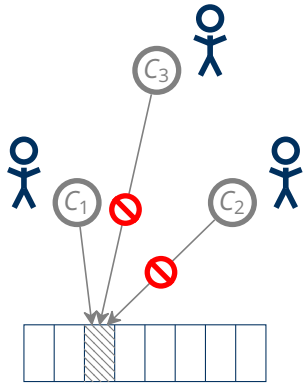
Capability Systems



Capability Tree

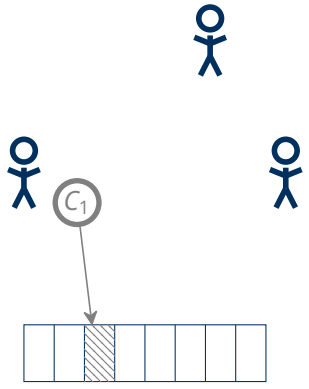


Capability Systems



Capability Tree

Capability Systems



Capability Tree

Characteristics of Capability Systems

Scope			
Enforcement			
Scalability			

Characteristics of Capability Systems

	L4		
Scope	Coherence Domain		
Enforcement	MMU / Kernel		
Scalability	Limited by Coherence Domain		

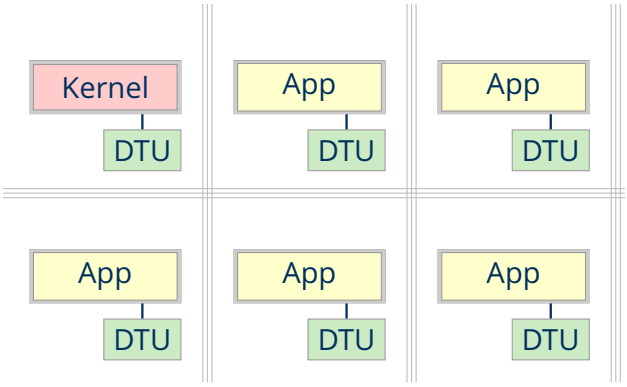
Characteristics of Capability Systems

	L4	M^3	
Scope	Coherence Domain	Machine	
Enforcement	MMU / Kernel	DTU / Kernel	
Scalability	Limited by Coherence Domain	Limited by Single Kernel	

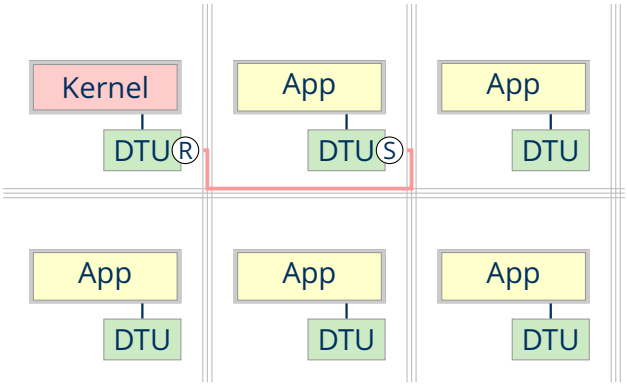
Characteristics of Capability Systems

	L4	M^3	Barrelfish
Scope	Coherence Domain	Machine	Machine
Enforcement	MMU / Kernel	DTU / Kernel	MMU / Kernel
Scalability	Limited by Coherence Domain	Limited by Single Kernel	Shown up to 32 Cores

