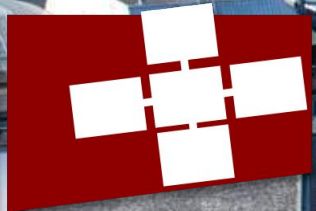


MARCIN COPIK, ALEXANDRU CALOTOIU, GYORGY RETHY, ROMAN BÖHRINGER, RODRIGO BRUNO, TORSTEN HOEFLE

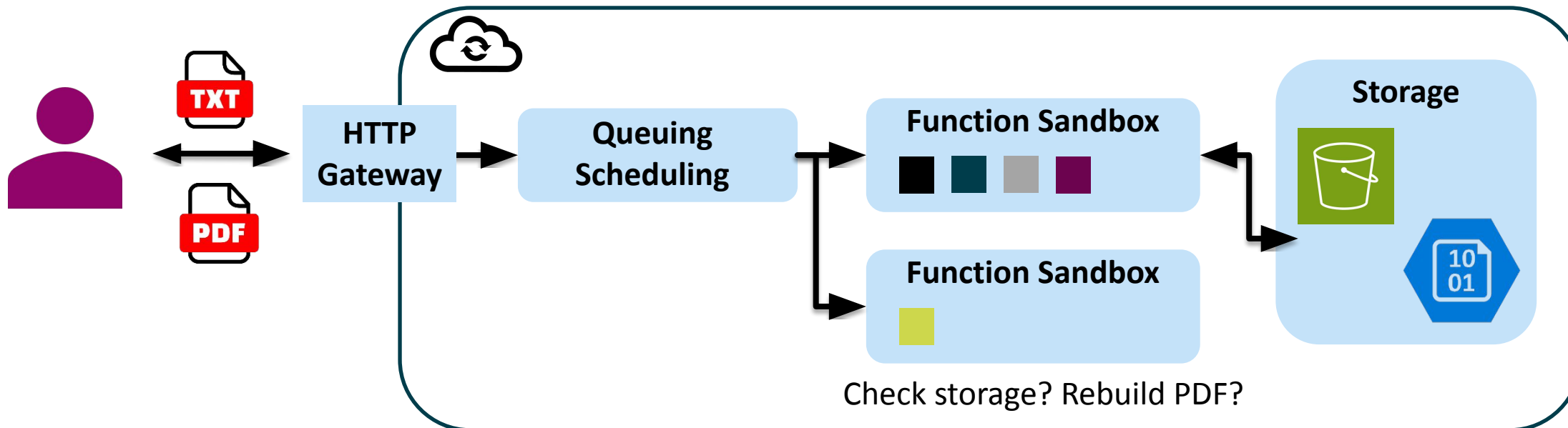
# Process-as-a-Service: Unifying Elastic and Stateful Clouds with Serverless Processes





# LaTeX Microservice in Serverless

Is cache up to date?  
Did other function edit file of this client?



## Serverless Service

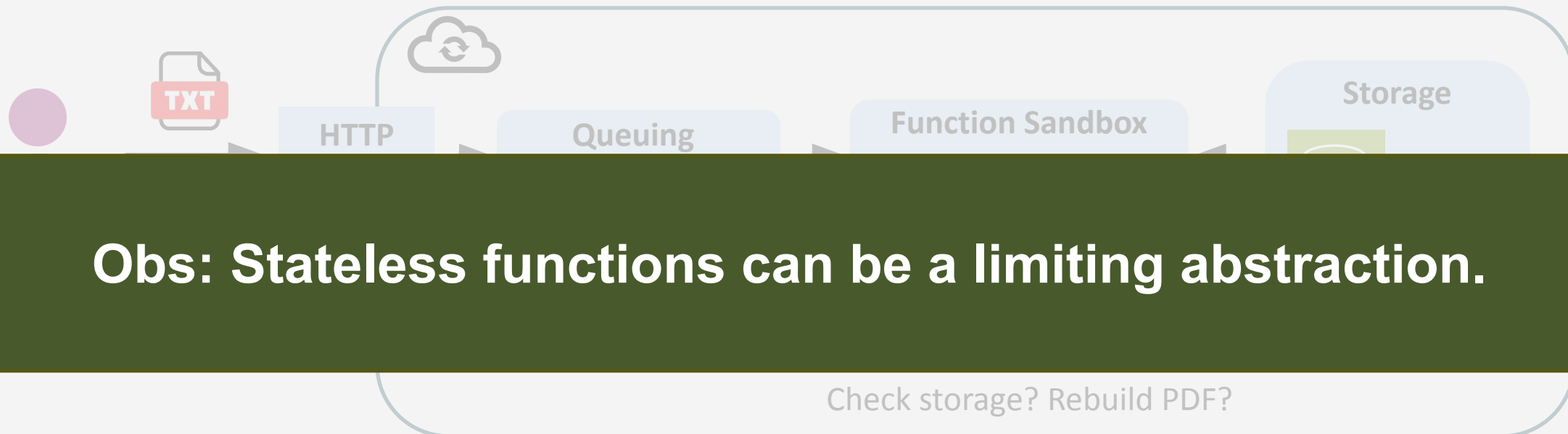
- ✗ Cached function state is not reliable
- ✗ Request context has no influence on scheduling

