KOLLAPS

Kollaps/Thunderstorm: Reproducible Evaluation of Distributed Systems

Miguel Matos

U. Lisbon & INESC-ID Portugal

Tutorial @ DISCOTEC/DAIS 2020

OUTLINE

- Overview of Kollaps & Thunderstorm
- Hands-on tutorial: basics
- Hands-on tutorial: advanced features

MOTIVATION

Amazon Found Every 100ms of Latency Cost them 1% in Sales

Nov 10, 2016, 11:43am EST

Why Brands Are Fighting Over Milliseconds

Video news

Buffering reduces video watch time by ~40%, according to research

September 14, 2016 (4 years ago)

Post Mortem: What Yesterday's Network Outage Looked Like

Zalando saw a 0.7% increase in revenue when they shaved 100ms off their load time.

MOTIVATION

- Performance depends heavily on underlying network
- **Variability** and **Failures** are the norm

- Need for tools for systematic evaluation of distributed applications

- Ability to answer key questions:
  - What is the impact of halving the network latency in application throughput?
  - What is the effect of packet loss?
  - What if …

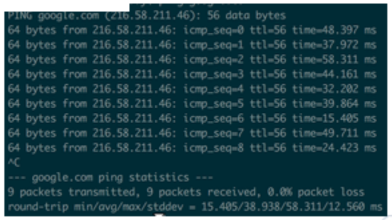
RELATED WORK

Name	Year	Mode	HW ind.	Orchestration	Concurrent deployments	Path congestion	Link-Level emulation capabilities				Any Language	Topology dynamics	Unit
							Bandwidth	Delay	Packet loss	Jitter			
DelayLine [47]	1994	User	✓	Centralized	✗	✗	✗	✓	✓	✗	✓	✗	P
ModelNet [81]	2002	Kernel	✓	Centralized	✗	✓	✓ϟ	✓ϟ	✓ϟ	✗	✓	✓	P
Nist NET [33]	2003	Kernel	✓	Centralized	✗	✗	✓ϟ	✓ϟ	✓ϟ	✓ϟ	✓	✗	P
NetEm [45]	2005	Kernel	✓	*(N/A: single link emulation only)*			✗	✓	✓	✓	✓	✗	P
Trickle [39]	2005	User	✓	*(N/A: single link emulation only)*			✓	✓	✗	✗	✓	✗	P
EmuSocket [23]	2006	User	✓	*(N/A: single lin* [illegible]			[illegible]	✓ϟ	✗	✗	✓	✗	P
ACIM/FlexLab [71]	2007	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✓ϟ	✓	✓	V
NCTUns [85]	2007	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓	✓	✓	✓	✗	P
Emulab [46, 88]	2008	Kernel	✗	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✗	✓	✓	V
IMUNES [70]	2008	Kernel	✗	Centralized	[illegible]	[illegible]	[illegible]	✓	✓	✗	✓	✗	P
MyP2P-World [75]	2008	User	✓	Centralized	[illegible]	[illegible]	[illegible]	✓	✓	✗	✗	✗	P
P2PLab [61]	2008	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓	✓	✗	✗	✗	P
Netkit [67]	2008	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✗	✓	✗	V
DFS [79]	2009	User	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓	✓	✗	✓	P
Dummynet [32]	2010	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✗	✓	✗	P
Mininet [53]	2010	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✓ϟ	✓	✓	P
SliceTime [86]	2011	Kernel	✗	Centralized	[illegible]	[illegible]	[illegible]	✓	✗	✗	✓	✓	V
Mininet-HiFi [44]	2012	Kernel	✓	Centralized	[illegible]	[illegible]	[illegible]	✓ϟ	✓ϟ	✓ϟ	✓	✓	C
SplayNet [76]	2013	User	✓	Decentralized	✓	✓	✓	✓	✓	✗	✗	✓	P
MaxiNet [87]	2014	Kernel	✓	Centralized	✗	✓	✓ϟ	✓ϟ	✓ϟ	✓ϟ	✓	✓	P
Dockemu [80]	2015	User	✓	Centralized	✗	✗	✓	✓	✓	✓	✓	✗	C
EvalBox [77]	2015	Kernel	✓	Centralized	✗	✗	✓ϟ	✓ϟ	✓ϟ	✓ϟ	✓	✓	P
ContainerNet [65]	2016	Kernel	✓	Centralized	✗	✓	✓ϟ	✓ϟ	✓ϟ	✓ϟ	✓	✓	C,V
Kathará [30]	2018	Kernel	✓	Centralized	✗	✓	✓ϟ	✓ϟ	✓ϟ	✗	✓	✗	C
Kollaps	**2020**	**Kernel**	✓	**Decentralized**	✓	✓	✓ϟ	✓ϟ	✓ϟ	✓ϟ	✓	✓	**C,V**