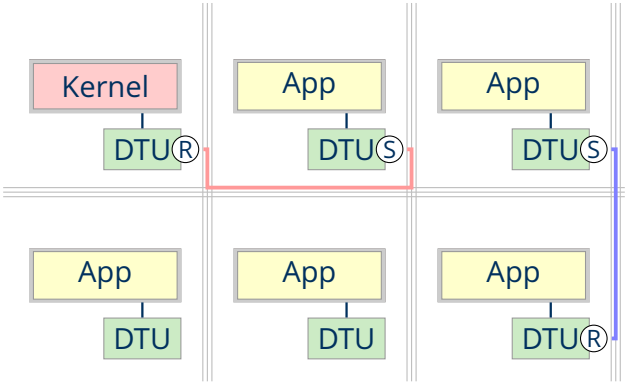
Base Capability System – M³



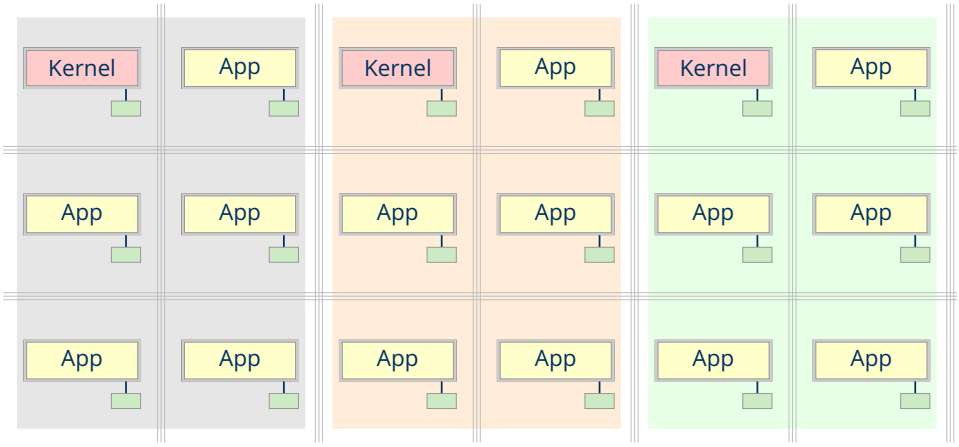
Base Capability System – M³



Base Capability System – M³

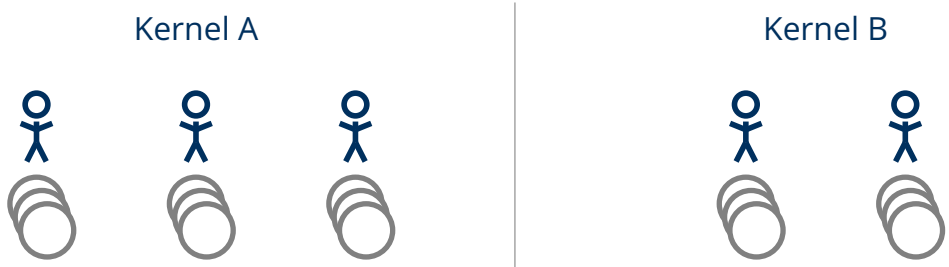


SemperOS– A Distributed Capability System

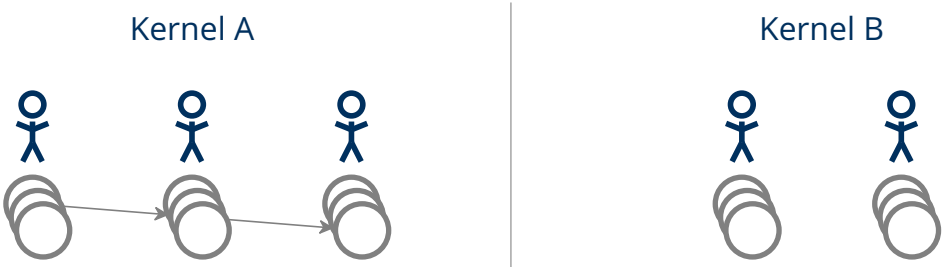


The Distributed Data Lookup – DDL

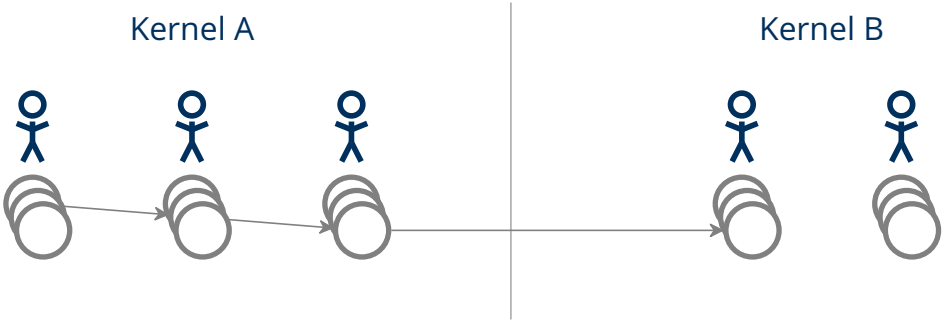
The Distributed Data Lookup – DDL



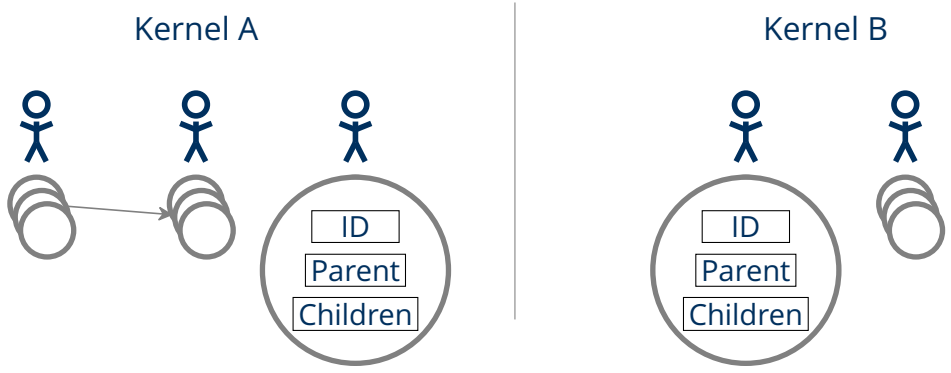