## Standard Microservice

- ✓ Persistent caching
- ✓ Load balancing with sticky sessions

# LaTeX Microservice in Serverless

Is cache up to date?  
Did other function edit file of this client?



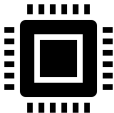
## Serverless Service

- ✗ Cached function state is not reliable
- ✗ Request context has no influence on scheduling

## Standard Microservice

- ✓ Extensive caching
- ✓ Load balancing with sticky sessions

# Cloud Service Models

|                                                                                     | IaaS                 | CaaS                                  | FaaS                     |
|-------------------------------------------------------------------------------------|----------------------|---------------------------------------|--------------------------|
|    | Persistent           | Persistent                            | Ephemeral                |
|    | Provisioned          | Provisioned                           | Pay-as-you-go            |
|    | Manual,<br>minutes   | Semi-automatic,<br>seconds to minutes | Automatic,<br>ms to secs |
|  | Persistent,<br>local | Persistent,<br>local                  | Ephemeral,<br>remote     |

# Cloud Service Models

IaaS

CaaS

FaaS



Persistent

Persistent

Ephemeral

Q: Is persistent state incompatible with ephemeral resources?



Manual,  
minutes

Semi-automatic,  
minutes & seconds

Automatic,  
seconds

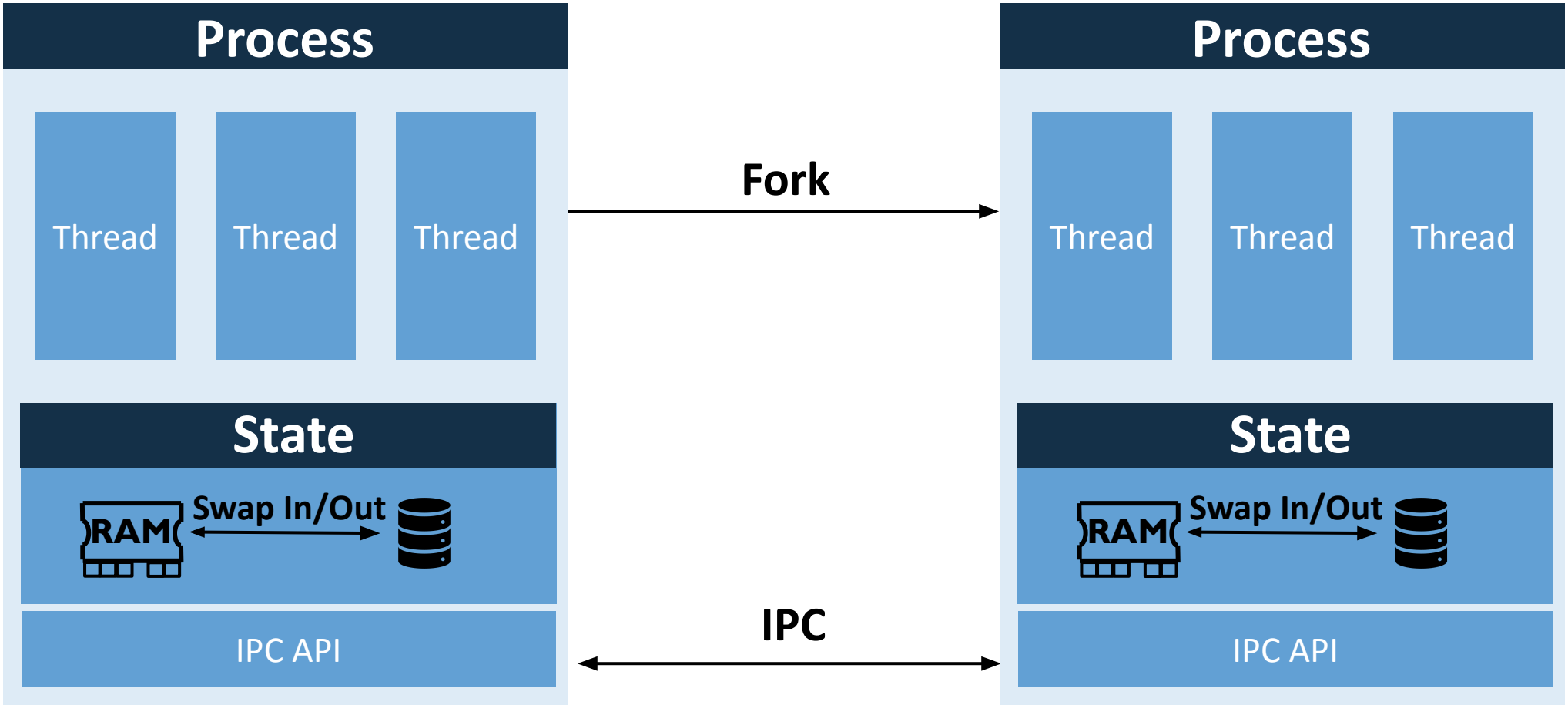


Persistent,  
local

Persistent,  
local

Ephemeral,  
remote

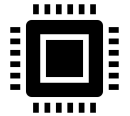
# POSIX Process Model



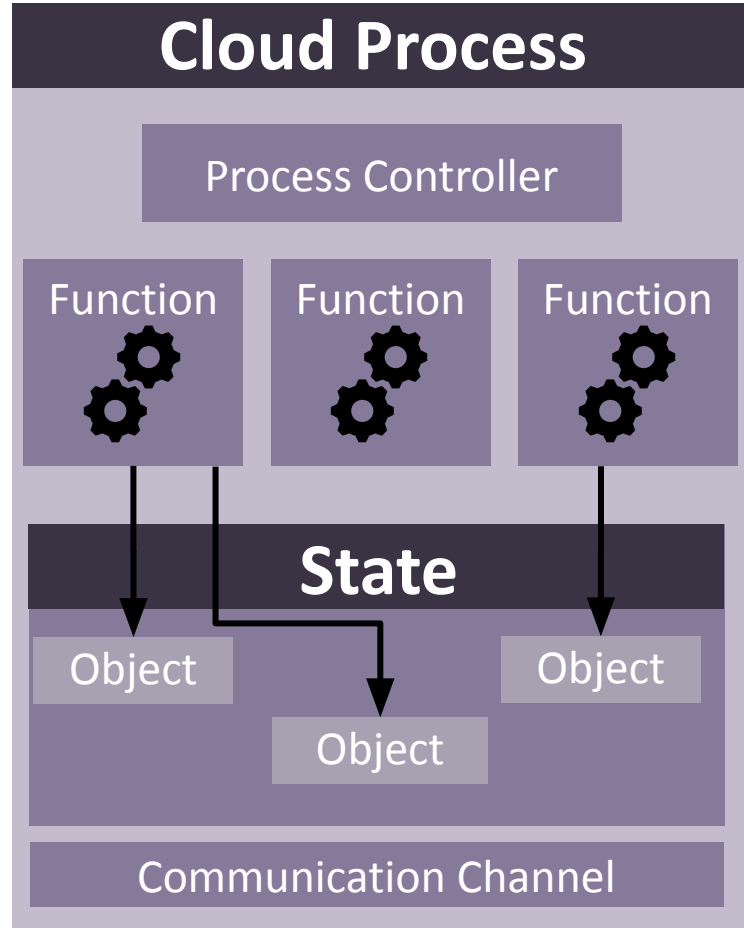
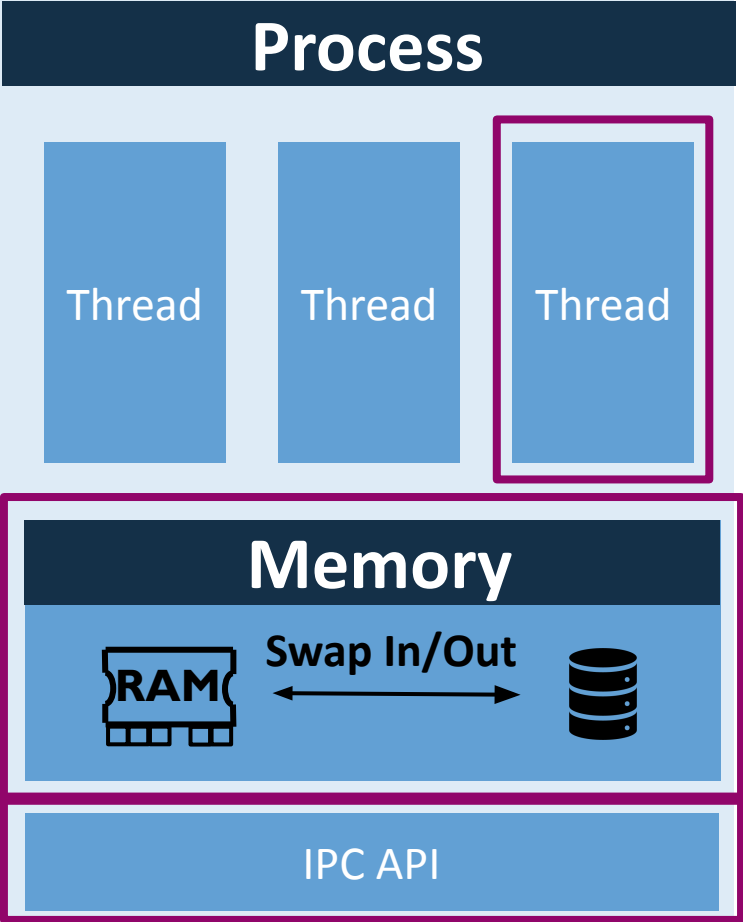
**Ephemeral resources may be preempted but state is not lost!**

