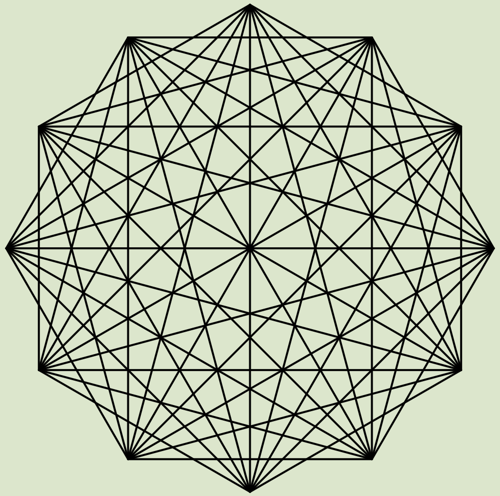
Serverless as a Bridge Between HPC and Clouds

Marcin Copik, Alexandru Calotoiu, Torsten Hoefler

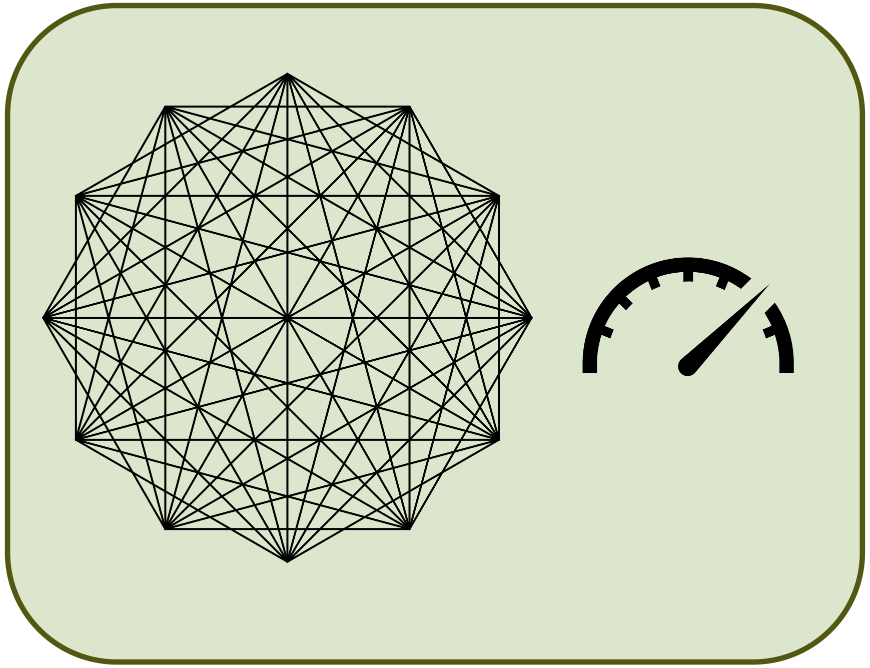
What is HPC?

What is HPC?

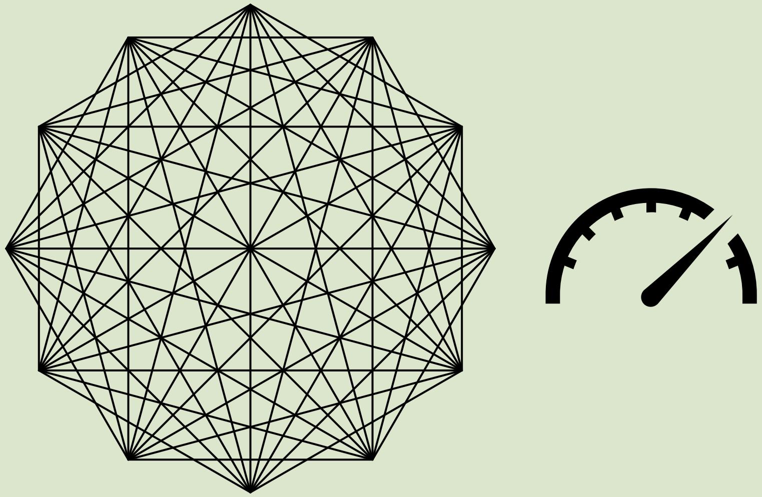
What is HPC?

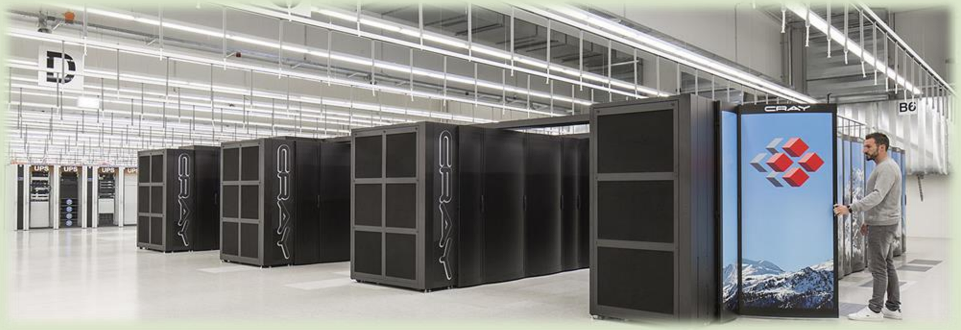


What is HPC?



What is HPC?





Piz Daint

- 5704 XC50 nodes – CPU + GPU, 64 GB memory.
- 1813 XC40 nodes – CPU, 64/128 GB memory.

Tracking Wasted Money in HPC

Tracking Wasted Money in HPC

Job Characteristics on Large-Scale Systems: Long-Term Analysis, Quantification, and Implications*

Tirthak Patel, Northeastern University
Zhengchun Liu, Raj Kettimuthu, Argonne National Laboratory
Paul Rich, William Allcock
Devesh Tiwari

A Case For Intra-rack Resource Disaggregation in HPC

GEORGE MICHELOGIANNAKIS, Lawrence Berkeley National Laboratory, USA
BENJAMIN KLENK, NVIDIA, USA
BRANDON COOK, Lawrence Berkeley National Laboratory, USA
MIN YEE TEH and MADELEINE GLICK, Columbia University, USA
LARRY DENNISON, NVIDIA, USA
KEREN BERGMAN, Columbia University, USA
JOHN SHALF, Lawrence Berkeley National Laboratory, USA

TACO, 2022

FINAL REPORT

...v, Jeffrey ...for

...Enos, and ...cations

...iv, 2017

Quantifying Memory Underutilization in HPC Systems and Using it to Improve Performance via Architecture Support

Gagandeep Panwar*, Virginia Tech, Blacksburg, USA, gpanwar@vt.edu
Da Zhang*, Virginia Tech, Blacksburg, USA, daz3@vt.edu
Yihan Pang*, Virginia Tech, Blacksburg, USA, pyihan1@vt.edu
Mai Dahshan, Virginia Tech, Blacksburg, USA, mdahshan@vt.edu
Nathan DeBardeleben, Los Alamos National Laboratory, Los Alamos, USA, ndebard@lanl.gov
Binoy Ravindran, Virginia Tech, Blacksburg, USA, binoy@vt.edu
Xun Jian, Virginia Tech, Blacksburg, USA, xunj@vt.edu

MICRO, 2019

A Holistic View of Memory Utilization on HPC Systems: Current and Future Trends

Ivy B. Peng*, peng8@llnl.gov, Lawrence Livermore National Laboratory, USA
Ian Karlin, karlin1@llnl.gov, Lawrence Livermore National Laboratory, USA
Maya B. Gokhale, gokhale2@llnl.gov, Lawrence Livermore National Laboratory, USA
Kathleen Shoga, Shoga1@llnl.gov, Lawrence Livermore National Laboratory, USA
Matthew Legendre, legendre1@llnl.gov, Lawrence Livermore National Laboratory, USA
Todd Gamblin, gamblin2@llnl.gov, Lawrence Livermore National Laboratory, USA

MEMSYS, 2021

an...

University of Tennessee, Knoxville, TN 37996, USA
{hyou,haozhang}@utk.edu

JSSPP, 2012

Tracking Wasted Money in HPC

FINAL REPORT

v, Jeffrey
for

Enos, and
cations

iv, 2017

Quantifying Memory Underutilization in HPC Systems and Using it to Improve Performance via Architecture Support

Gagandeep Panwar*
Virginia Tech
Blacksburg, USA
gpanwar@vt.edu

Da Zhang*
Virginia Tech
Blacksburg, USA
daz3@vt.edu

Yihan Pang*
Virginia Tech
Blacksburg, USA
pyihan1@vt.edu

Mai Dahshan
Virginia Tech
Blacksburg, USA
mdahshan@vt.edu

Nathan DeBardeleben
Los Alamos National Laboratory
Los Alamos, USA
ndebard@lanl.gov

Binoy Ravindran
Virginia Tech
Blacksburg, USA
binoy@vt.edu

Xun Jian
Virginia Tech
Blacksburg, USA
xunj@vt.edu

MICRO, 2019

Job Characteristics on Large-Scale Systems: Long-Term Analysis, Quantification, and Implications*

Tirthak Patel
Northeastern University

Zhengchun Liu, Raj Kettimuthu
Argonne National Laboratory

Paul Rich, William Allcock

Devesh Tiwari

A Case For Intra-rack Resource Disaggregation in HPC

GEORGE MICHELOGIANNAKIS, Lawrence Berkeley National Laboratory, USA
BENJAMIN KLENK, NVIDIA, USA
BRANDON COOK, Lawrence Berkeley National Laboratory, USA
MIN YEE TEH and MADELEINE GLICK, Columbia University, USA
LARRY DENNISON, NVIDIA, USA
KEREN BERGMAN, Columbia University, USA
JOHN SHALF, Lawrence Berkeley National Laboratory, USA

TACO, 2022

and

A Holistic View of Memory Utilization on HPC Systems: Current and Future Trends

Ivy B. Peng*
peng8@llnl.gov
Lawrence Livermore National Laboratory
USA

Ian Karlin
karlin1@llnl.gov
Lawrence Livermore National Laboratory
USA

Maya B. Gokhale
gokhale2@llnl.gov
Lawrence Livermore National Laboratory
USA

Kathleen Shoga
Shoga1@llnl.gov
Lawrence Livermore National Laboratory
USA

Matthew Legendre
legendre1@llnl.gov
Lawrence Livermore National Laboratory
USA

Todd Gamblin
gamblin2@llnl.gov
Lawrence Livermore National Laboratory
USA

MEMSYS, 2021

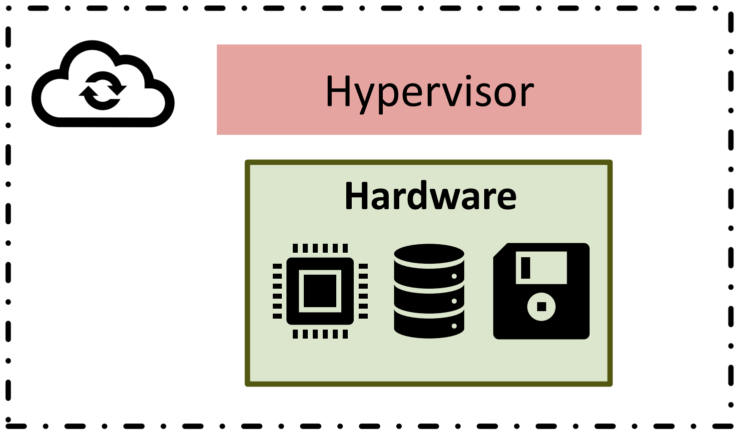
University of Tennessee, Knoxville, TN 37996, USA
{hyou,haozhang}@utk.edu

JSSPP, 2012

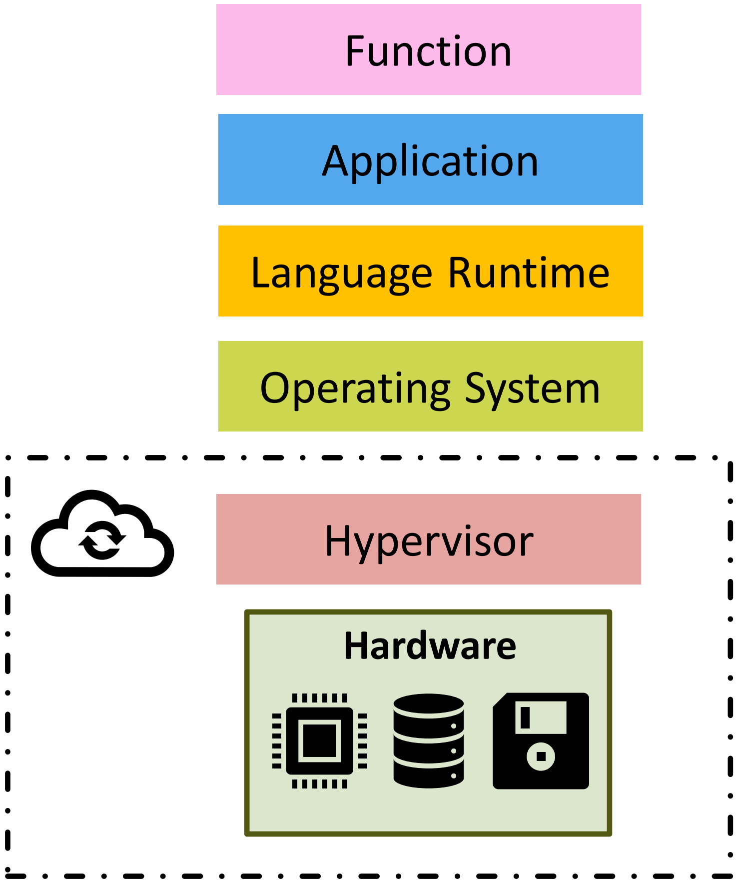
Can we solve underutilization with sharing and fine-grained allocations?