The Distributed Data Lookup – DDL



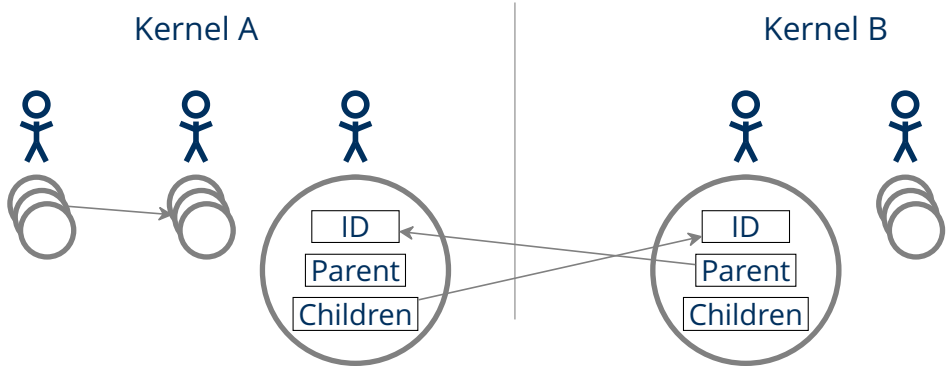
The Distributed Data Lookup – DDL



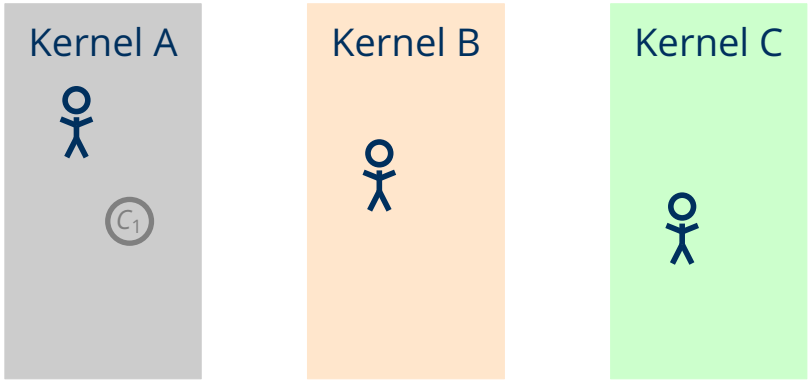
The Distributed Data Lookup – DDL



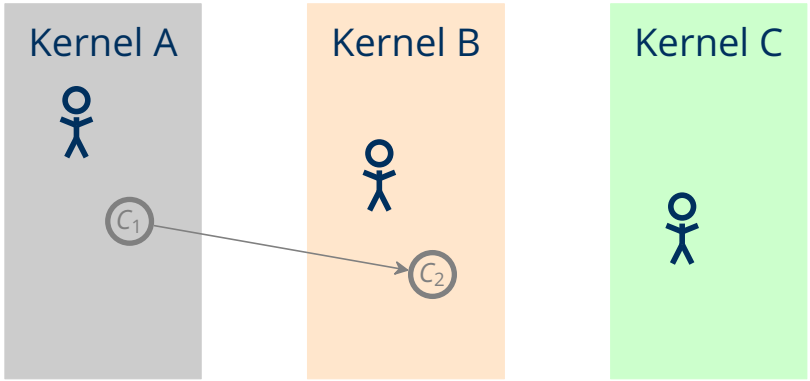
The Distributed Data Lookup – DDL



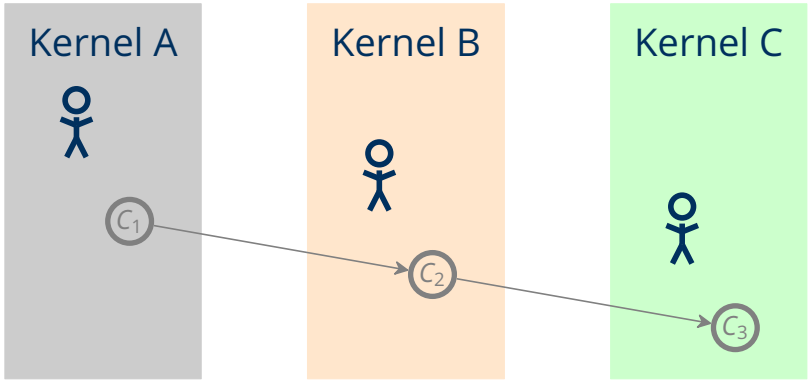
Distributed Capabilities – Revocation



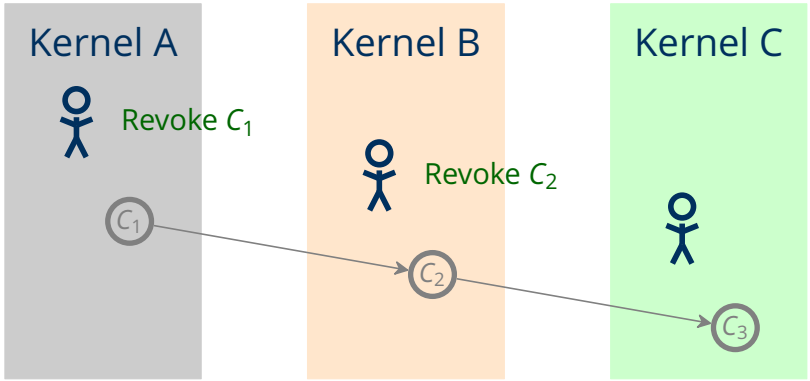
Distributed Capabilities – Revocation