Main limitations:
- scalability
- accuracy
- dynamics

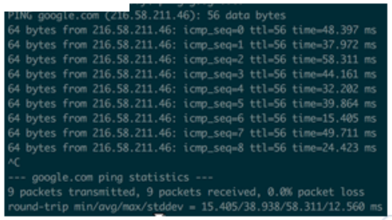
KOLLAPS IN A NUTSHELL

- Applications are concerned about the end-to-end properties:
  - latency, jitter, bandwidth, packet loss
- Rather than the internal network state leading to those properties



KOLLAPS IN A NUTSHELL

- Applications are concerned about the end-to-end properties:
  - latency, jitter, bandwidth, packet loss
- Rather than the internal network state leading to those properties

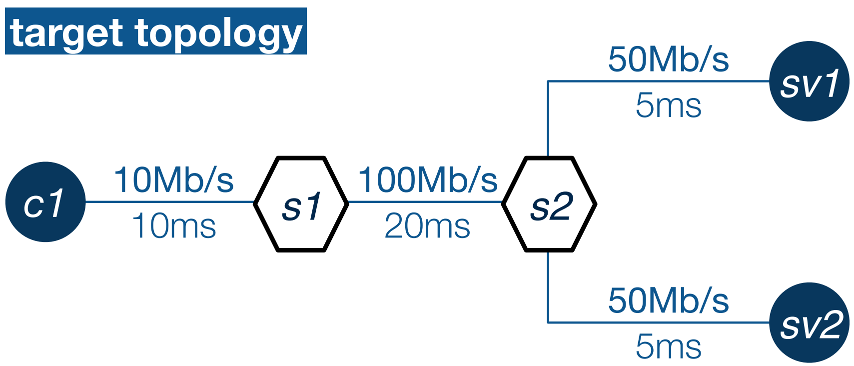


- Emulate **only** end-to-end properties
- Allows decentralized highly scalable emulation