Cloud and Serverless

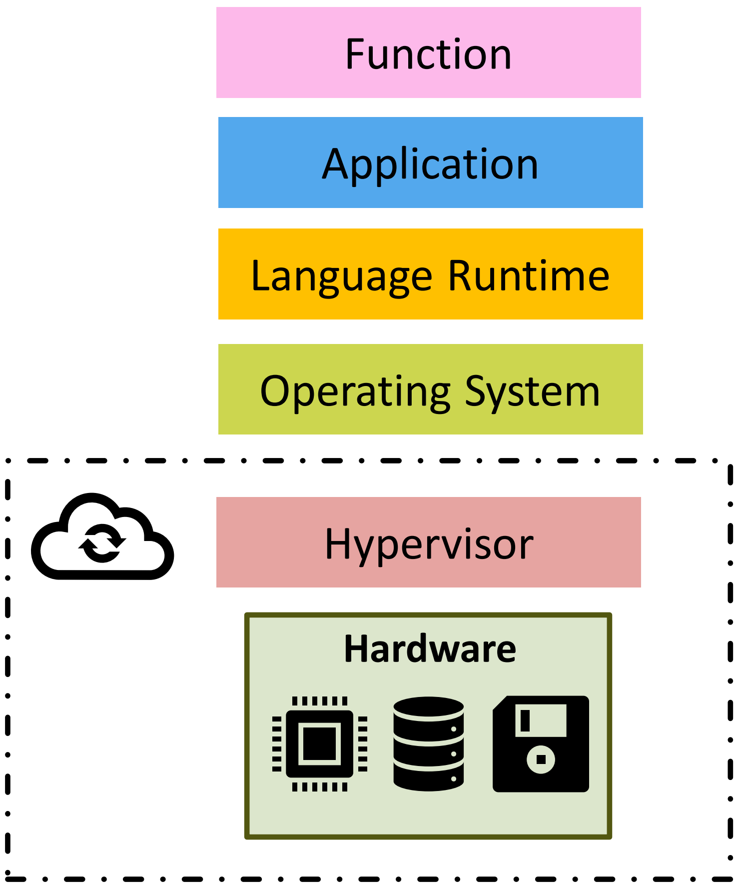
Cloud and Serverless



Cloud and Serverless



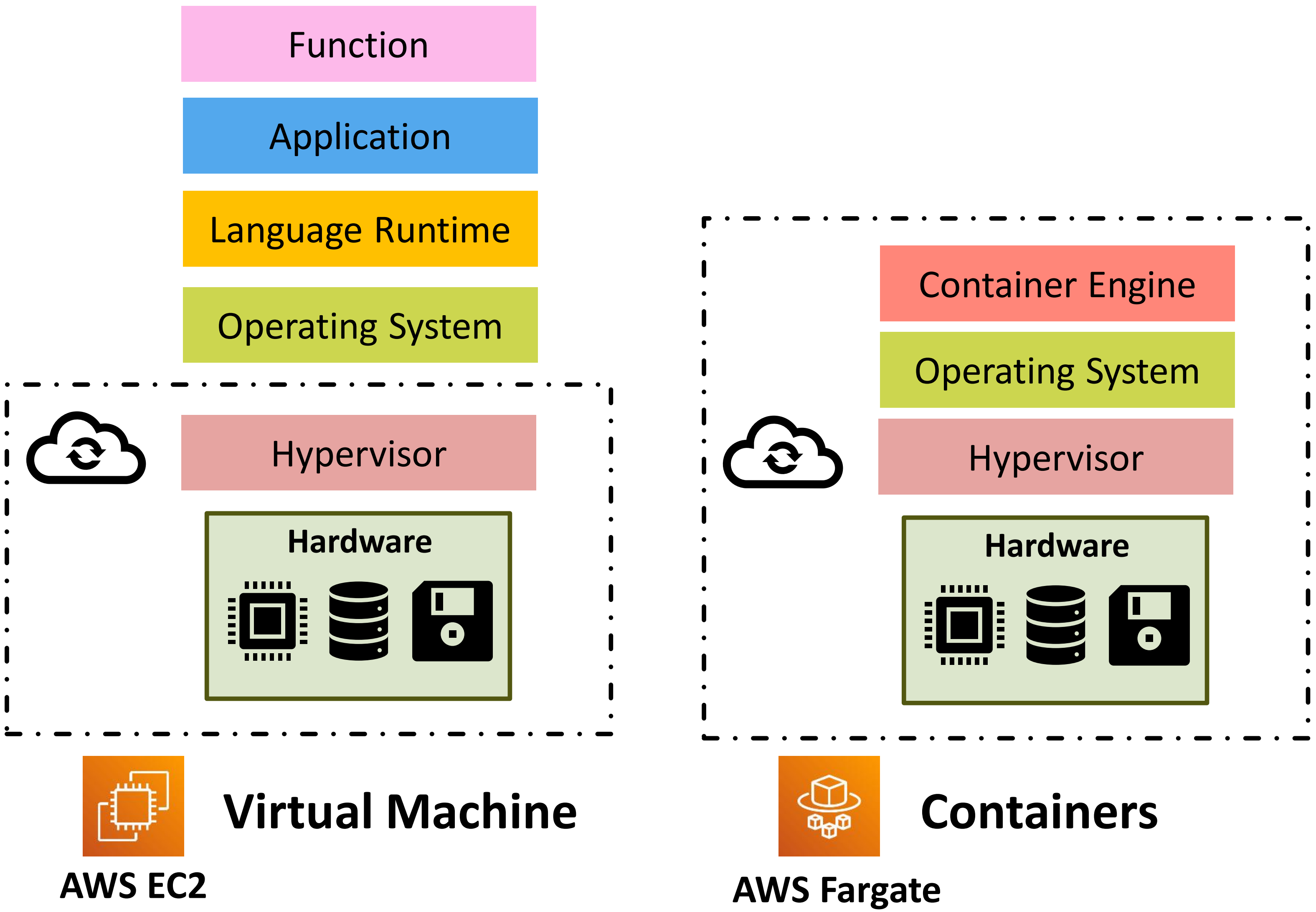
Cloud and Serverless



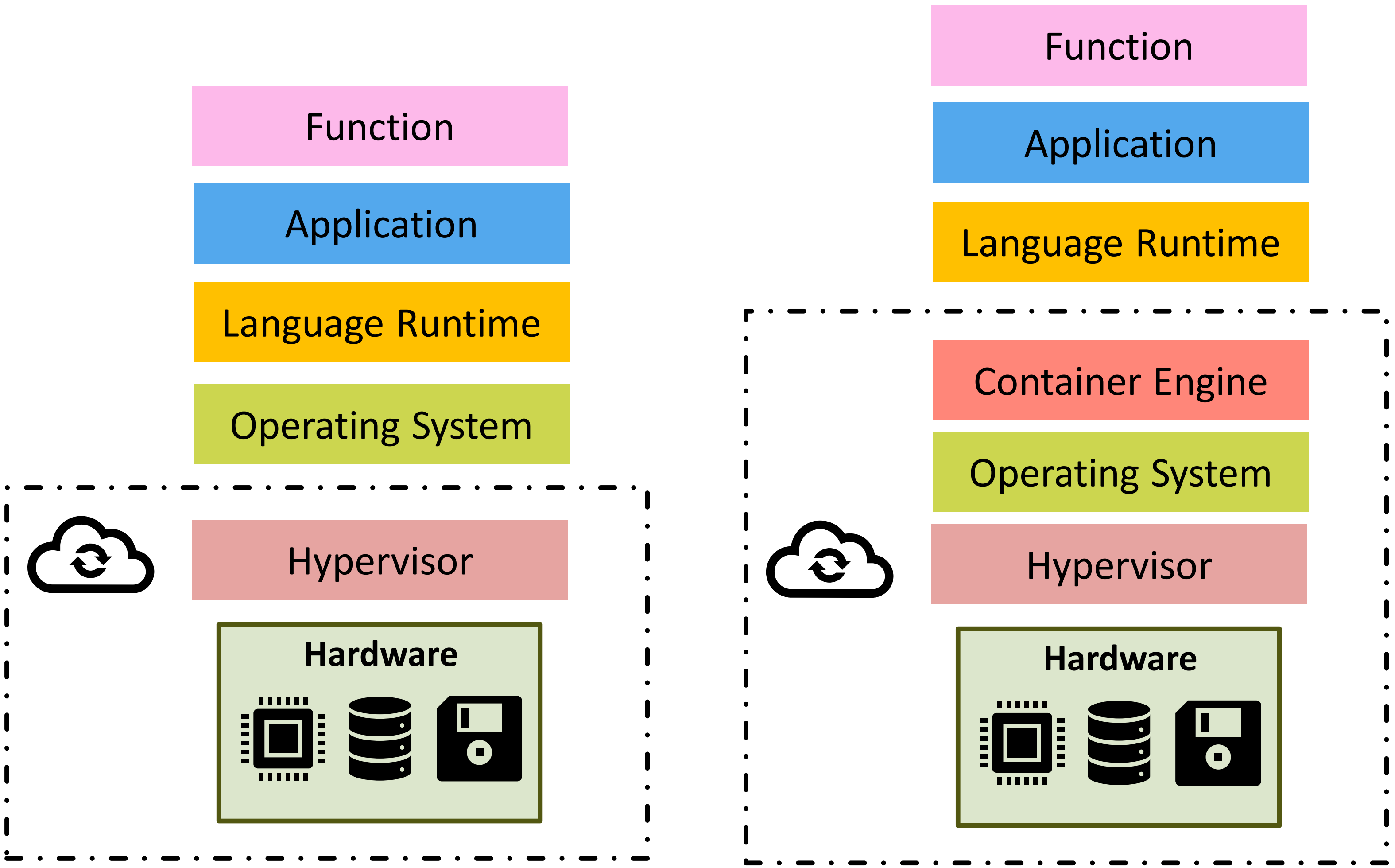
Virtual Machine

AWS EC2

Cloud and Serverless



Cloud and Serverless



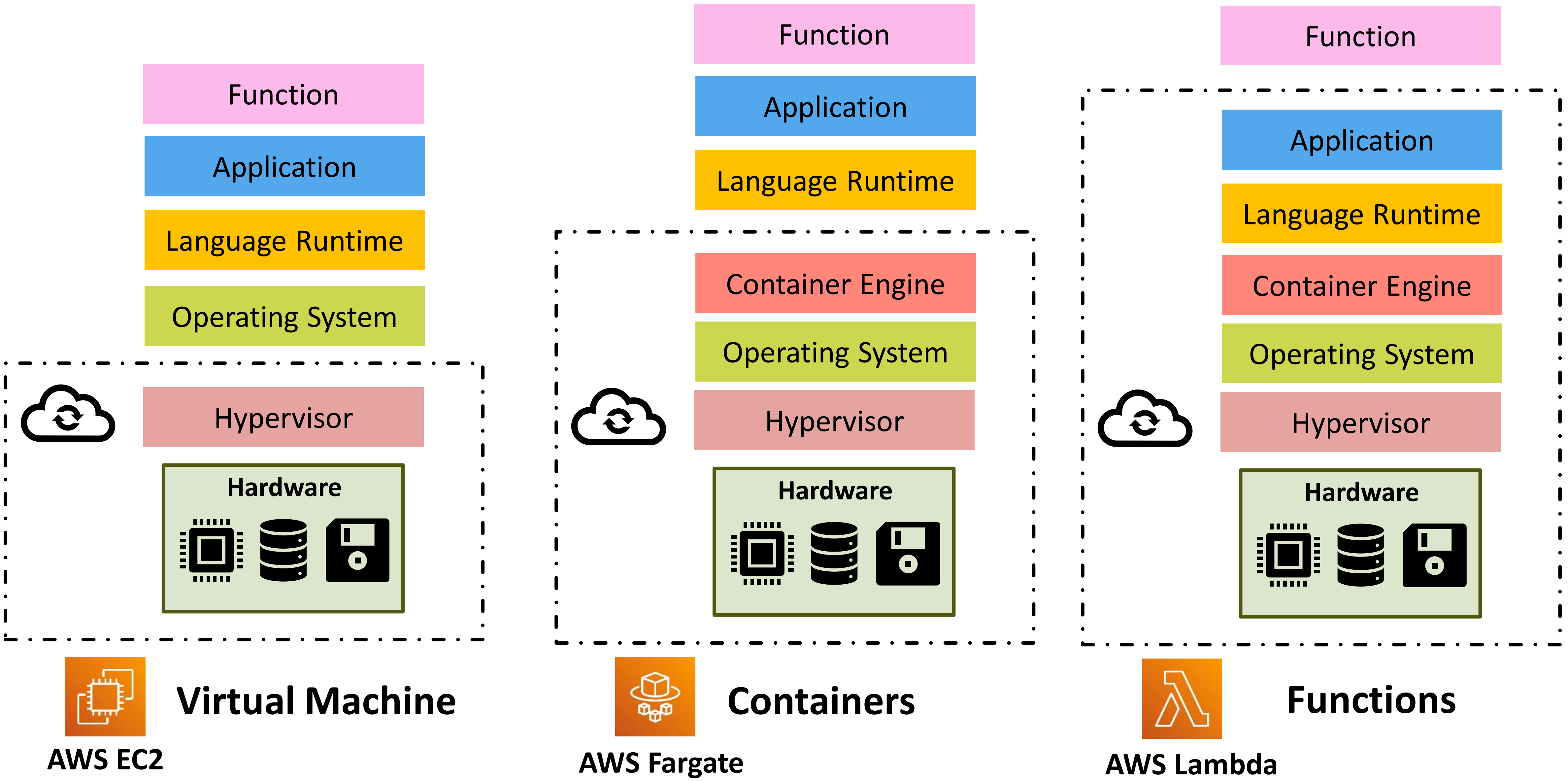
Virtual Machine

AWS EC2

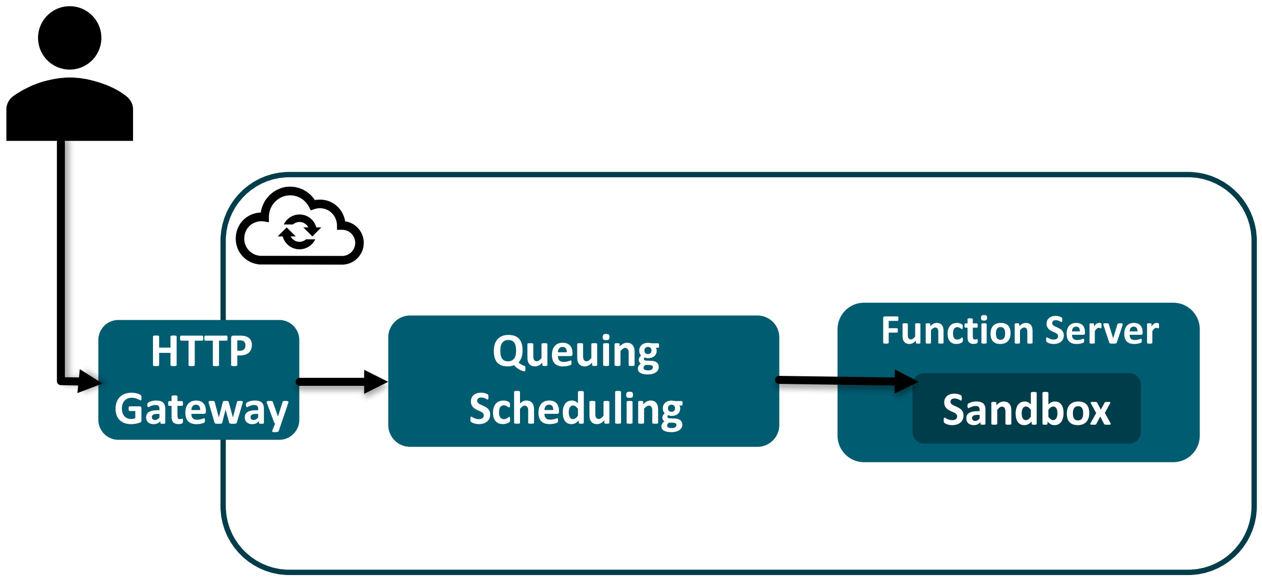
Containers

AWS Fargate

Cloud and Serverless

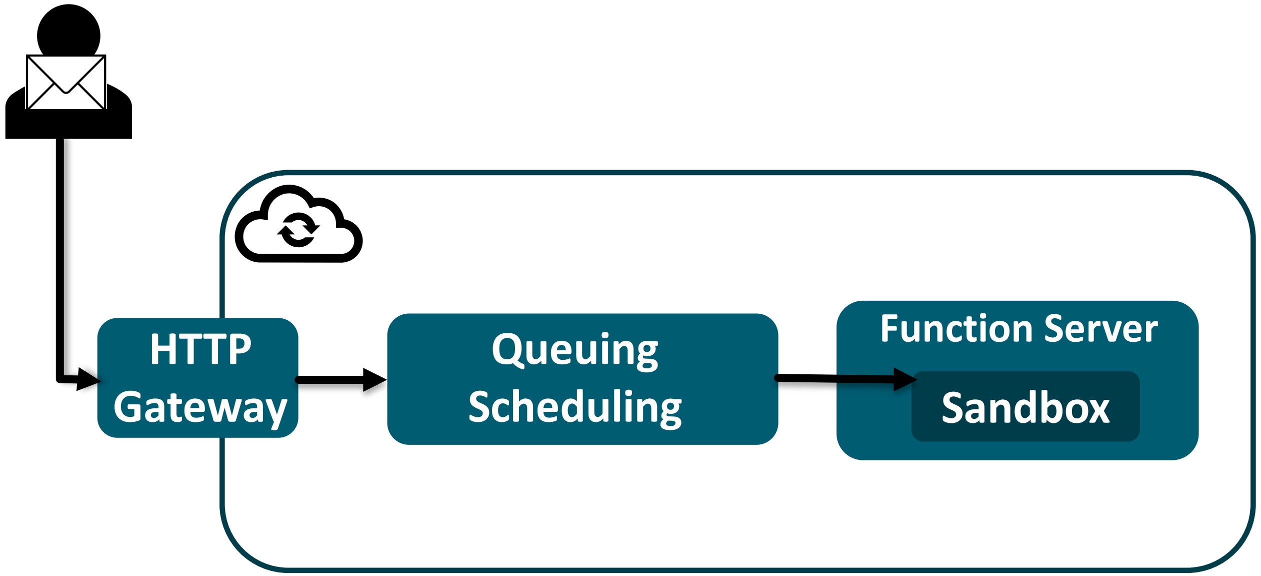


How does Function-as-a-Service (FaaS) work?



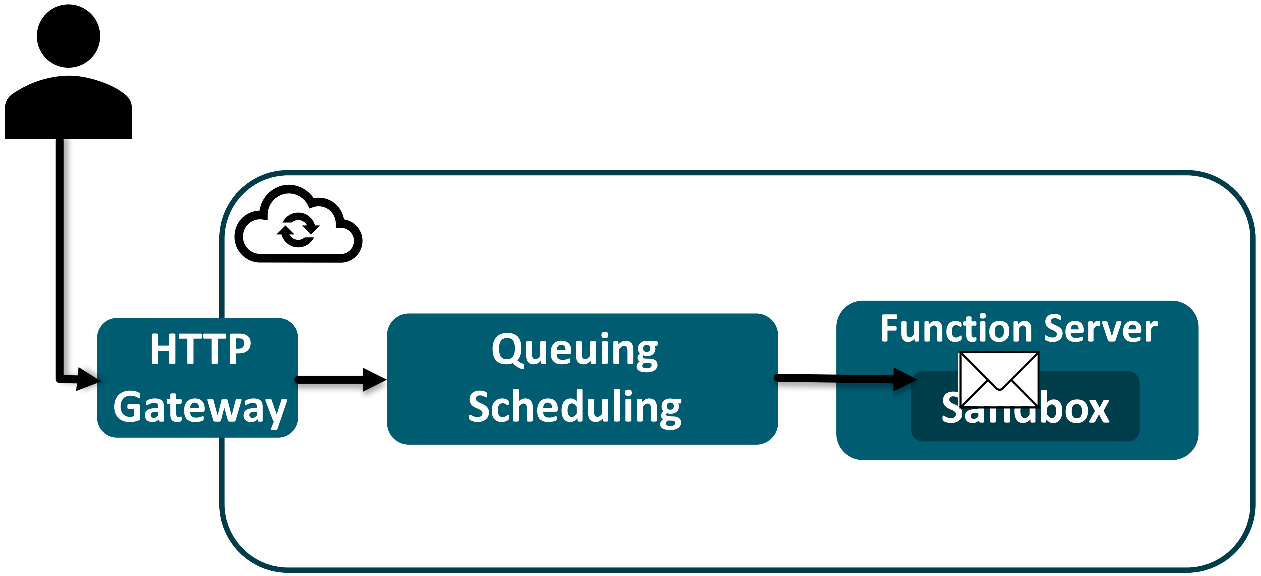
"SeBS: a Serverless Benchmark Suite for Function-as-a-Service Computing", Middleware 2021

How does Function-as-a-Service (FaaS) work?



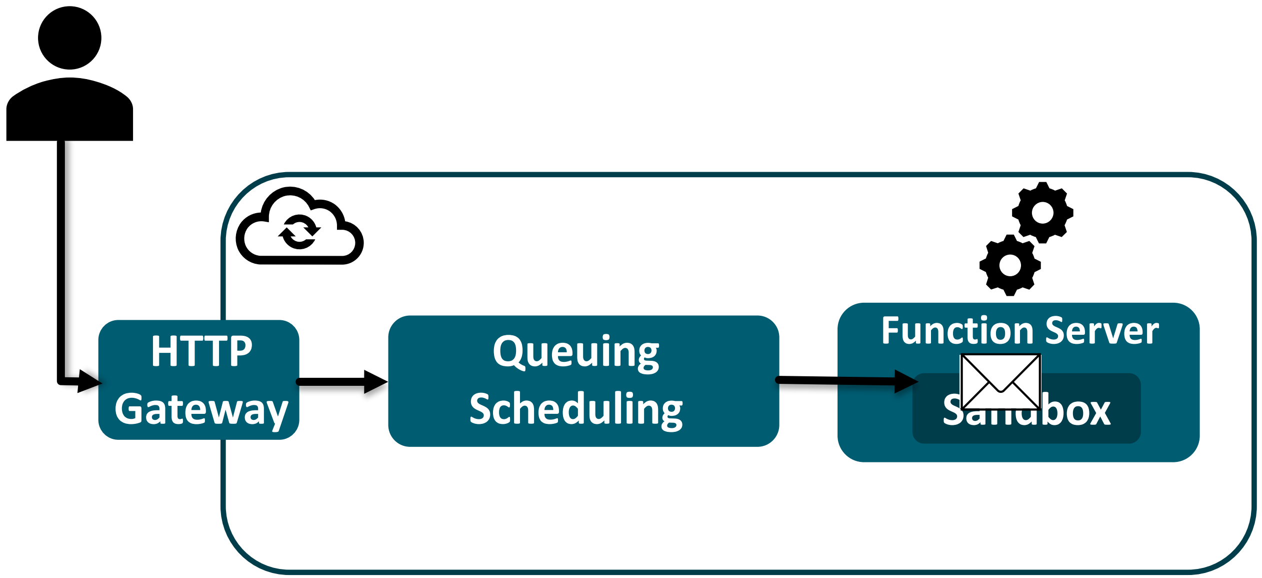
"SeBS: a Serverless Benchmark Suite for Function-as-a-Service Computing", Middleware 2021

How does Function-as-a-Service (FaaS) work?



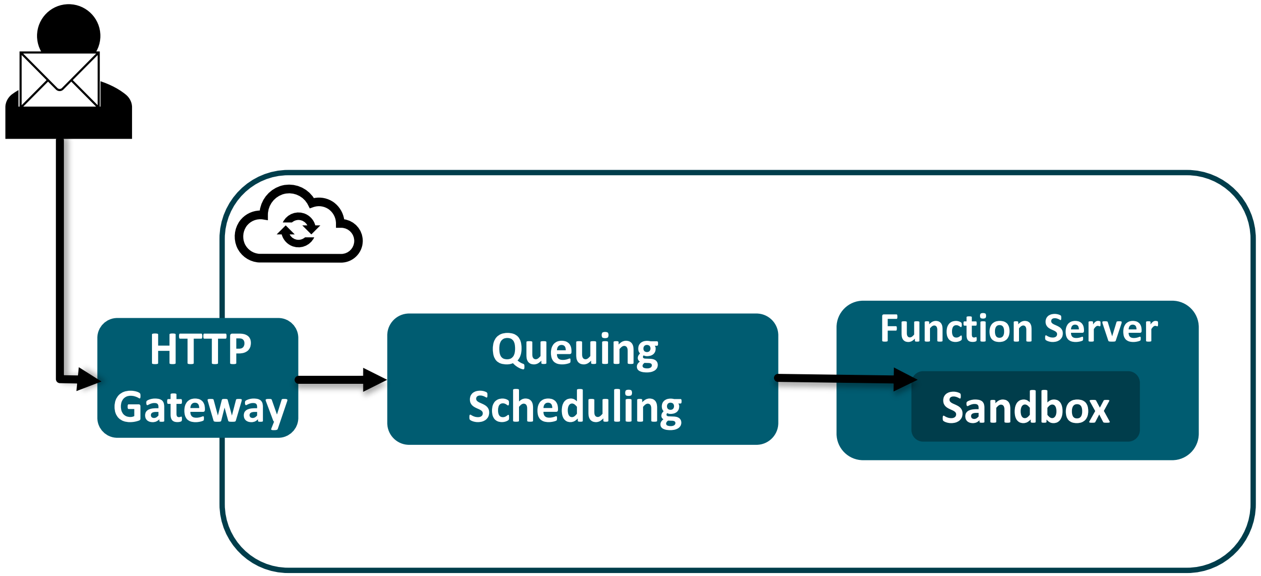
“SeBS: a Serverless Benchmark Suite for Function-as-a-Service Computing”, Middleware 2021

How does Function-as-a-Service (FaaS) work?



“SeBS: a Serverless Benchmark Suite for Function-as-a-Service Computing”, Middleware 2021