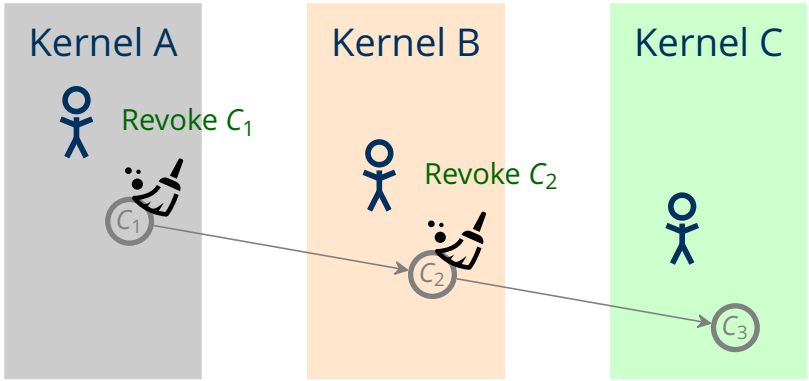
Distributed Capabilities – Revocation



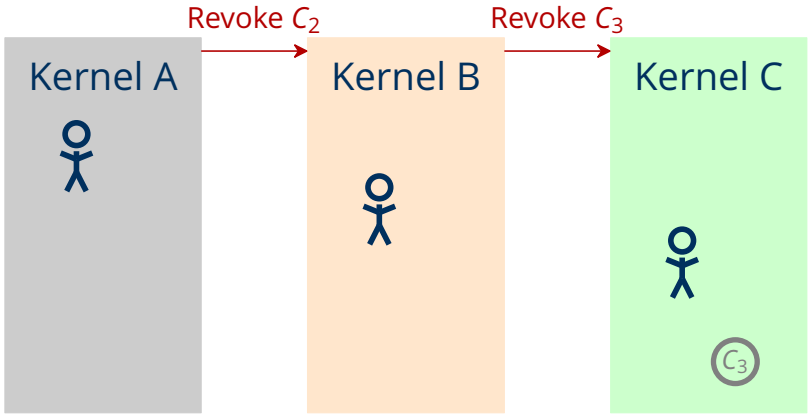
Distributed Capabilities – Revocation



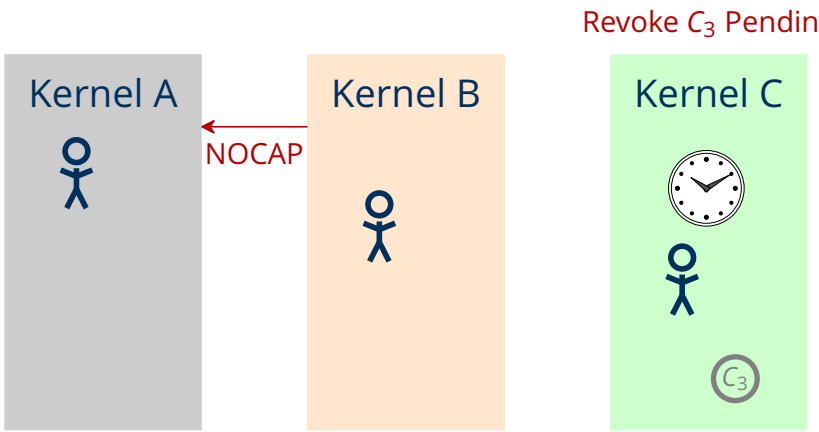
Distributed Capabilities – Revocation



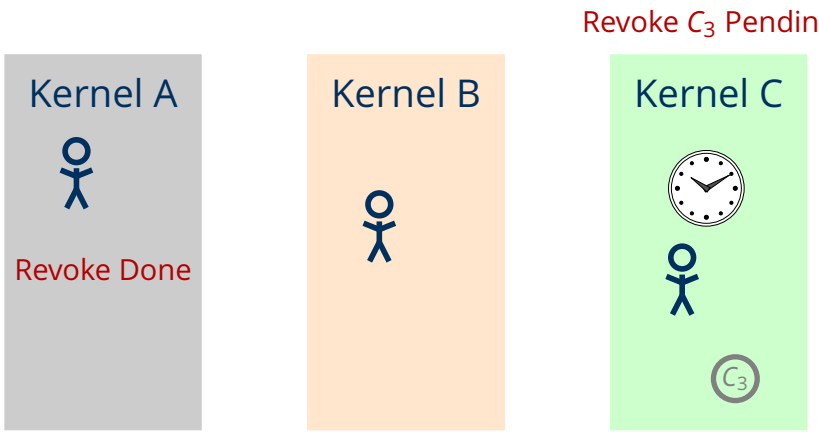
Distributed Capabilities – Revocation



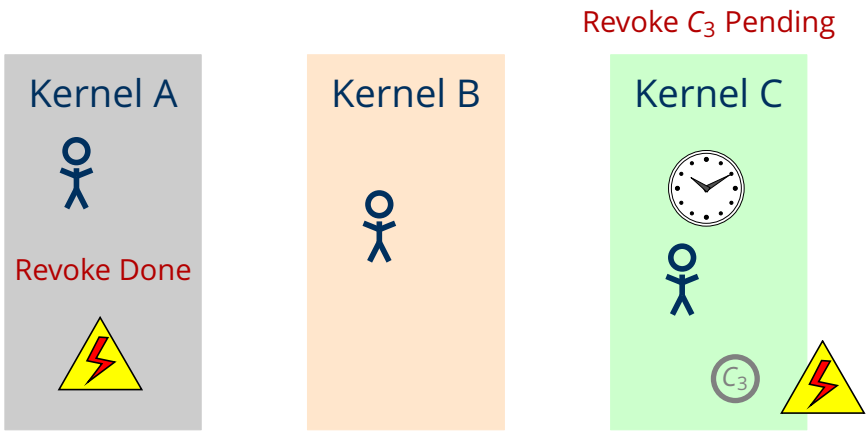
Distributed Capabilities – Revocation



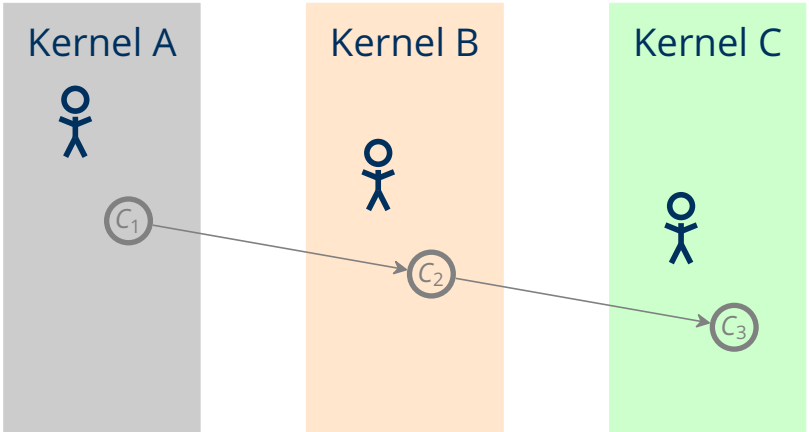