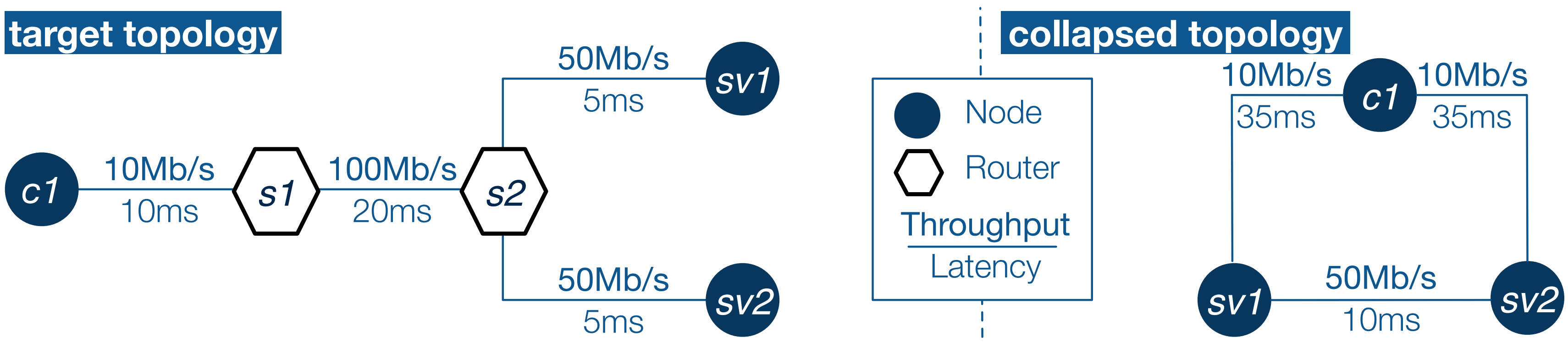
ANGAINOR

inesc id lisboa

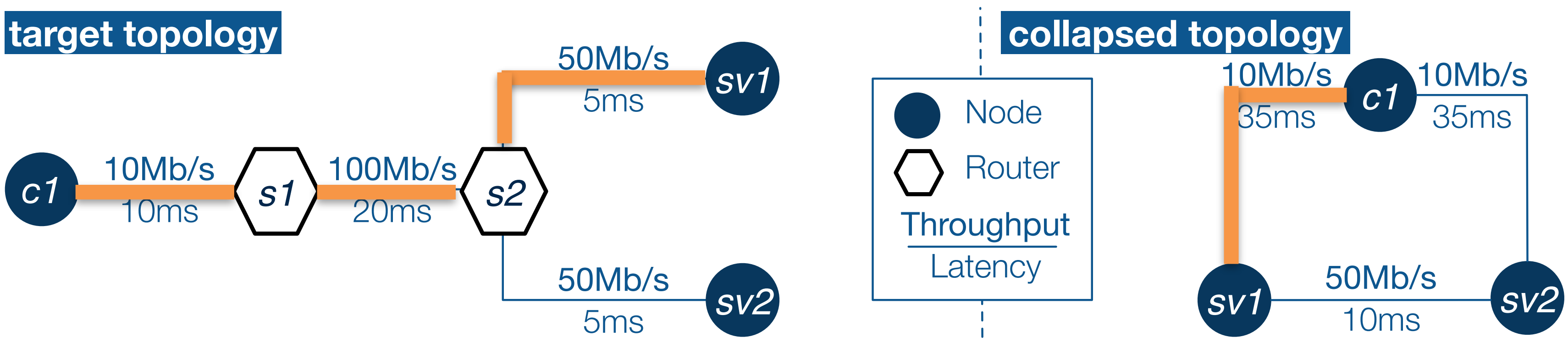
NETWORK COLLAPSING



NETWORK COLLAPSING