How does Function-as-a-Service (FaaS) work?



"SeBS: a Serverless Benchmark Suite for Function-as-a-Service Computing", Middleware 2021

“But serverless is slow and expensive”

"But serverless is slow and expensive"

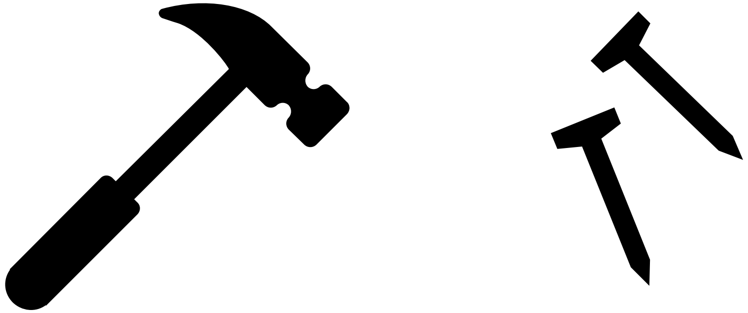
Scaling up the Prime Video audio/video monitoring service and reducing costs by 90%

The move from a distributed microservices architecture to a monolith application helped achieve higher scale, resilience, and reduce costs.

"But serverless is slow and expensive"

Scaling up the Prime Video audio/video monitoring service and reducing costs by 90%

The move from a distributed microservices architecture to a monolith application helped achieve higher scale, resilience, and reduce costs.



Serverless for High-Performance Applications

Serverless for High-Performance Applications

Serverless is slow

Serverless for High-Performance Applications

Serverless is slow

Communication is slow and restricted

Serverless for High-Performance Applications

Serverless is slow

Communication is slow and restricted

Serverless is hard to program.