Distributed Capabilities – Revocation



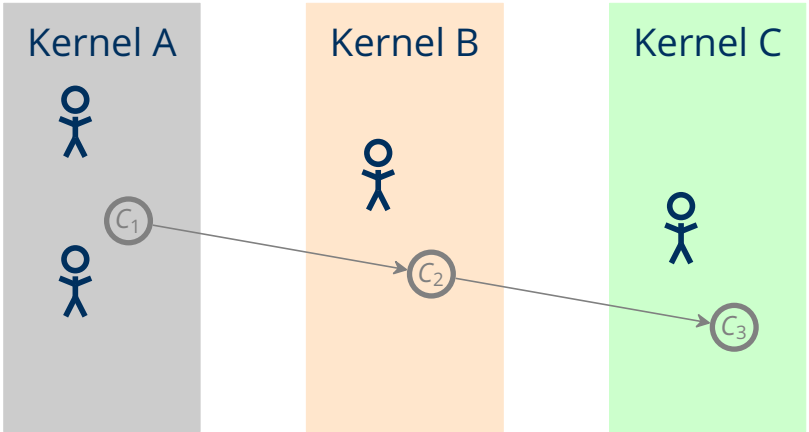
Distributed Capabilities – Revocation



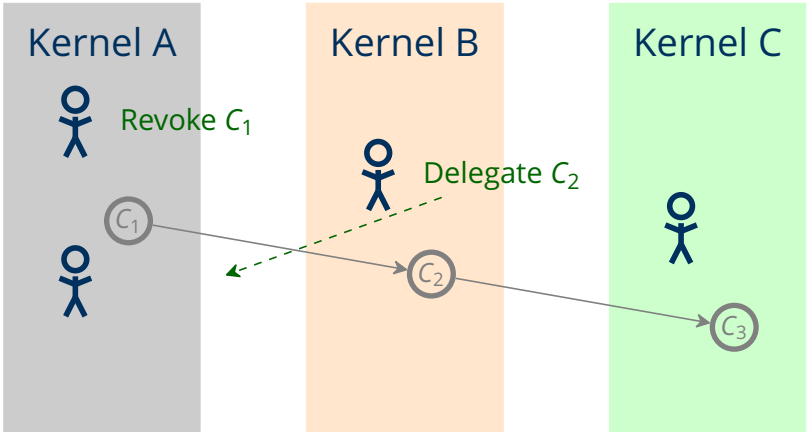
Distributed Capabilities – Delegation



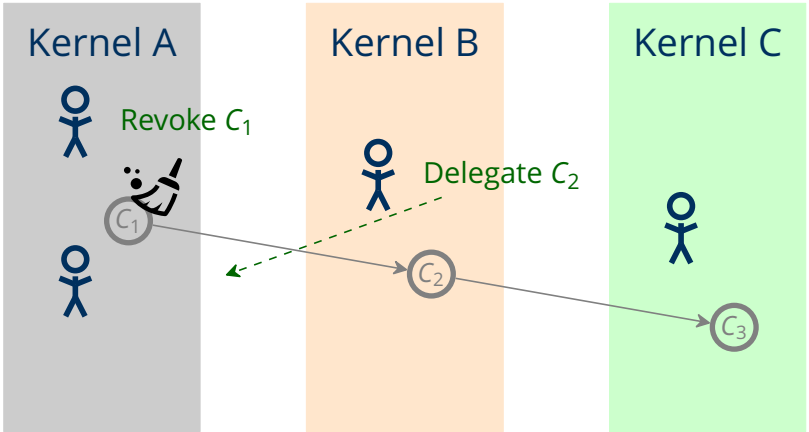
Distributed Capabilities – Delegation



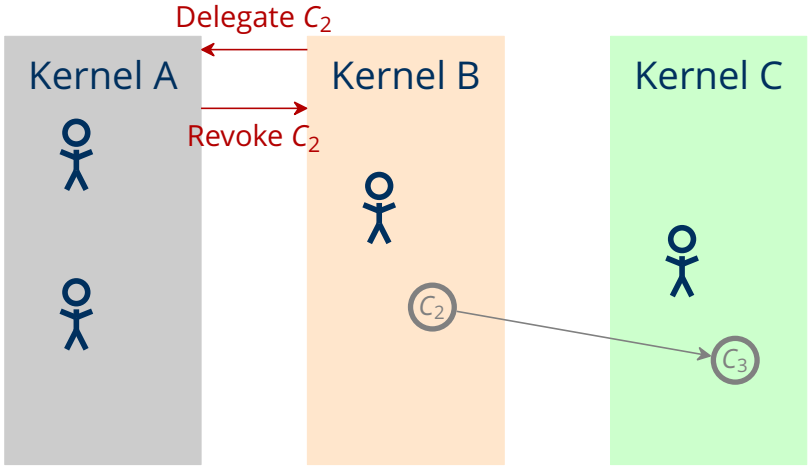
Distributed Capabilities – Delegation



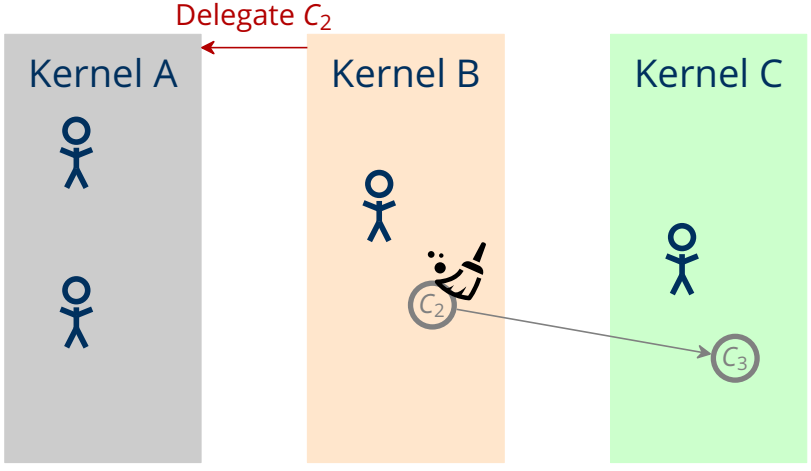