# Cloud Process

Durable state of a container and **ephemeral compute** of serverless.

|                                                                                     | IaaS                 | CaaS                                 | <u>PraaS</u>          | FaaS                  |
|-------------------------------------------------------------------------------------|----------------------|--------------------------------------|-----------------------|-----------------------|
|    | Persistent           | Persistent                           | Ephemeral             | Ephemeral             |
|    | Provisioned          | Provisioned                          | Pay-as-you-go         | Pay-as-you-go         |
|    | Manual,<br>minutes   | Semi-automatic,<br>minutes & seconds | Automatic,<br>seconds | Automatic,<br>seconds |
|  | Persistent,<br>local | Persistent,<br>local                 | Persistent,<br>local  | Ephemeral,<br>remote  |

# Cloud Process Model

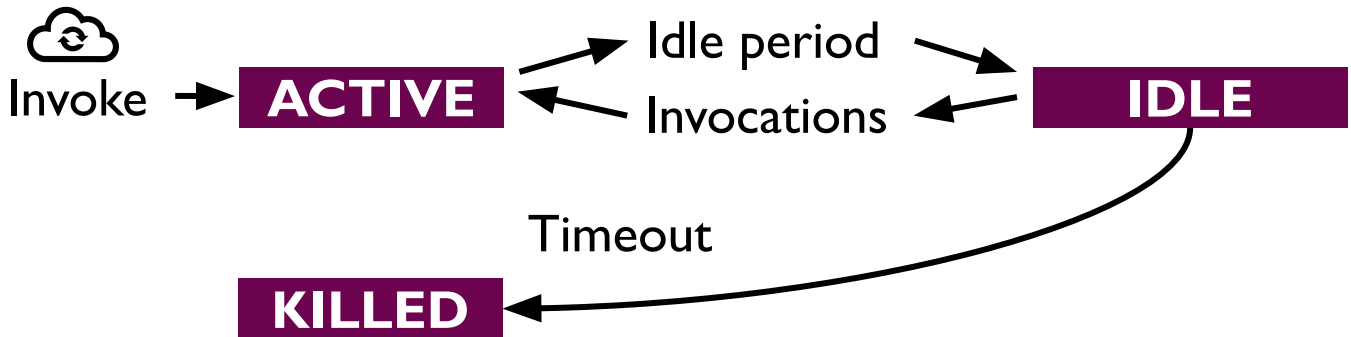


State is represented as a pool for actual data in the process.