Serverless for High-Performance Applications

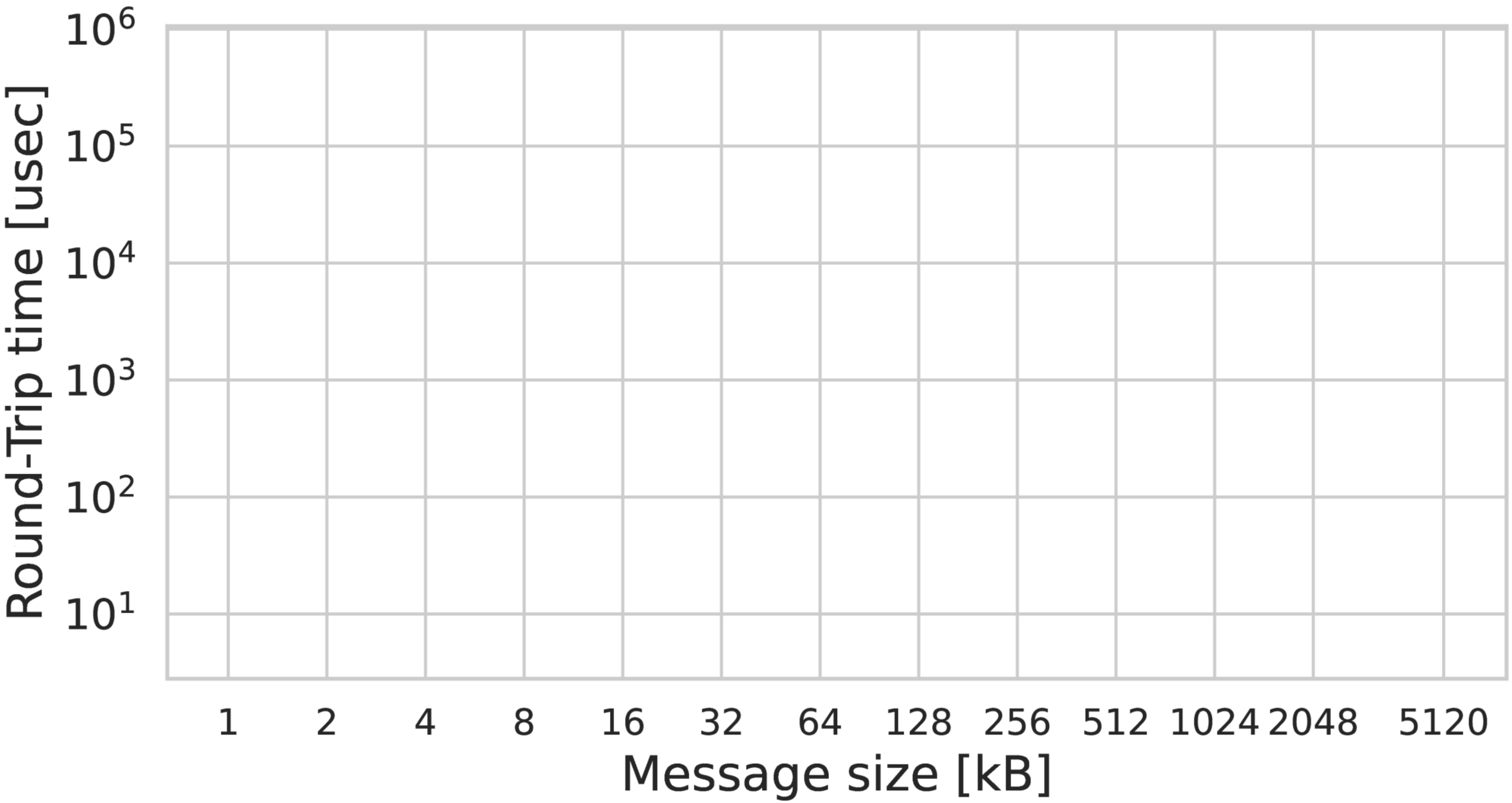
Serverless is slow

Communication is slow and restricted

Answer: rFaaS

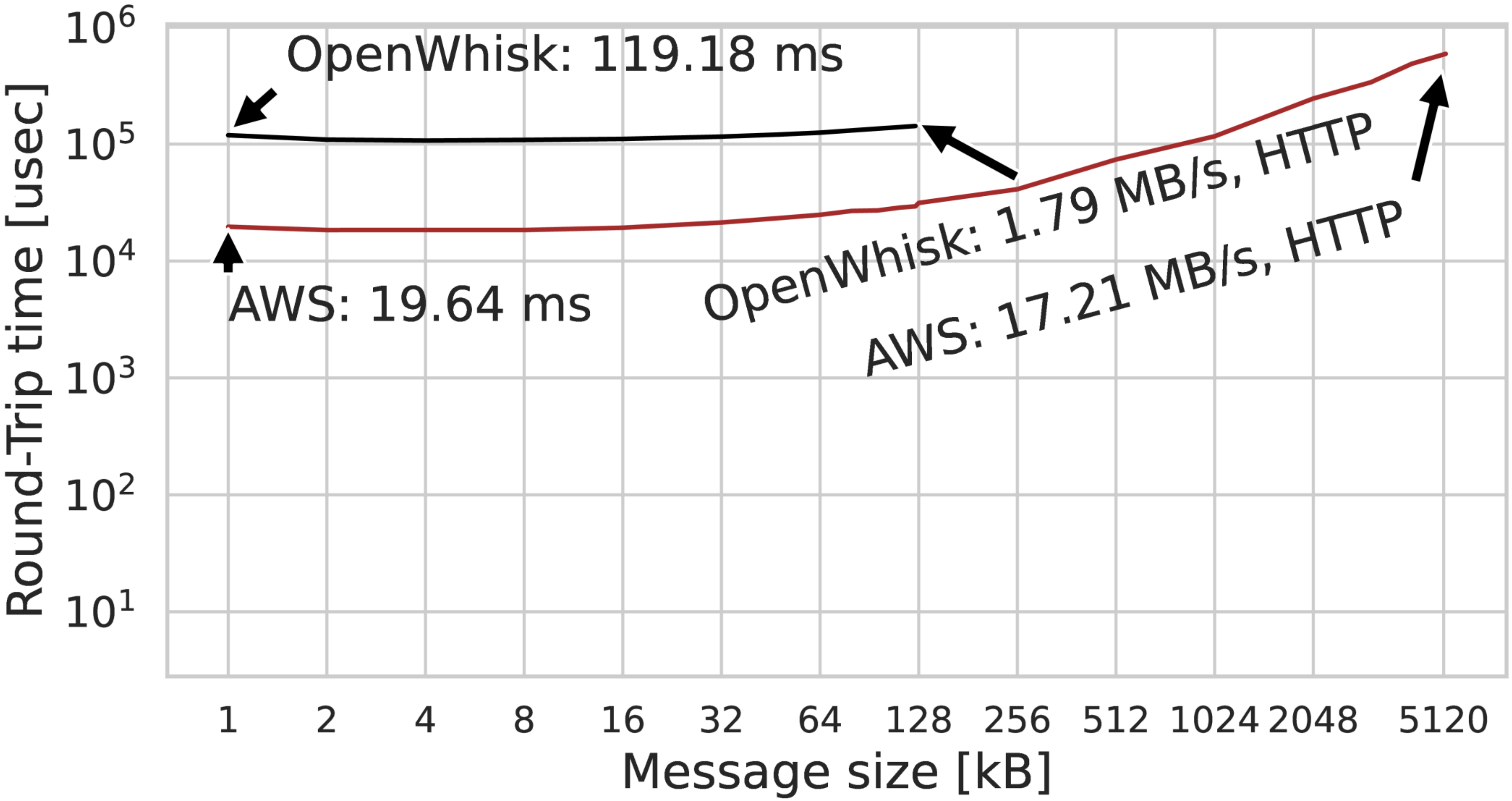
Serverless is hard to program.

How fast are invocations in FaaS?



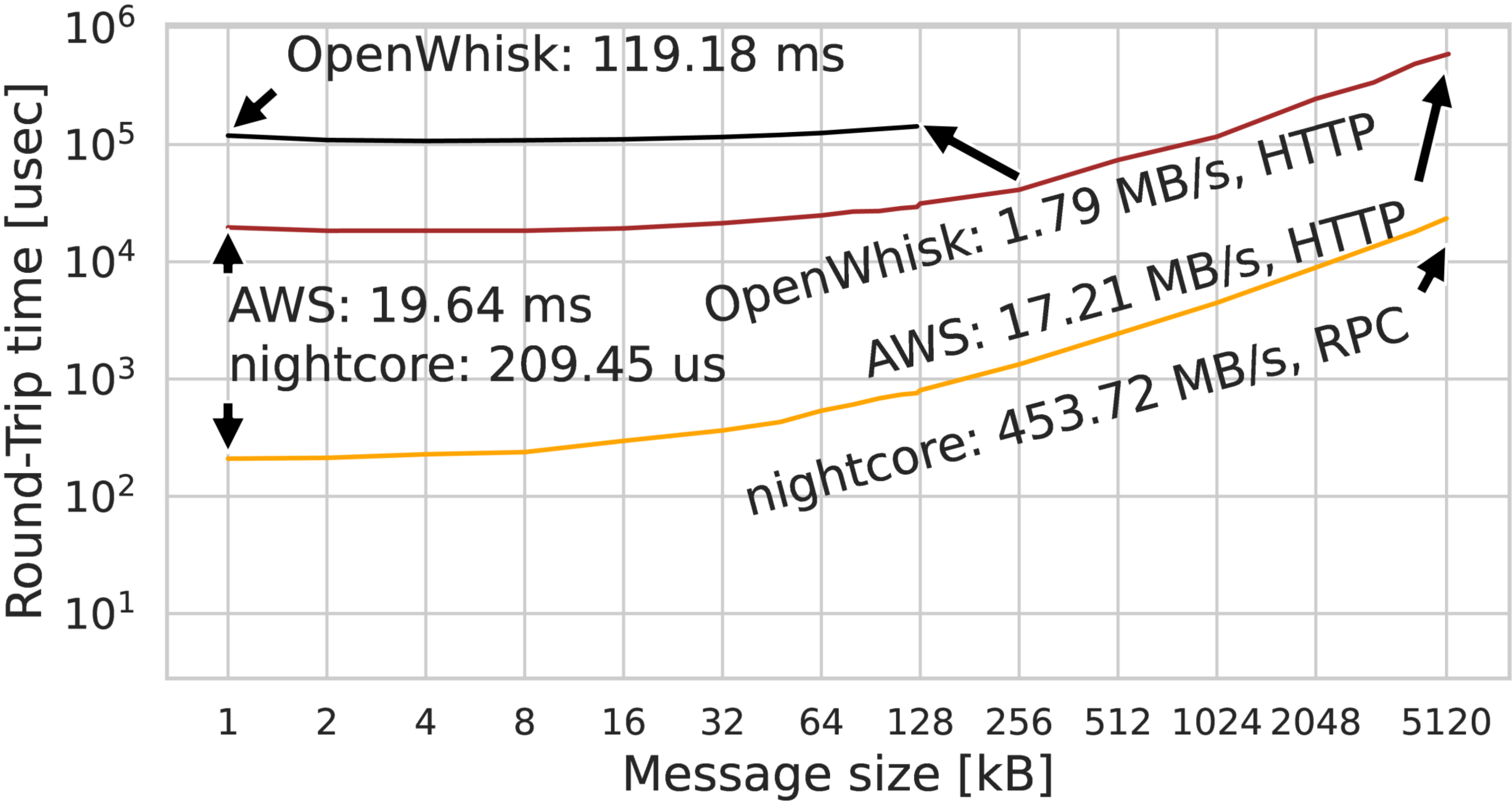
"rFaaS: Enabling High Performance Serverless with RDMA and Leases", IPDPS'23

How fast are invocations in FaaS?



"rFaaS: Enabling High Performance Serverless with RDMA and Leases", IPDPS'23

How fast are invocations in FaaS?

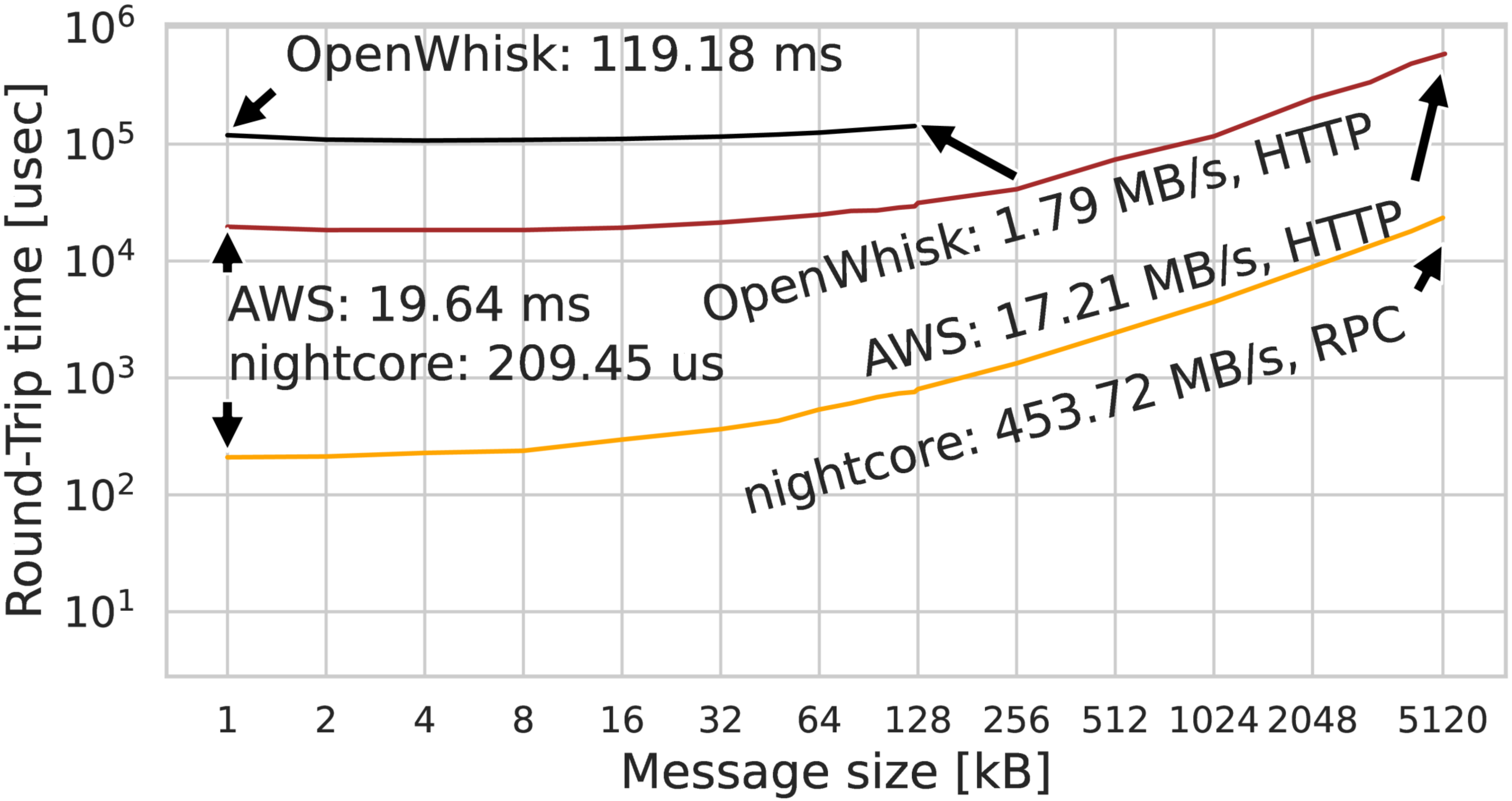


"rFaaS: Enabling High Performance Serverless with RDMA and Leases", IPDPS'23

How fast are invocations in FaaS?