Distributed Capabilities – Delegation



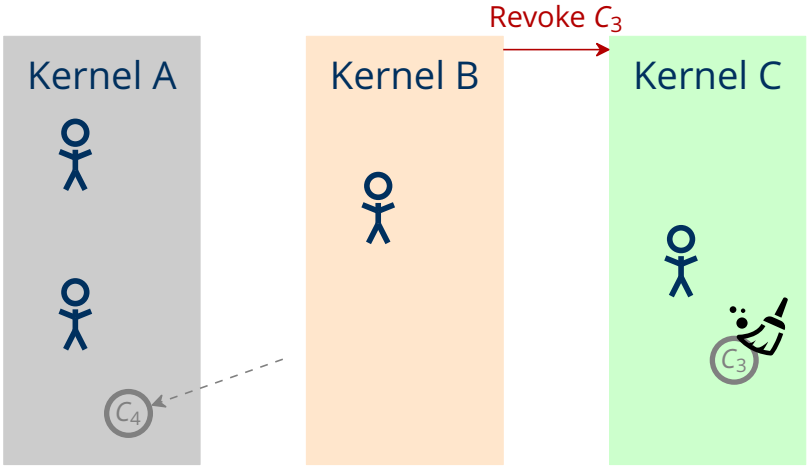
Distributed Capabilities – Delegation



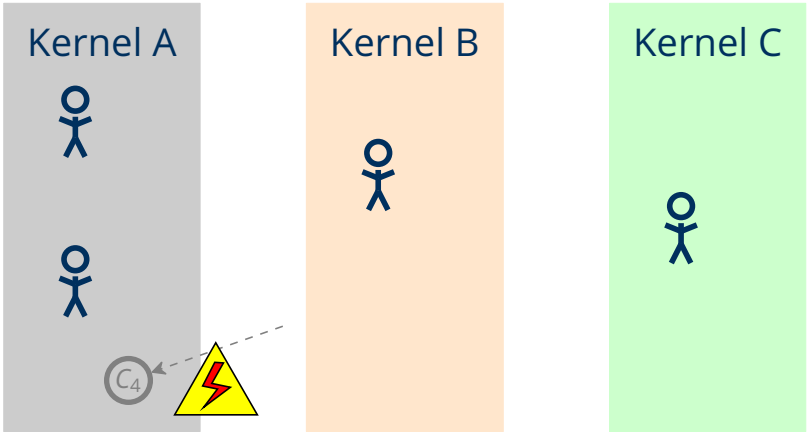
Distributed Capabilities – Delegation



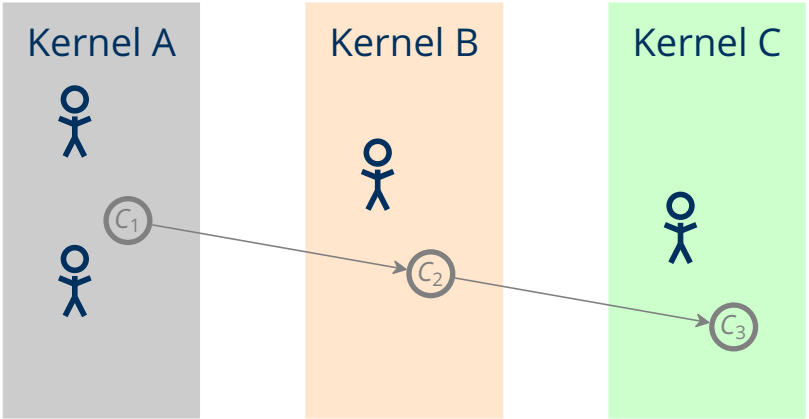
Distributed Capabilities – Delegation



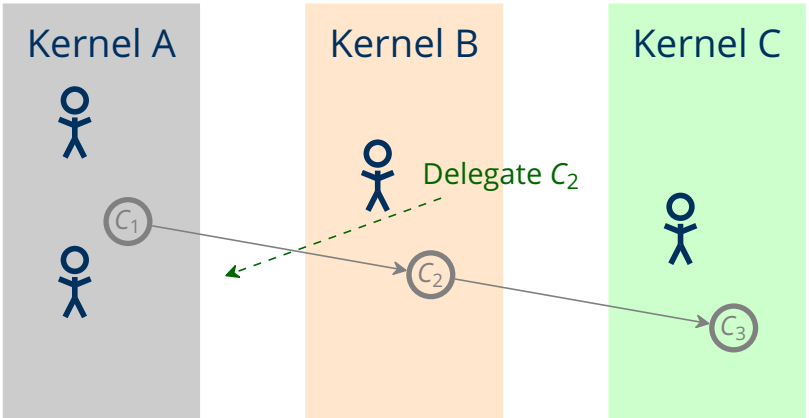
Distributed Capabilities – Delegation



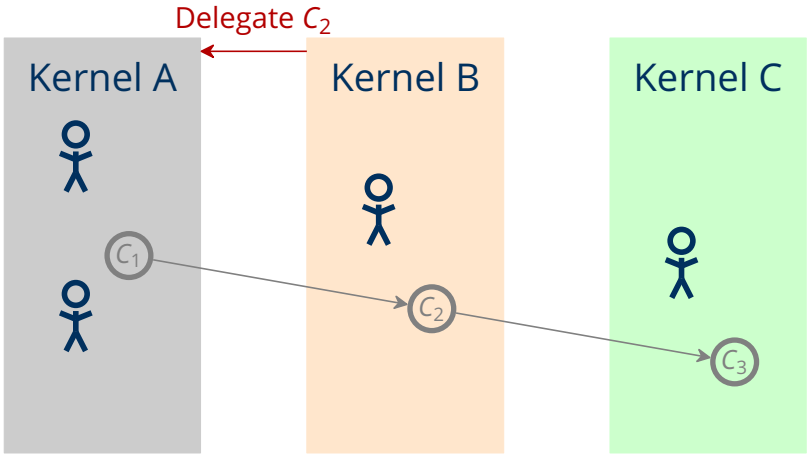
Distributed Capabilities – Delegation Cont'd