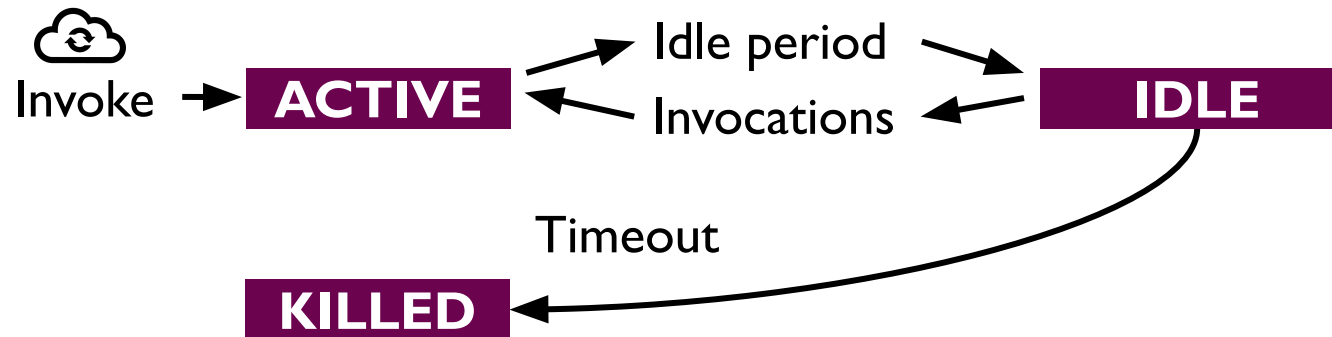
# Serverless Instance Lifecycle

FaaS

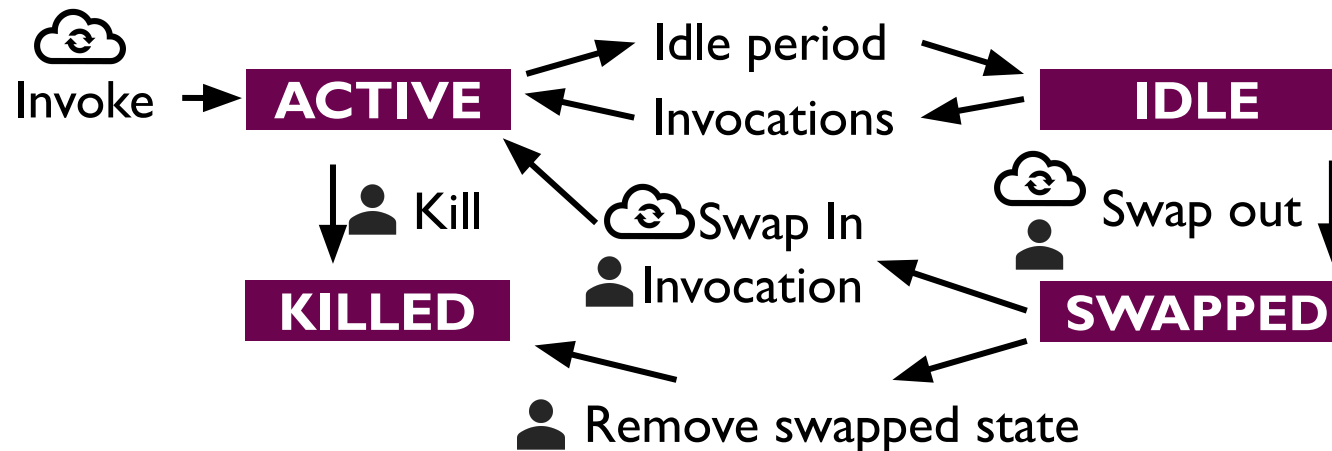


# Serverless Instance Lifecycle

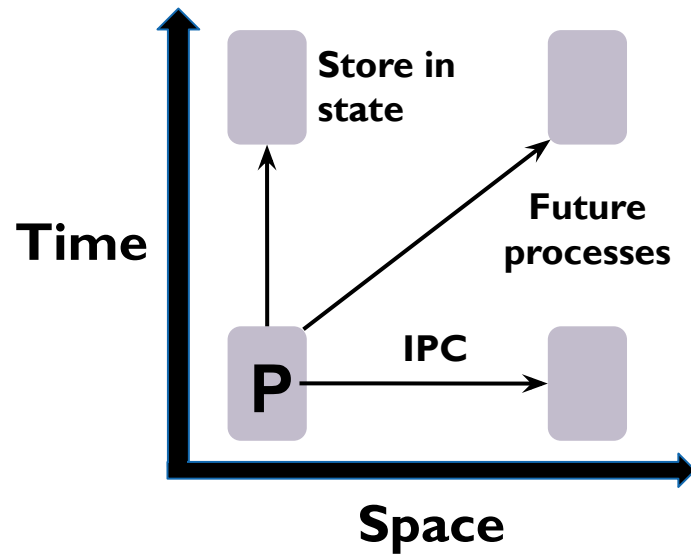
## FaaS



## PraaS



# PraaS API



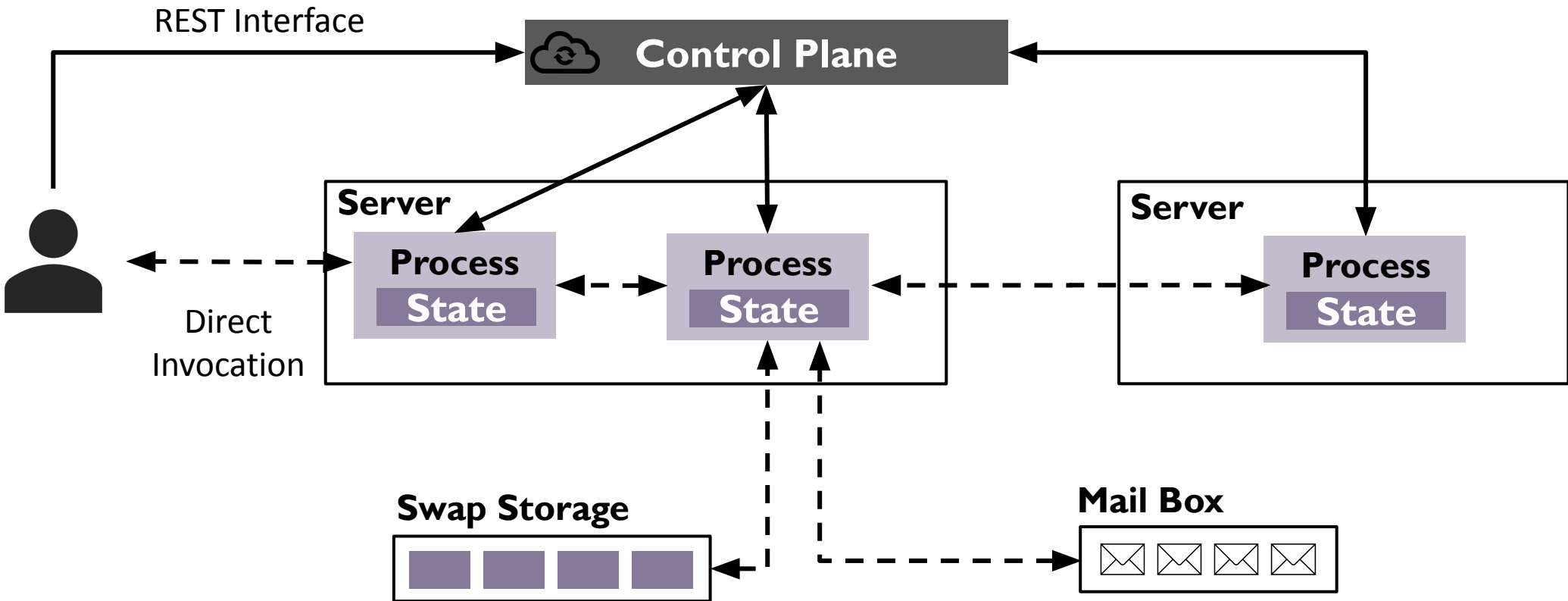