Reduced invocation critical path

Zero-copy RDMA networking

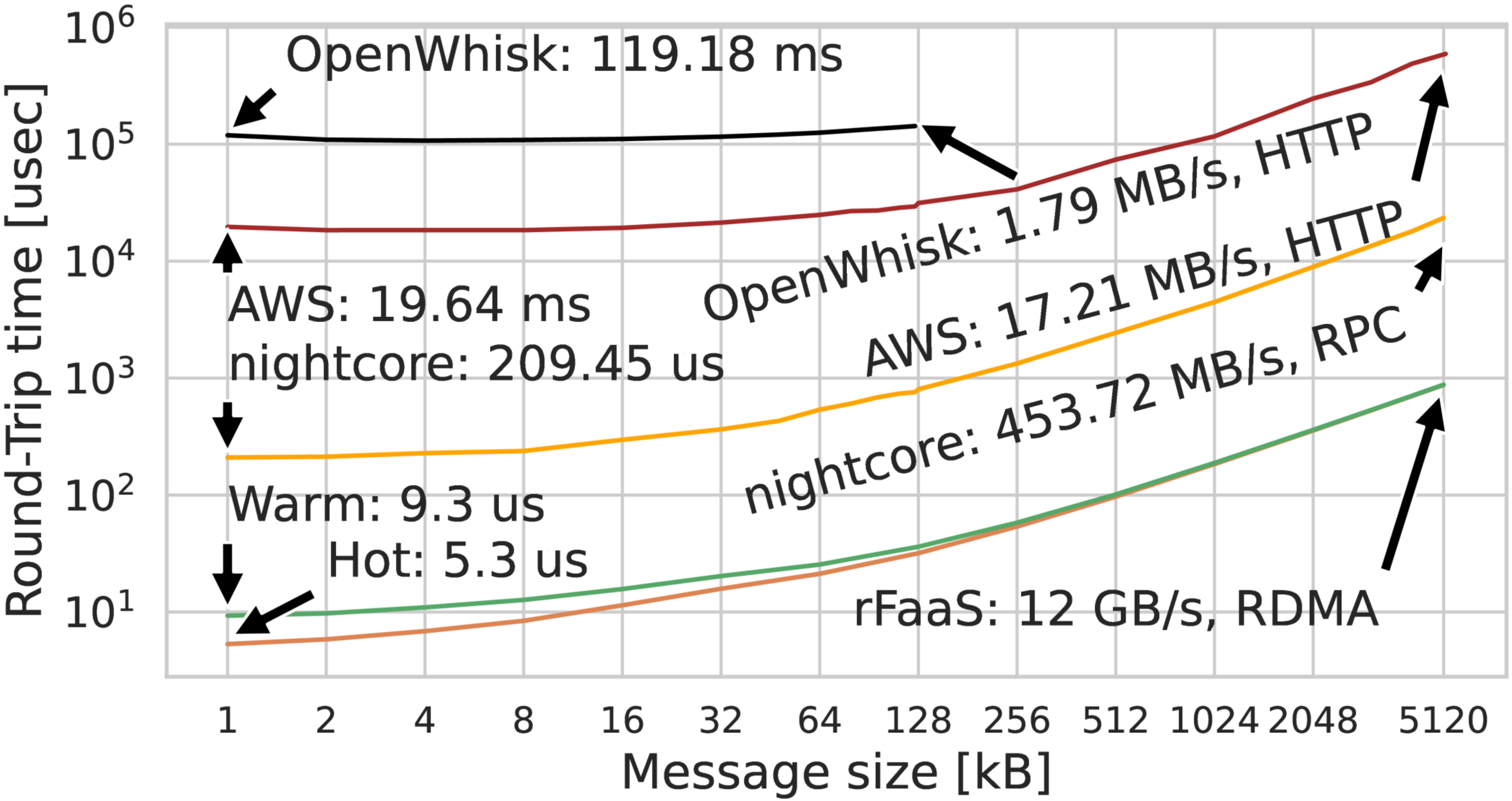


"rFaaS: Enabling High Performance Serverless with RDMA and Leases", IPDPS'23

How fast are invocations in FaaS?

Reduced invocation critical path

Zero-copy RDMA networking



"rFaaS: Enabling High Performance Serverless with RDMA and Leases", IPDPS'23

FaaS in High-Performance Applications

Serverless is slow

Communication is slow and restricted

Answer: rFaaS

Serverless is hard to program.

FaaS in High-Performance Applications

Serverless is slow

Answer:
rFaaS

Communication is slow and restricted

Answer:
FMI

Serverless is hard to program.

Communication in serverless

“FMI: Fast and Cheap Message Passing for Serverless Functions”, ICS’23

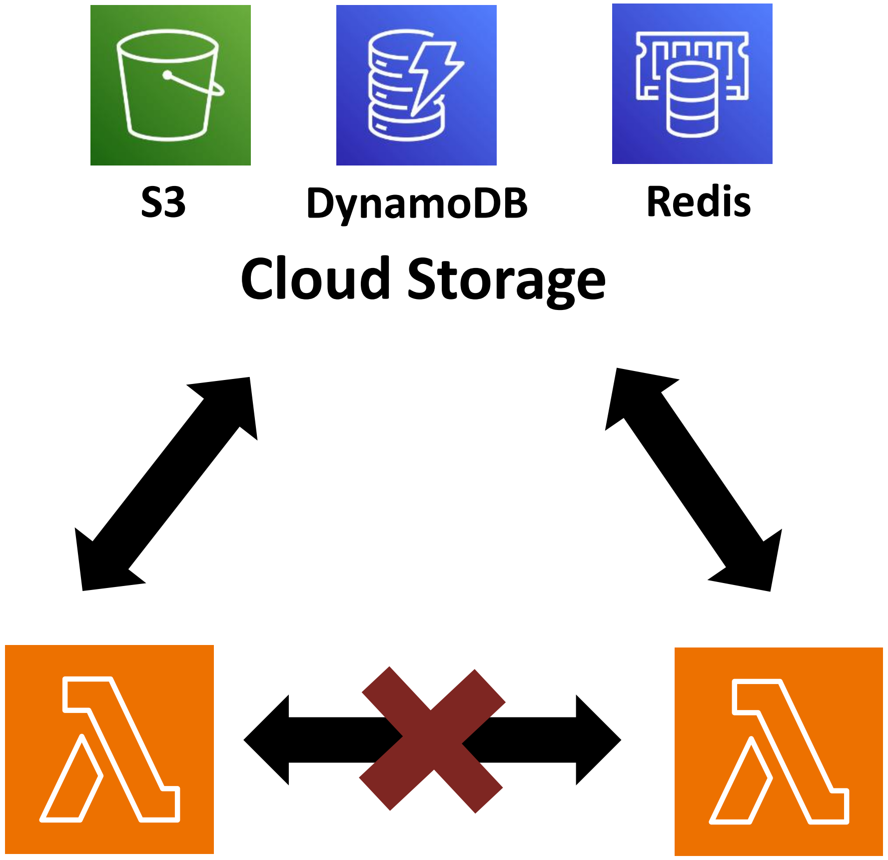
Communication in serverless

“FMI: Fast and Cheap Message Passing for Serverless Functions”, ICS’23

Communication in serverless

"FMI: Fast and Cheap Message Passing for Serverless Functions", ICS'23

Communication in serverless

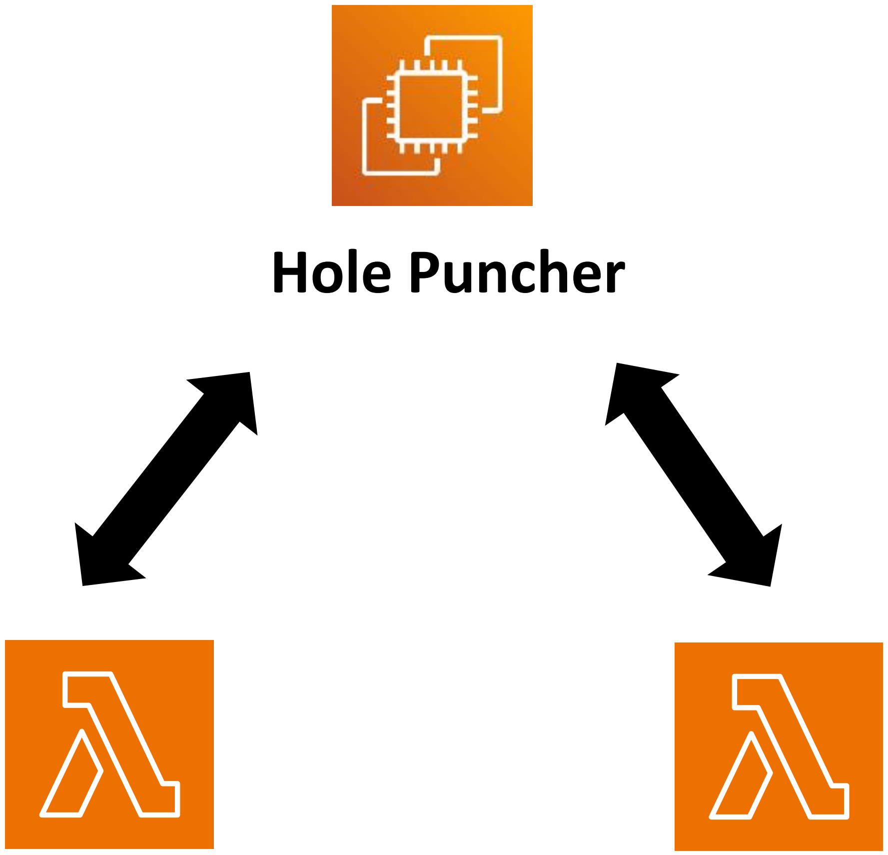


"FMI: Fast and Cheap Message Passing for Serverless Functions", ICS'23

Communication in serverless

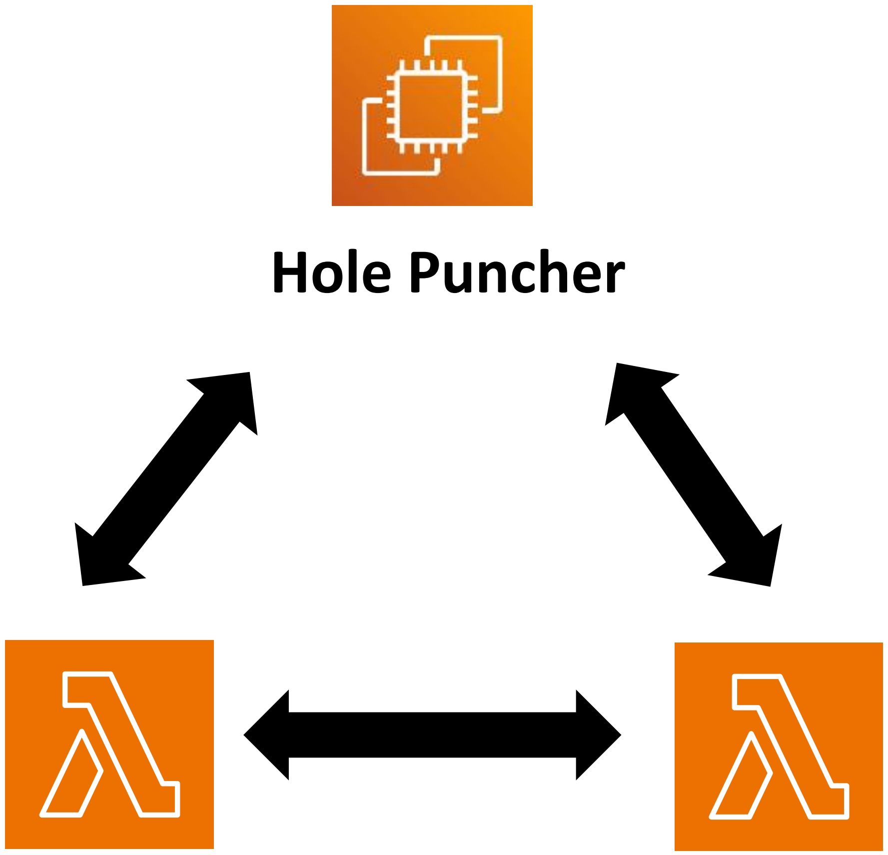
“FMI: Fast and Cheap Message Passing for Serverless Functions”, ICS’23

Communication in serverless



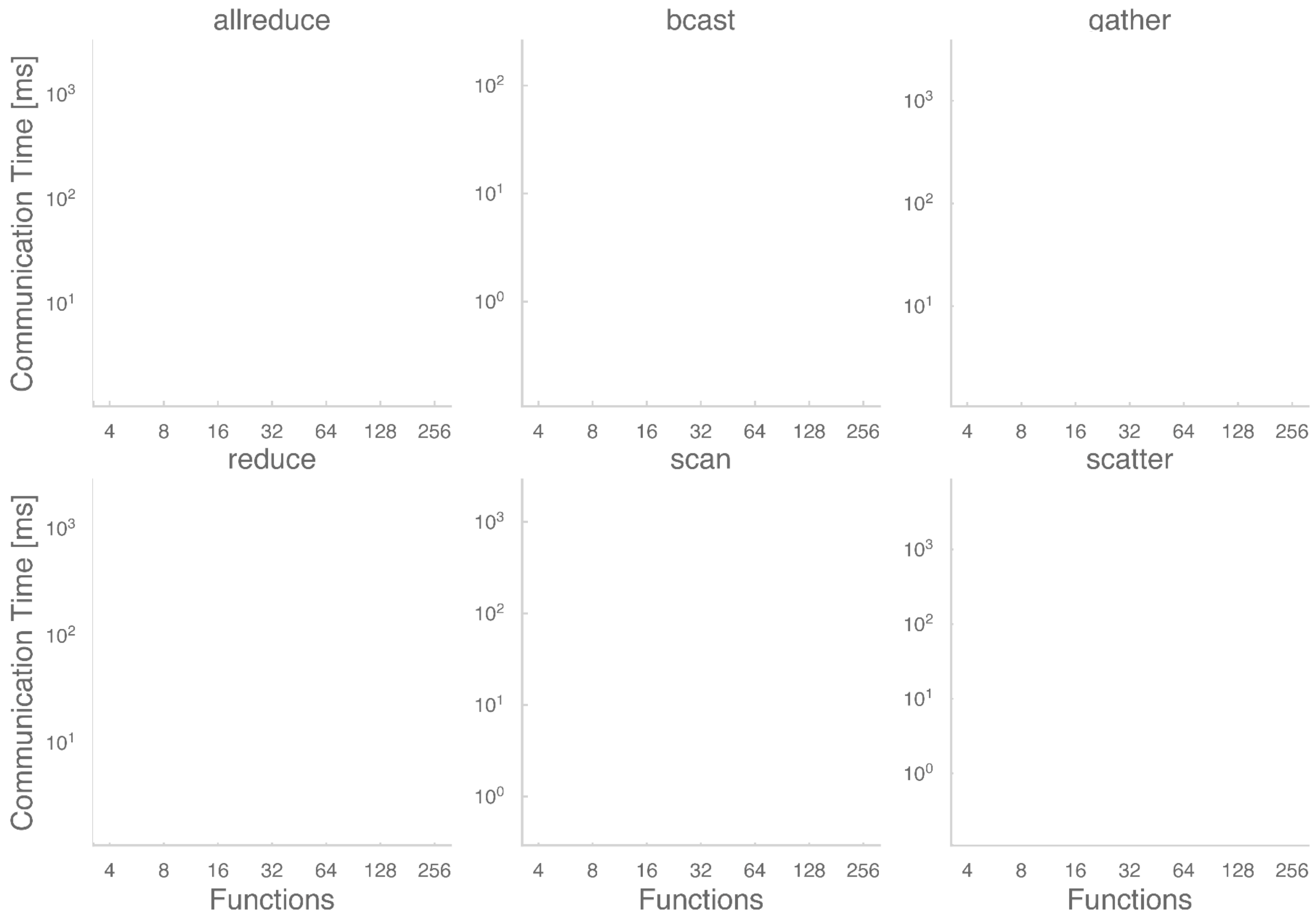
"FMI: Fast and Cheap Message Passing for Serverless Functions", ICS'23