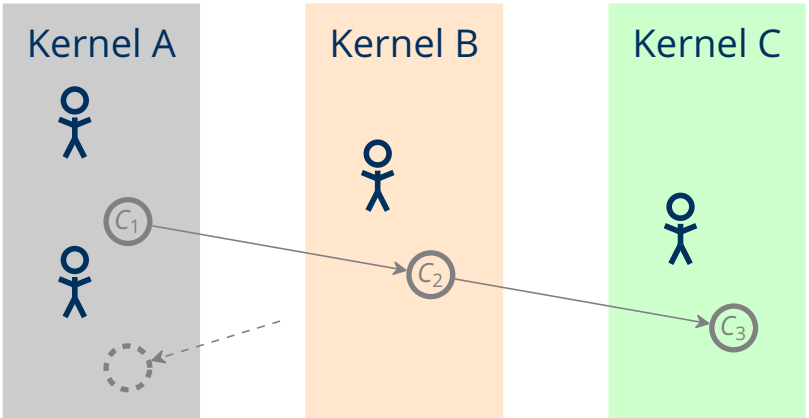
Distributed Capabilities – Delegation Cont'd



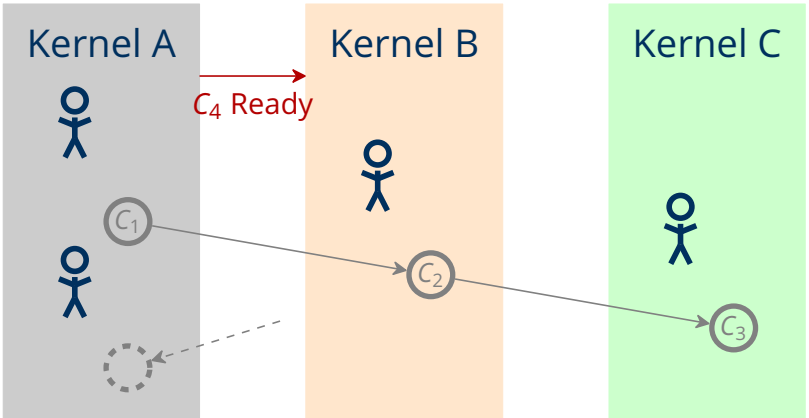
Distributed Capabilities – Delegation Cont'd



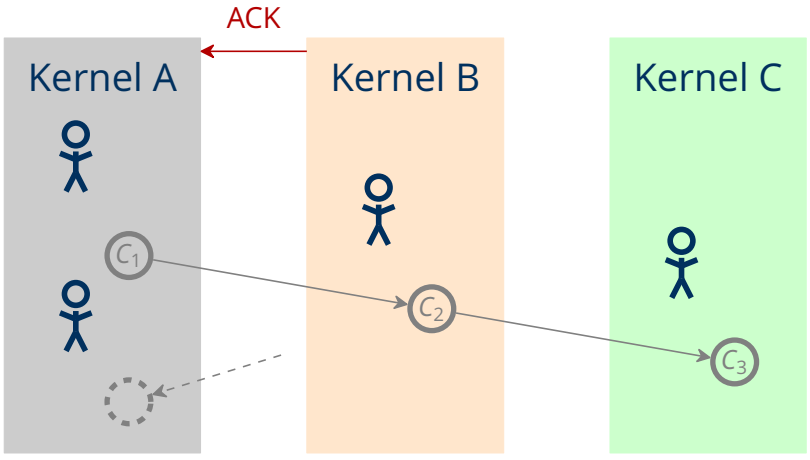
Distributed Capabilities – Delegation Cont'd



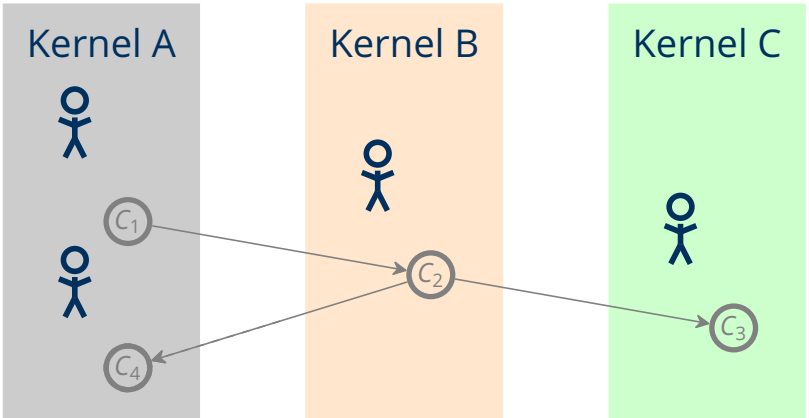
Distributed Capabilities – Delegation Cont'd



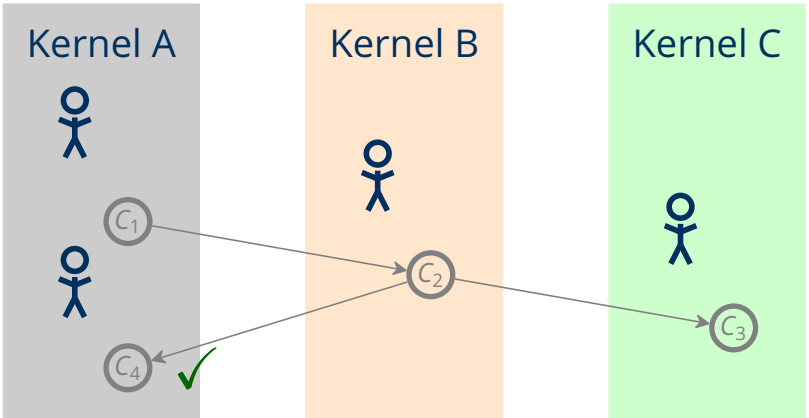
Distributed Capabilities – Delegation Cont'd