```
invoke([pid_destination], function, data)
```

```
state(key, data)  
data = state(key)
```

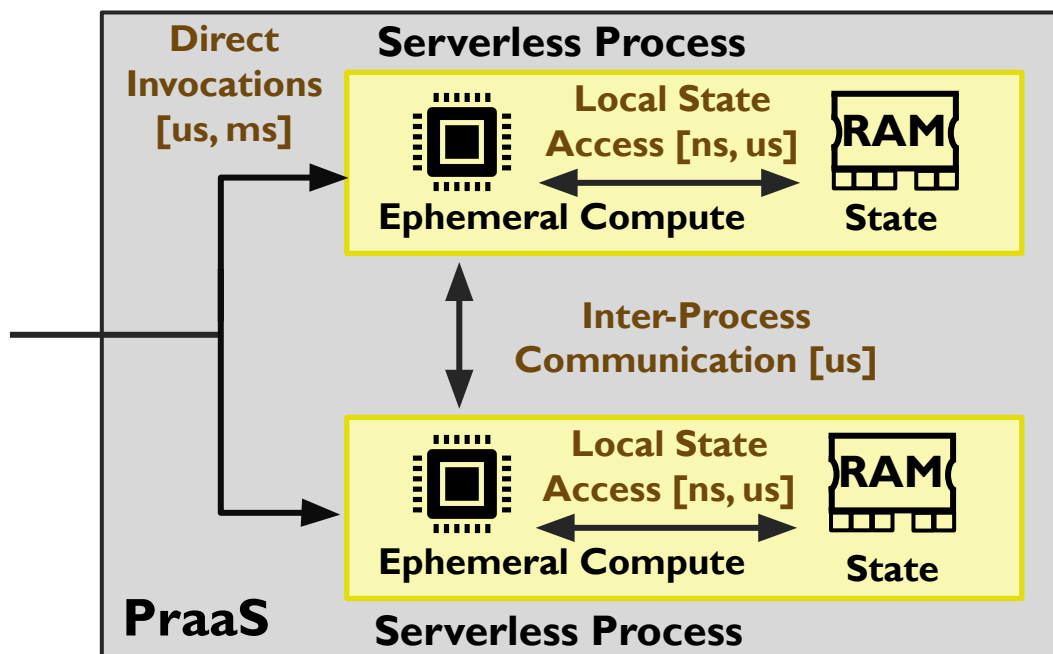
```
send(pid_destination, key, data)  
data = recv([pid_source], key)
```