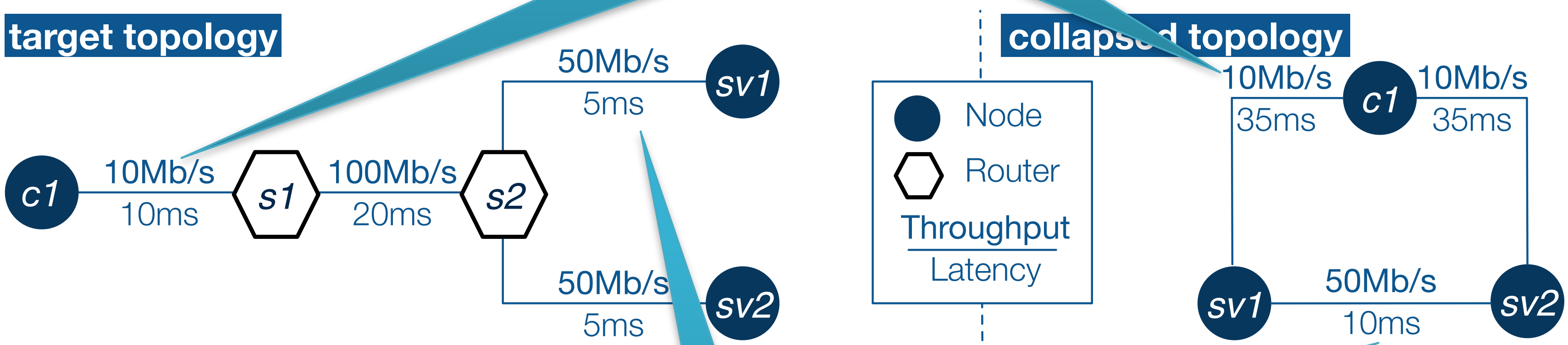


NETWORK COLLAPSING



NETWORK COLLAPSING

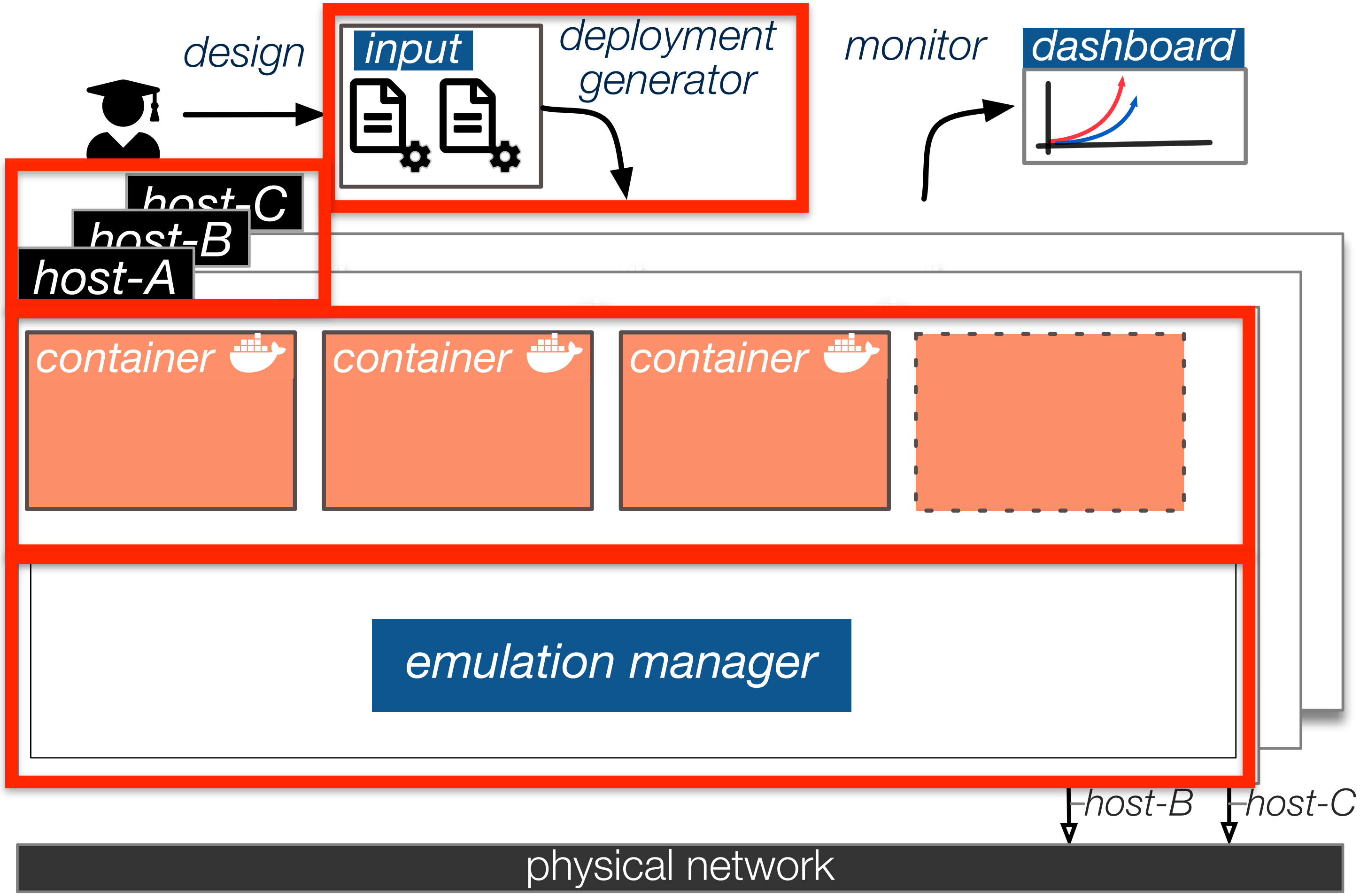
Minimum bandwidth of all links



Pre-computation of static properties