Communication in serverless



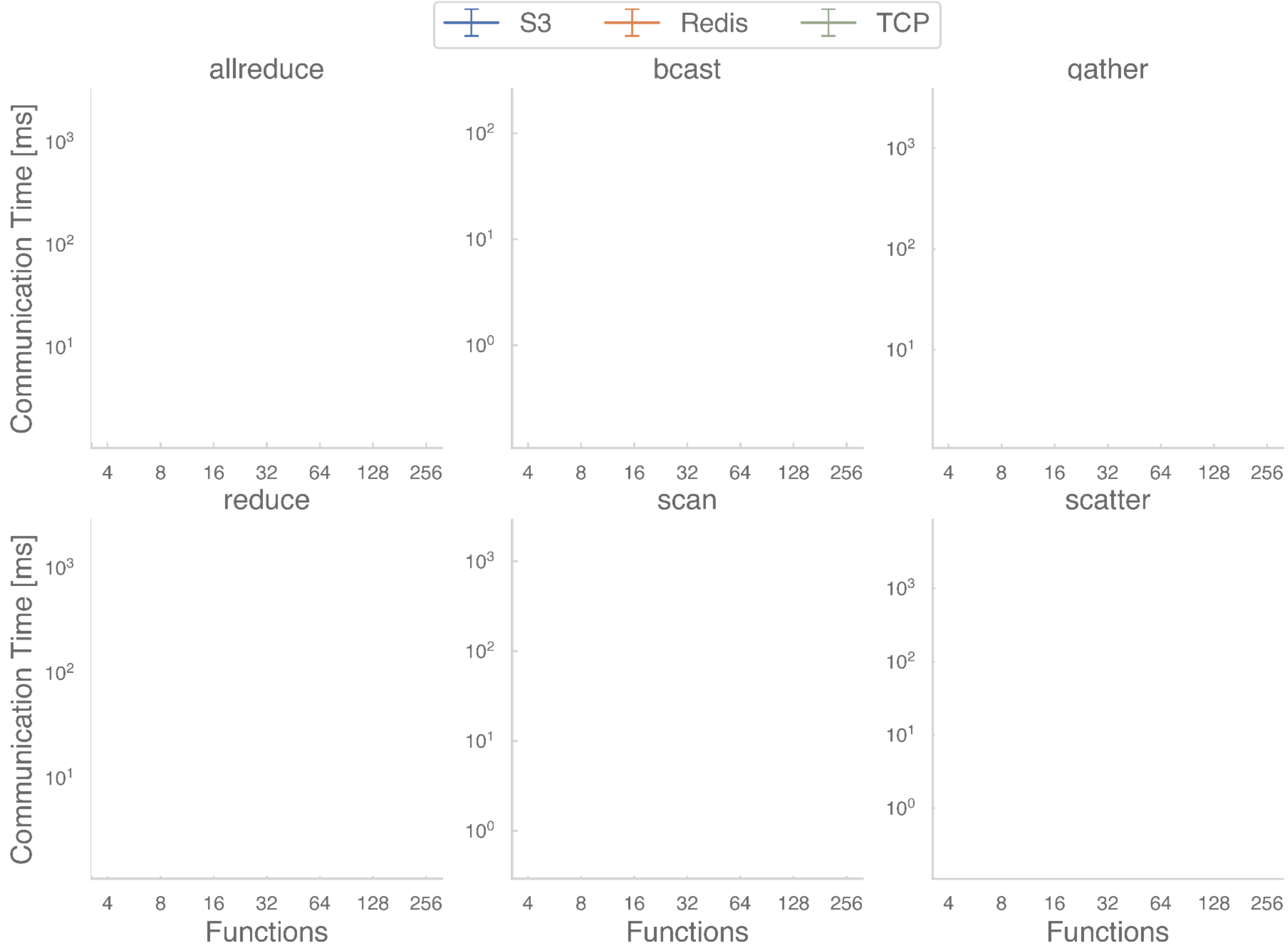
"FMI: Fast and Cheap Message Passing for Serverless Functions", ICS'23

FMI on AWS Lambda



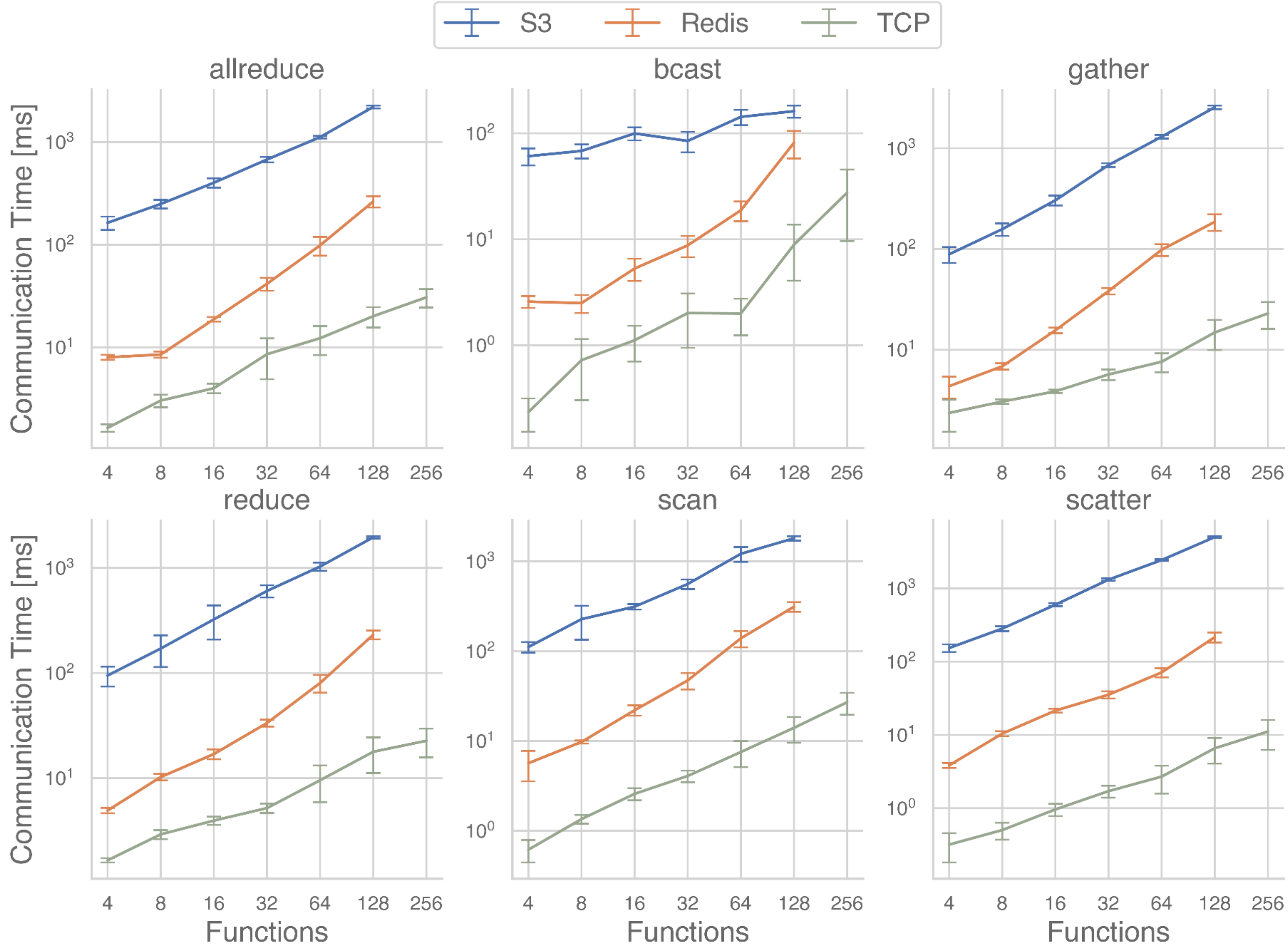
“FMI: Fast and Cheap Message Passing for Serverless Functions”, ICS’23

FMI on AWS Lambda



"FMI: Fast and Cheap Message Passing for Serverless Functions", ICS'23

FMI on AWS Lambda



“FMI: Fast and Cheap Message Passing for Serverless Functions”, ICS’23

FaaS in High-Performance Applications

Serverless is slow

Answer:
rFaaS

Serverless is hard to program.

Communication is slow and restricted

Answer:
FMI

FaaS in High-Performance Applications

Serverless is slow

Answer: rFaaS

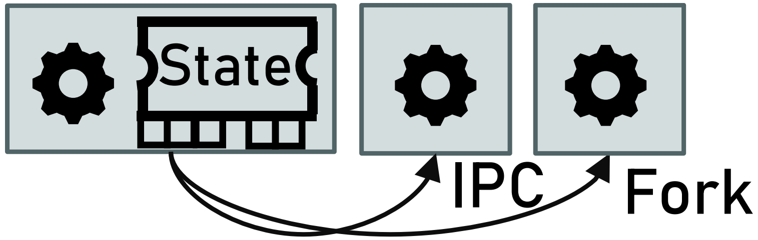
Communication is slow and restricted

Answer: FMI

Serverless is hard to program.

Answer: Serverless Processes

Serverless Process

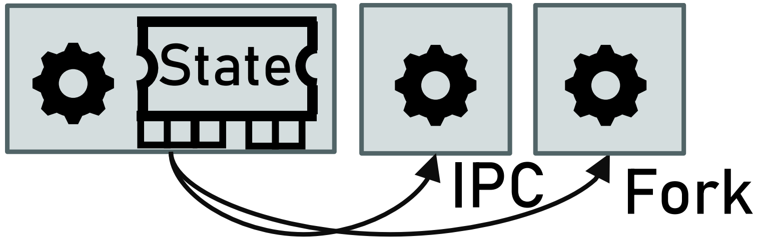


OS Process

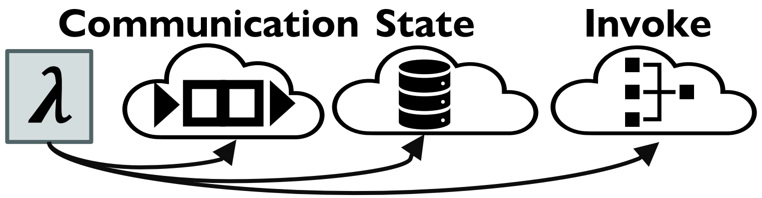
Nano- and micro-second latency of OS primitives.

"Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes", paper preprint.

Serverless Process



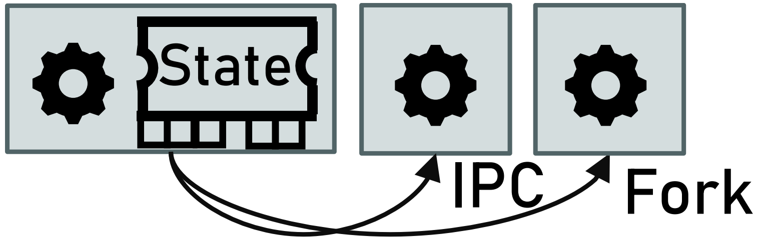
OS Process
Nano- and micro-second latency of OS primitives.



Serverless Function
Millisecond latency of cloud proxies.

"Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes", paper preprint.

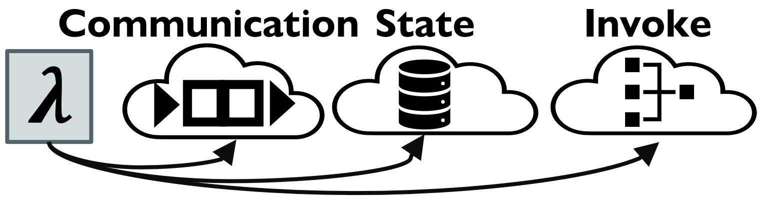
Serverless Process



OS Process

Nano- and micro-second latency of OS primitives.

+

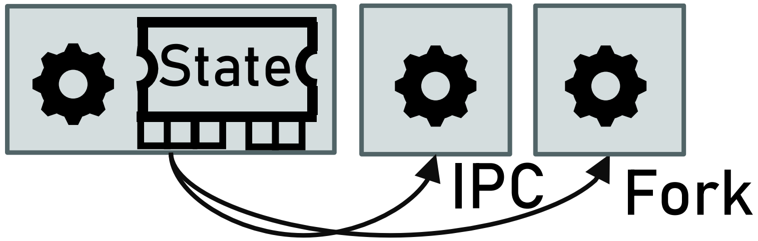


Serverless Function

Millisecond latency of cloud proxies.

"Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes", paper preprint.