Distributed Capabilities – Delegation Cont'd



Distributed Capabilities – Delegation Cont'd



Distributed Capabilities – Interferences

1st \ 2nd	Obtain	Delegate	Revoke/Crash
Obtain	✓	✓	!
Delegate	✓	✓	↯
Revoke	!	!	↯

Evaluation

- *gem5* simulation of 640 out-of-order cores
- Application traces recorded on Linux and replayed in SemperOS
- Replicated file system services

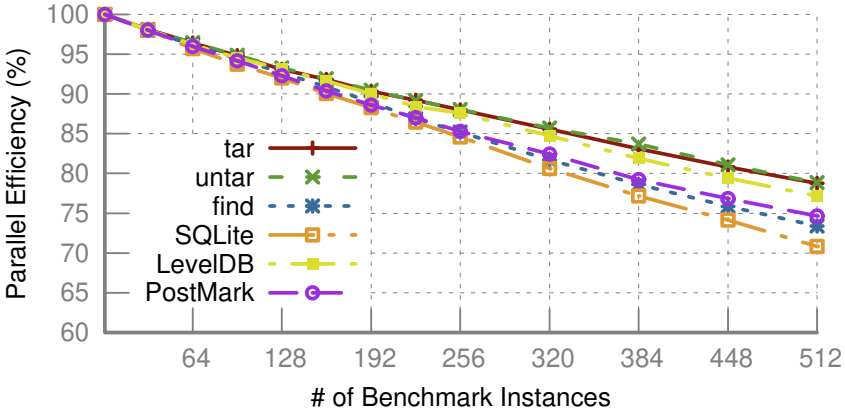
Using the Capability System – Services

- In-memory file system
- OPEN/READ/WRITE: Hand out capabilities to specific part of a file
- CLOSE: Revoke all capabilities to file

Application Benchmarks

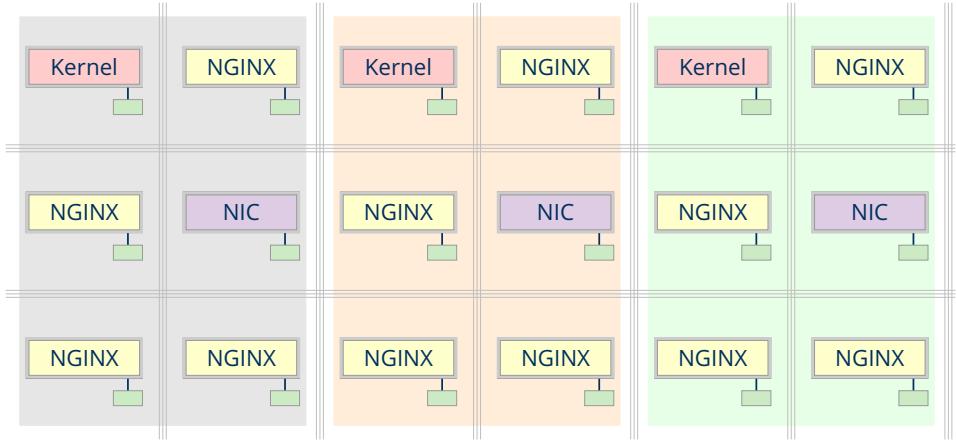
Benchmark	Cap. ops	Cap. ops/s	Cap. ops	Cap. ops/s
# of instances	1		512	
tar	21	7,295	10,752	191,703
untar	11	4,012	5,632	100,772
find	3	1,310	1,536	27,096
SQLite	24	5,987	12,288	207,072
LevelDB	22	8,749	11,264	201,204
PostMark	38	21,166	19,456	348,285

Application Benchmarks

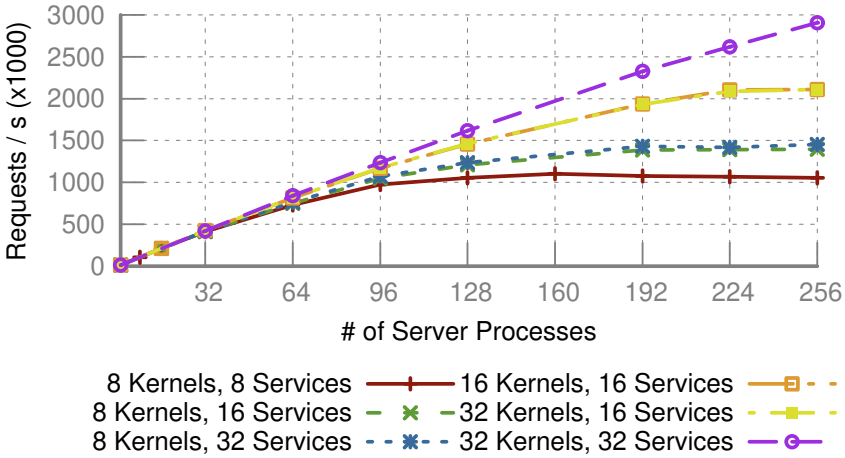


Parallel efficiency using 32 kernels and 32 file service instances.

Application Benchmarks – NGINX webserver



Application Benchmarks – NGINX webserver



SemperOS

- SemperOS implements a scalable distributed capability system
- Capabilities globally identified by DDL
- Up to 78% parallel efficiency when using 11% of the cores for the OS

Capability Systems Scale.