Sum of latencies of all links

ARCHITECTURE

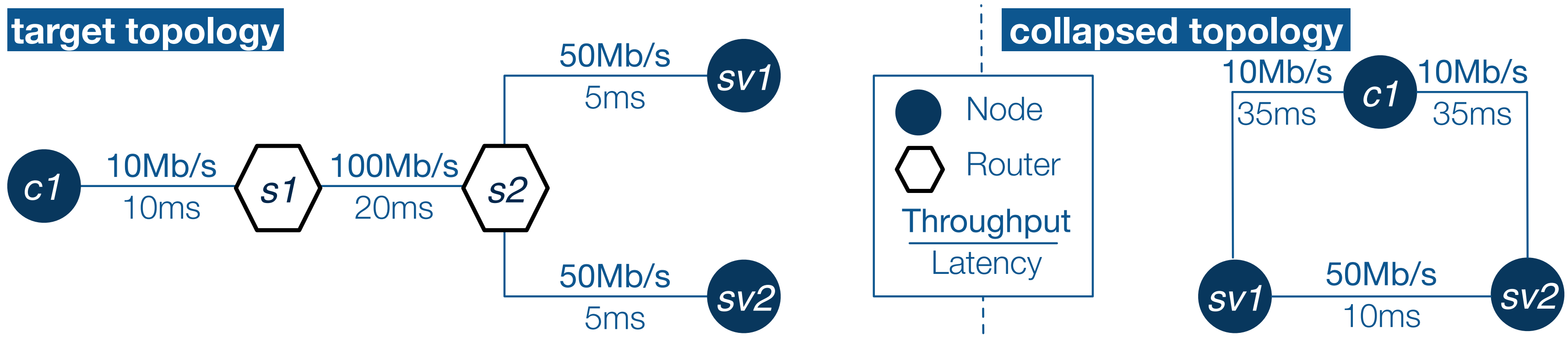


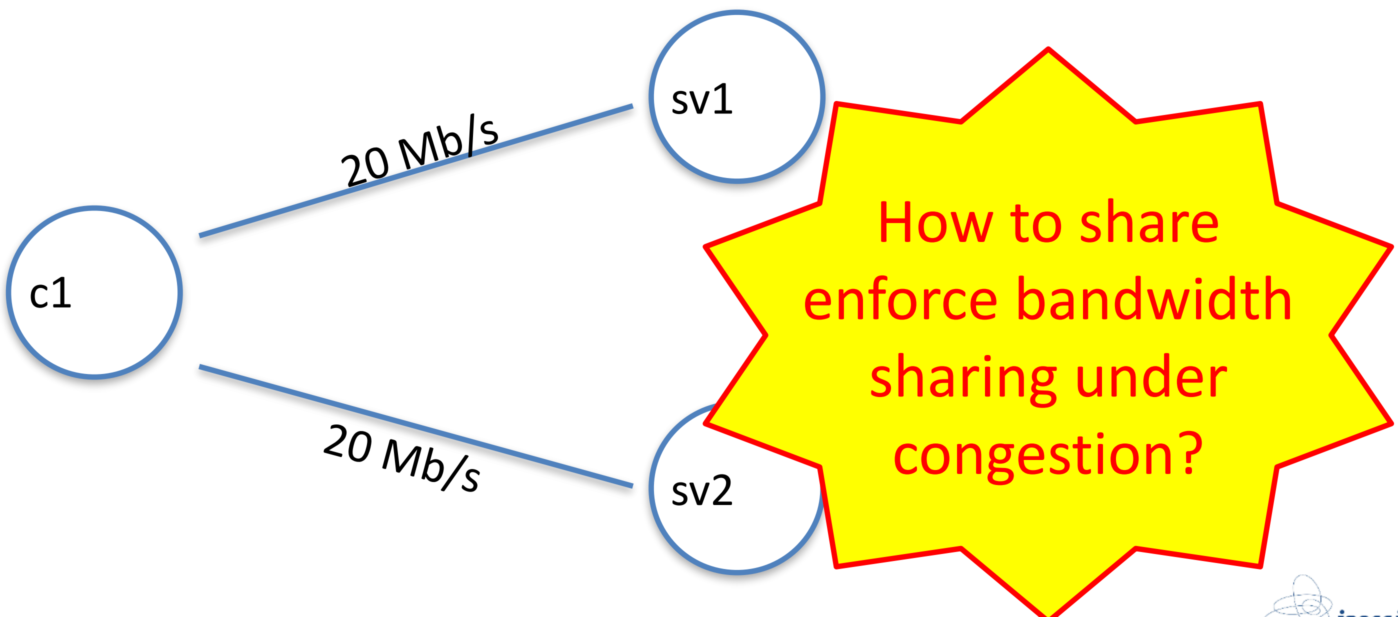
EMULATION MANAGER (EM)