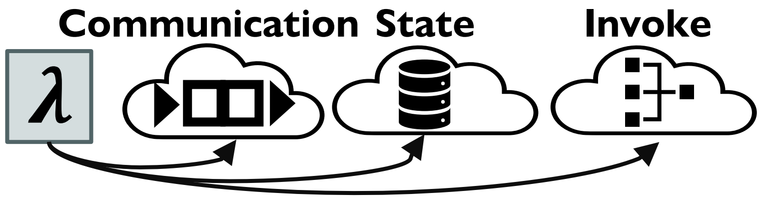
Serverless Process



OS Process

Nano- and micro-second latency of OS primitives.

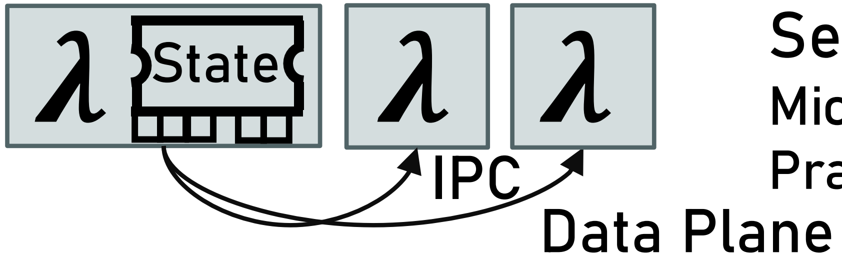
+



Serverless Function

Millisecond latency of cloud proxies.

=

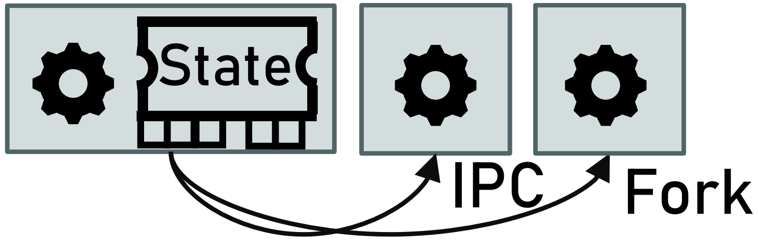


Serverless Process

Microsecond latency of PraaS backend.

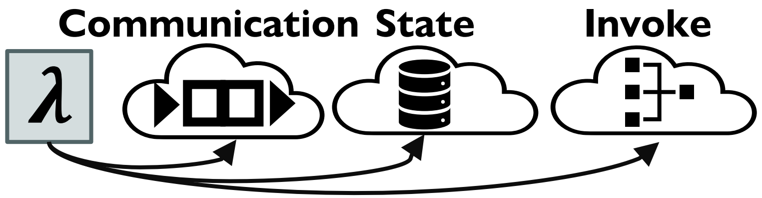
“Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes”, paper preprint.

Serverless Process



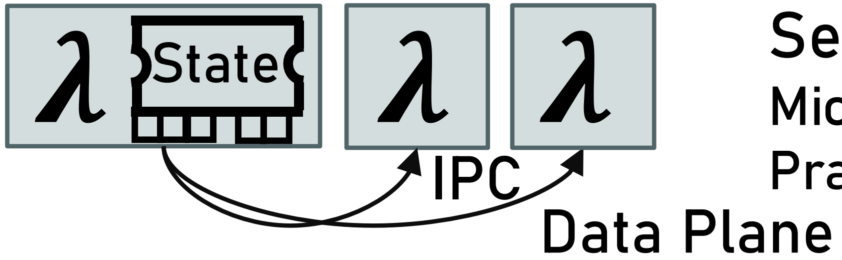
OS Process

Nano- and micro-second latency of OS primitives.



Serverless Function

Millisecond latency of cloud proxies.



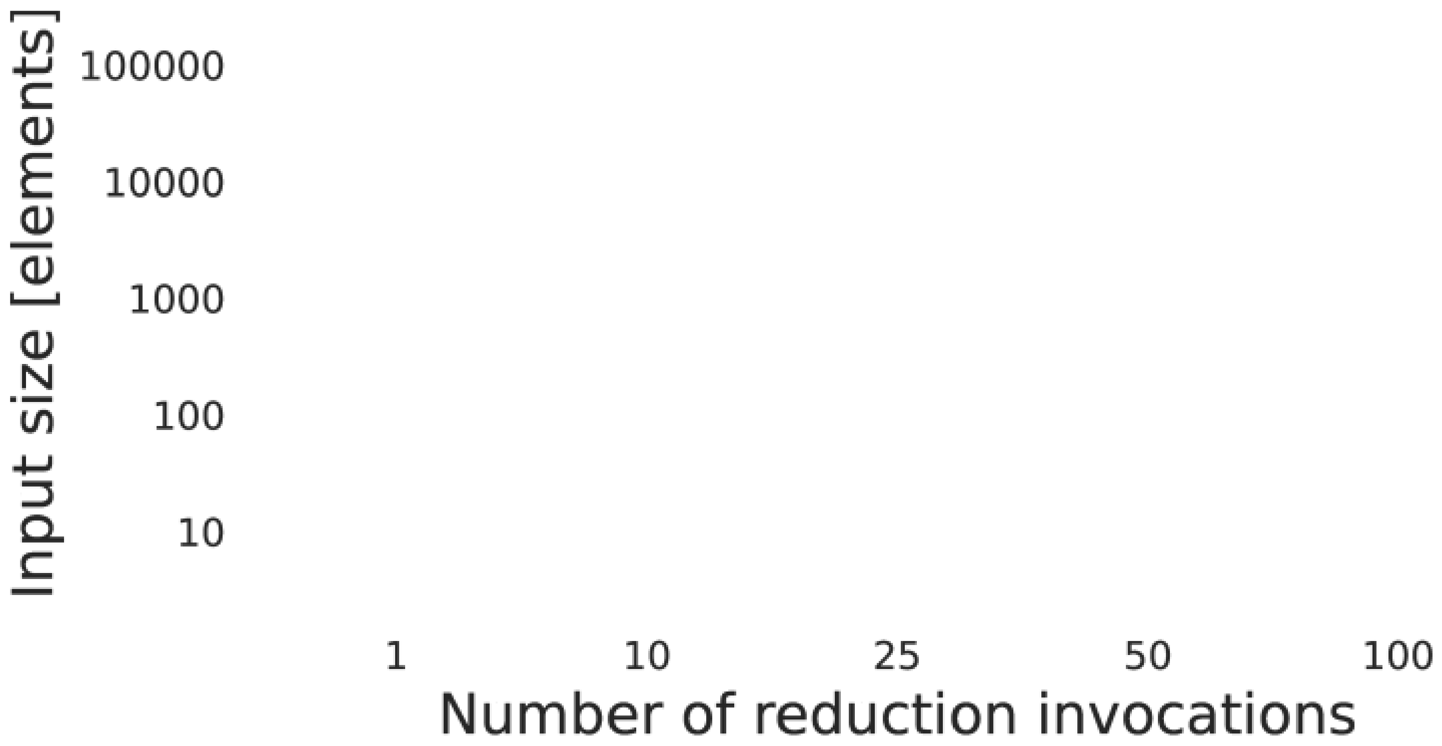
Serverless Process

Microsecond latency of PraaS backend.

Works on AWS Fargate, Knative, Kubernetes.

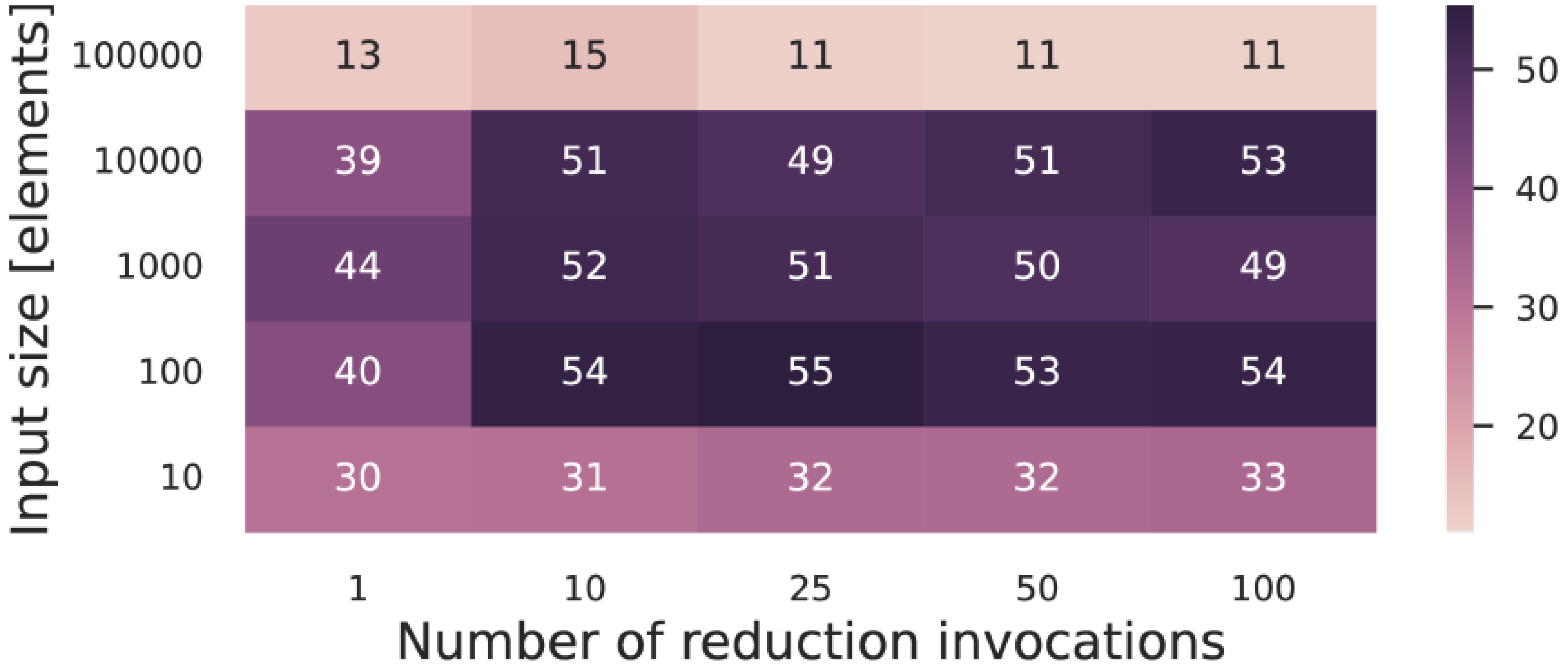
"Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes", paper preprint.

Reduction Benchmark: Process State vs S3



“Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes”, paper preprint.

Reduction Benchmark: Process State vs S3



“Process-as-a-Service: FaaSt Stateful Computing with Optimized Data Planes”, paper preprint.

Serverless Solutions for HPC

Serverless Solutions for HPC

spcl/serverless-benchmarks

Serverless Solutions for HPC

spcl/serverless-benchmarks

spcl/fmi

Serverless Solutions for HPC

spcl/serverless-benchmarks

spcl/fmi

spcl/rFaaS

Serverless Solutions for HPC

spcl/serverless-benchmarks

spcl/fmi

spcl/rFaaS

spcl/PraaS

Conclusions

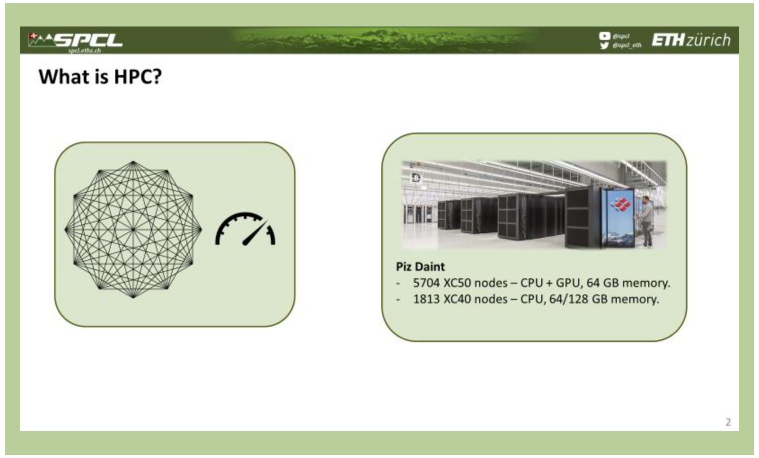
More of SPCL's research:

- youtube.com/@spcl — 150+ Talks
- twitter.com/spcl_eth — 1.2K+ Followers
- github.com/spcl — 2K+ Stars

… or spcl.ethz.ch

This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.

Conclusions



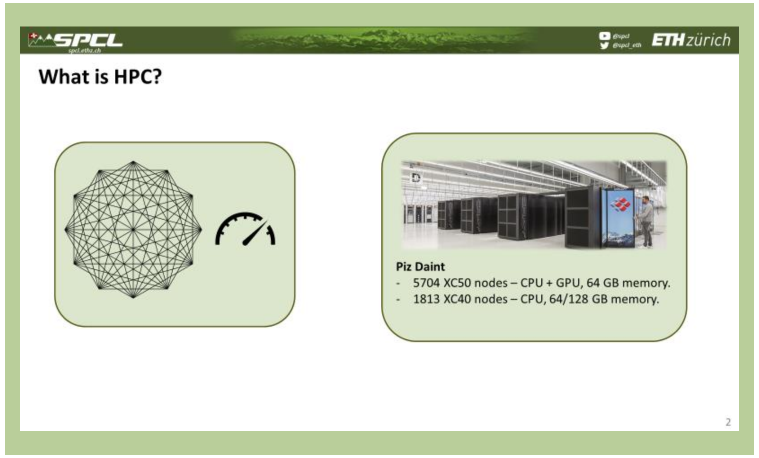
More of SPCL's research:

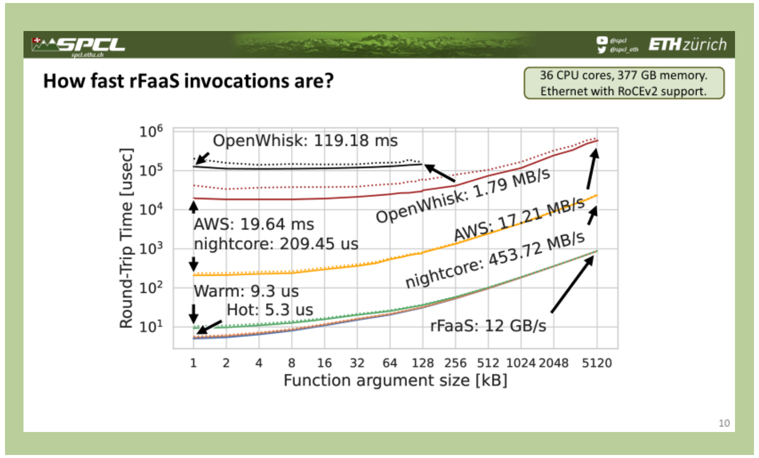
- youtube.com/@spcl — 150+ Talks
- twitter.com/spcl_eth — 1.2K+ Followers
- github.com/spcl — 2K+ Stars

… or spcl.ethz.ch

This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.