**PraaS generalizes over FaaS and captures stateful services.**

# PraaS Platform Architecture



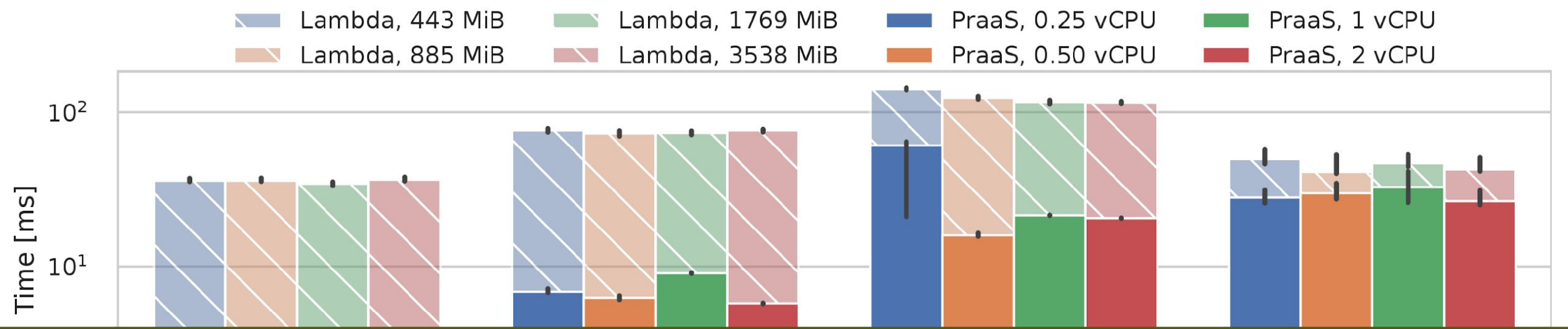
# PraaS in Practise



- Applications can use multiple processes
- Each process can host multiple functions
- Intra-process comm. over POSIX queues
- Inter-process is provider defined
  - (TCP in our prototypes)
- Two prototypes:
  - AWS Fargate
  - Knative (Kubernetes)



# Benchmark: LaTeX Microservice



**LaTeX files can be served up to 34.3x faster.**

get-file, Input 1

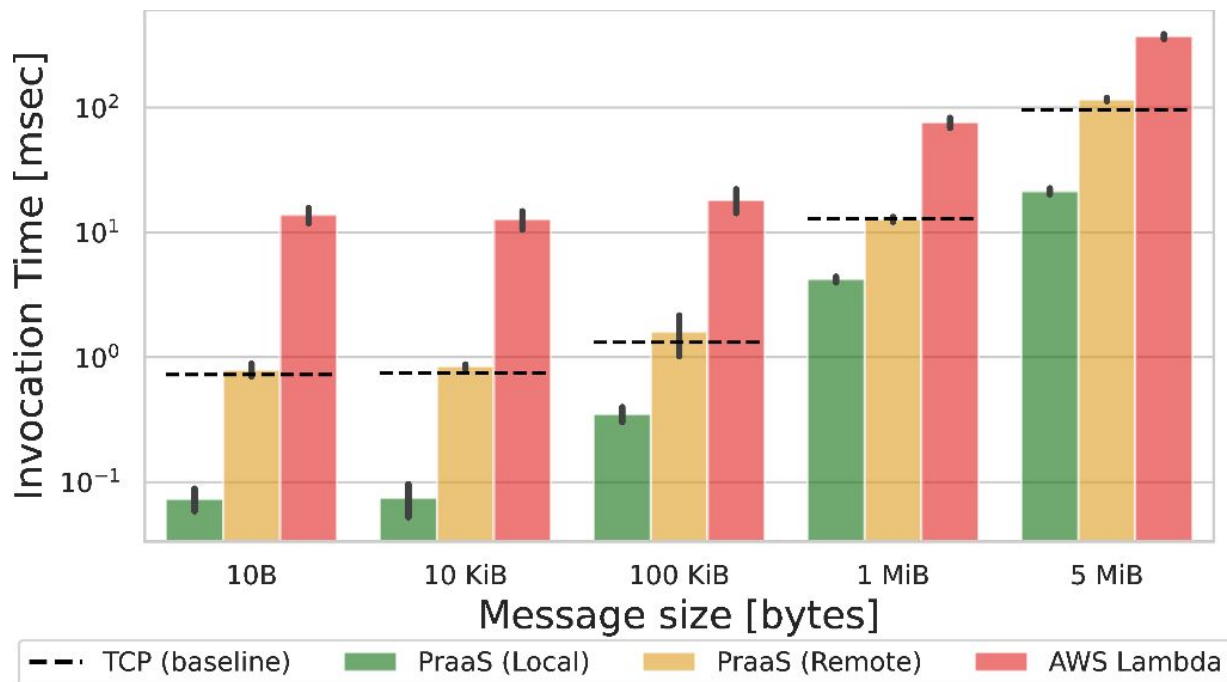
get-file, Input 4

get-pdf

update-file, Input 1

# Microbenchmarks Evaluation

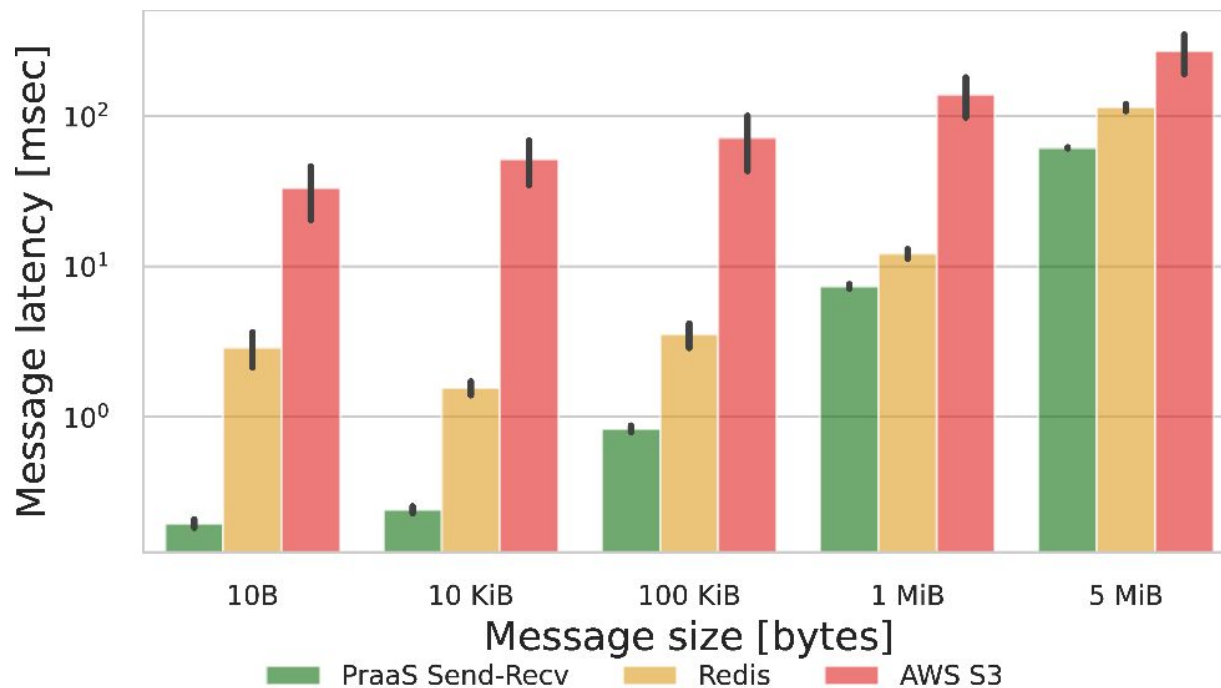
## Data Plane Invocations



**PraaS matches PaaS TCP for a no-op function.**

# Microbenchmarks Evaluation

## Inter-Process Communication



**Reduced latency compared to S3 (77%-99%) and Redis (39%-93%).**



**Marcin Copik**  
ETH Zurich



**Alexandru Calotoiu**  
ETH Zurich



**Gyorgy Rethy**  
Oracle Labs



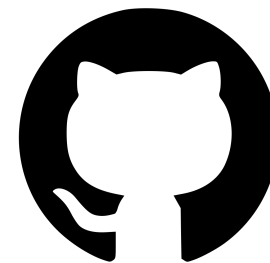
**Roman Böhringer**  
OpenCore GmbH



**Rodrigo Bruno**  
Univ. Lisbon & INESC-ID



**Torsten Hoefler**  
ETH Zurich



**spcl/PraaS**



**Paper**



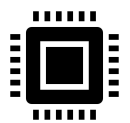
# Paper

## IaaS

## CaaS

## PraaS

## FaaS



Persistent

Persistent

Ephemeral

Ephemeral



Provisioned

Provisioned

Pay-as-you-go

Pay-as-you-go



Manual,  
minutes

Semi-automatic,  
minutes & seconds

Automatic,  
seconds

Automatic,  
seconds



Persistent,  
local

Persistent,  
local

Persistent,  
local

Ephemeral,  
remote

Thank You!