Conclusions





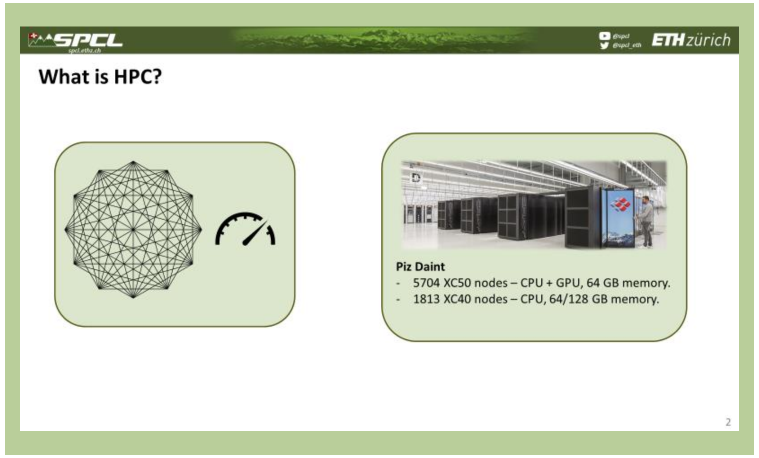
More of SPCL's research:

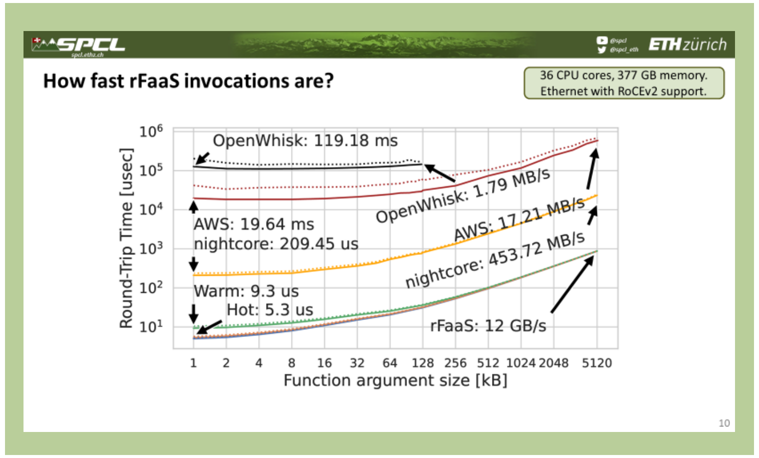
- youtube.com/@spcl — 150+ Talks
- twitter.com/spcl_eth — 1.2K+ Followers
- github.com/spcl — 2K+ Stars

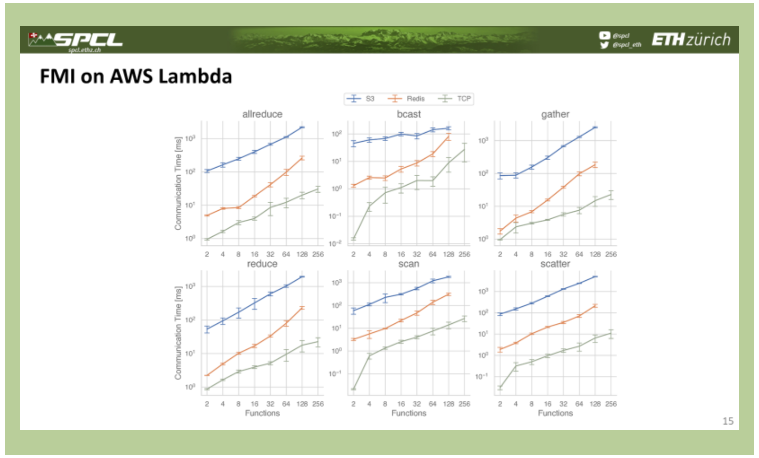
… or spcl.ethz.ch

This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.

Conclusions







More of SPCL's research:

youtube.com/@spcl — 150+ Talks

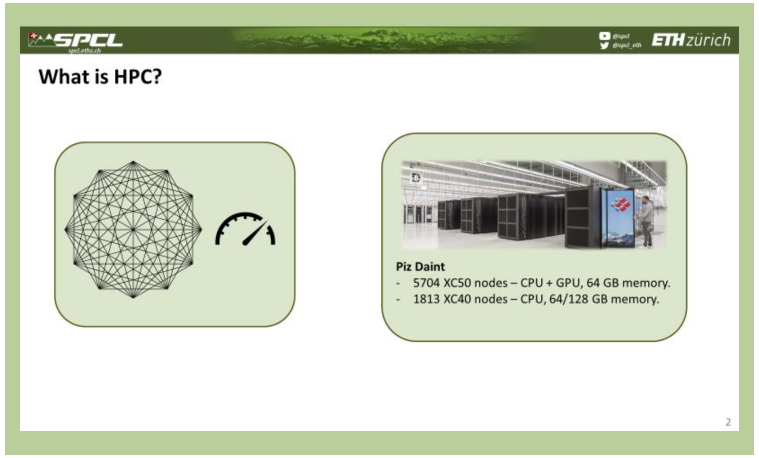
twitter.com/spcl_eth — 1.2K+ Followers

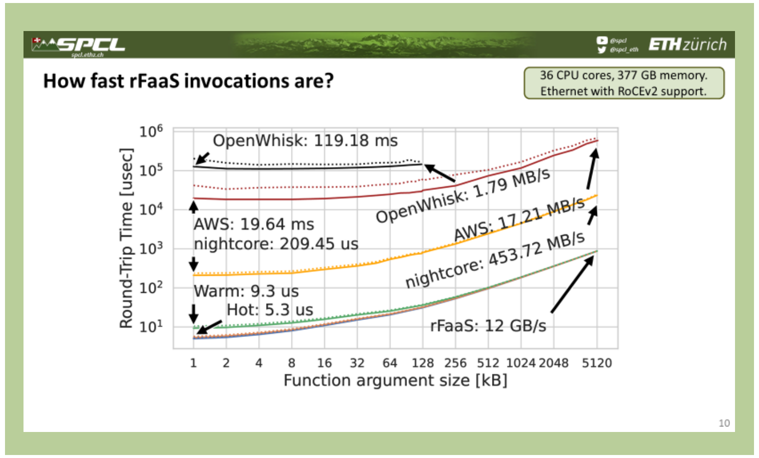
github.com/spcl — 2K+ Stars

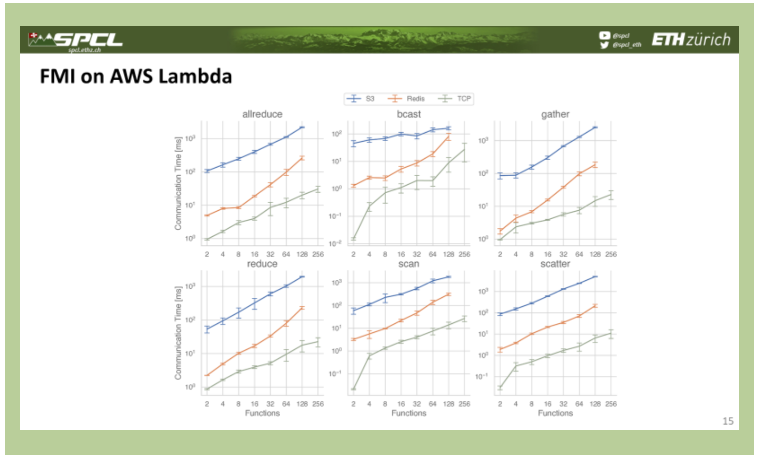
… or spcl.ethz.ch

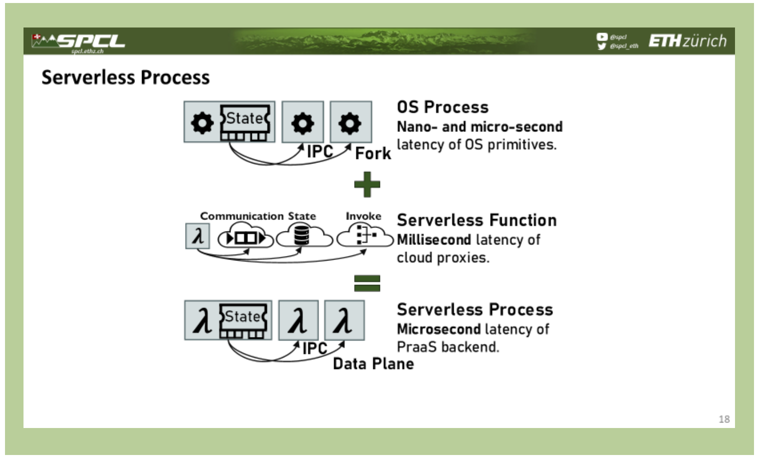
This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.

Conclusions









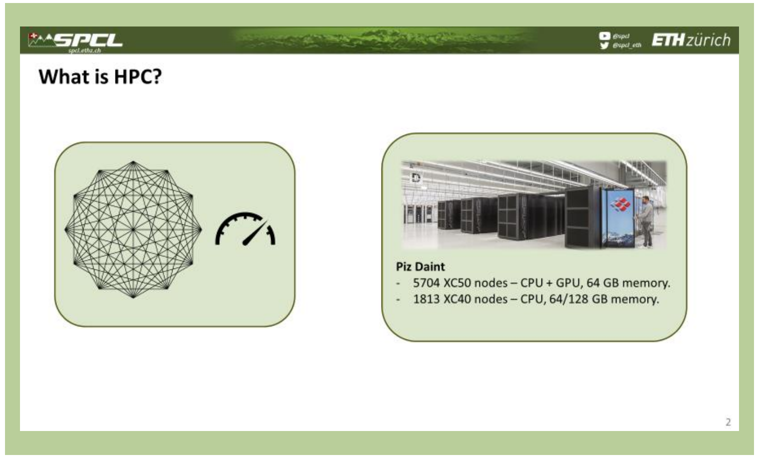
More of SPCL's research:

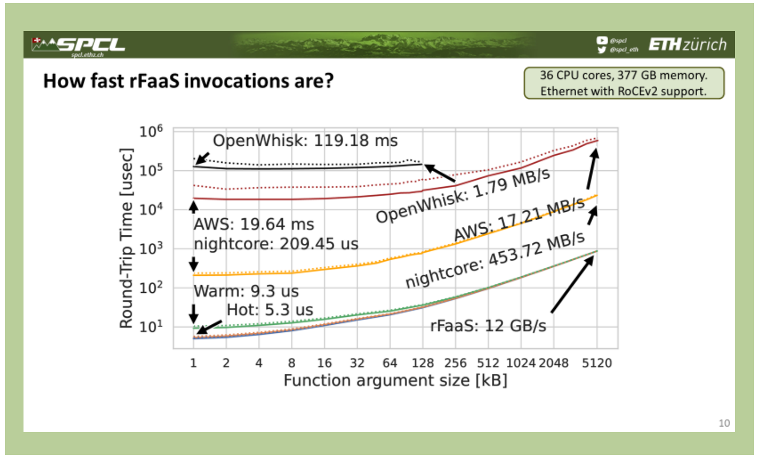
- youtube.com/@spcl — 150+ Talks
- twitter.com/spcl_eth — 1.2K+ Followers
- github.com/spcl — 2K+ Stars

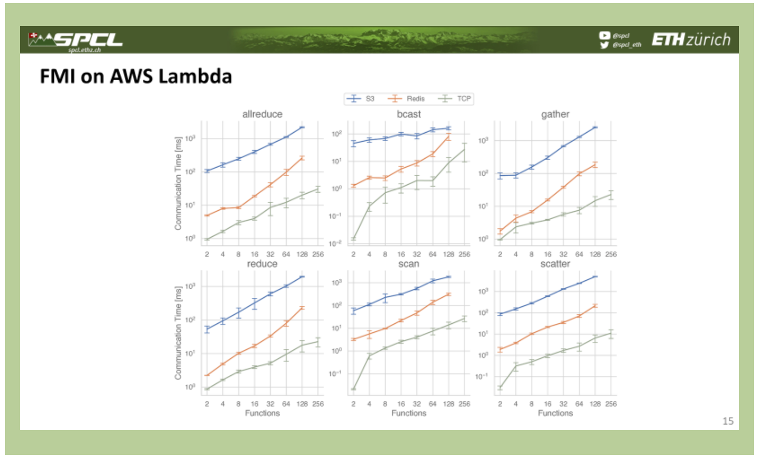
... or spcl.ethz.ch

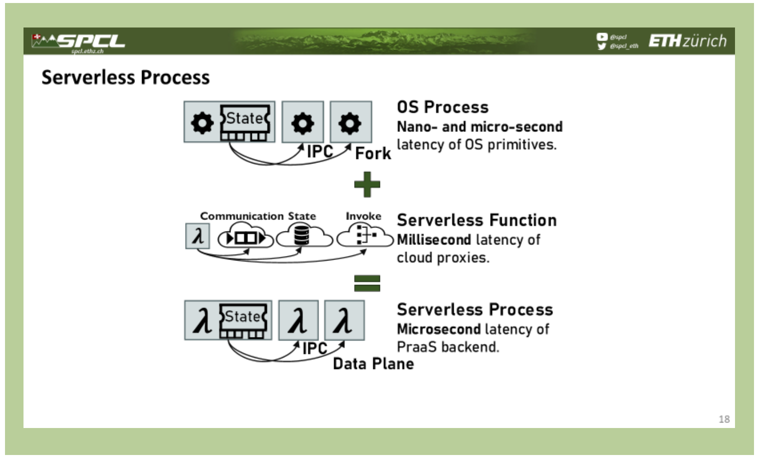
This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.

Conclusions









More of SPCL's research:

- youtube.com/@spcl — 150+ Talks
- twitter.com/spcl_eth — 1.2K+ Followers
- github.com/spcl — 2K+ Stars

… or spcl.ethz.ch

Poster — **Personal website**

This work has received funding from the European Research Council (ERC), Swiss National Science Foundation (SNF), and from Amazon Web Services through the AWS Cloud Credits for Research program.