- One instance per physical machine

- Enforces topology properties
  - static properties
  - dynamic properties

EM: DYNAMIC PROPERTIES





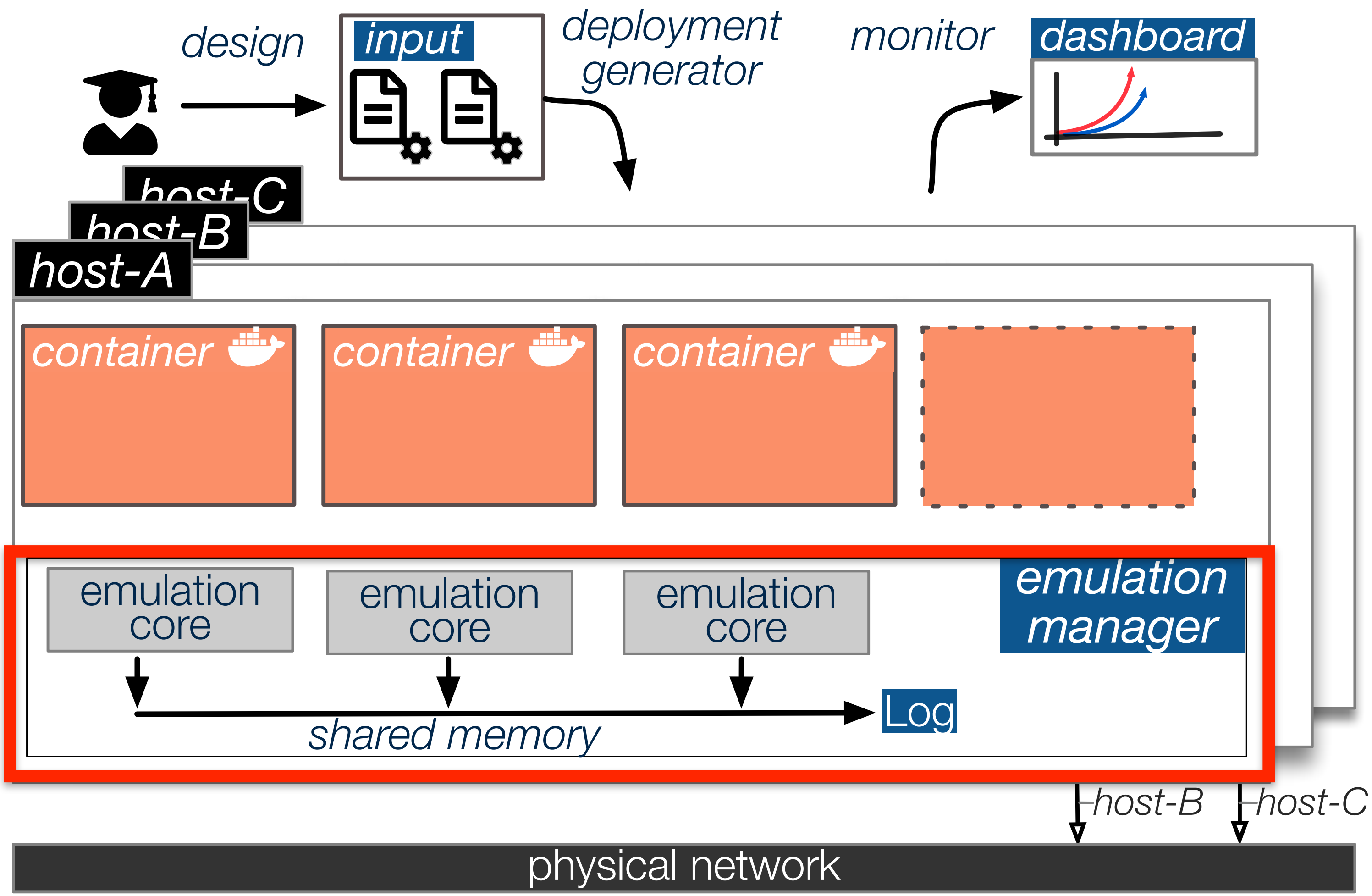
EM: DYNAMIC PROPERTIES

- RTT-Aware Min-Max model:

$$Share(f) = \left(RTT(f) \sum_{i=1}^{n} \frac{1}{RTT(f_i)} \right)^{-1}$$

- Intuition
  - Available bandwidth is inversely proportional to the RTT

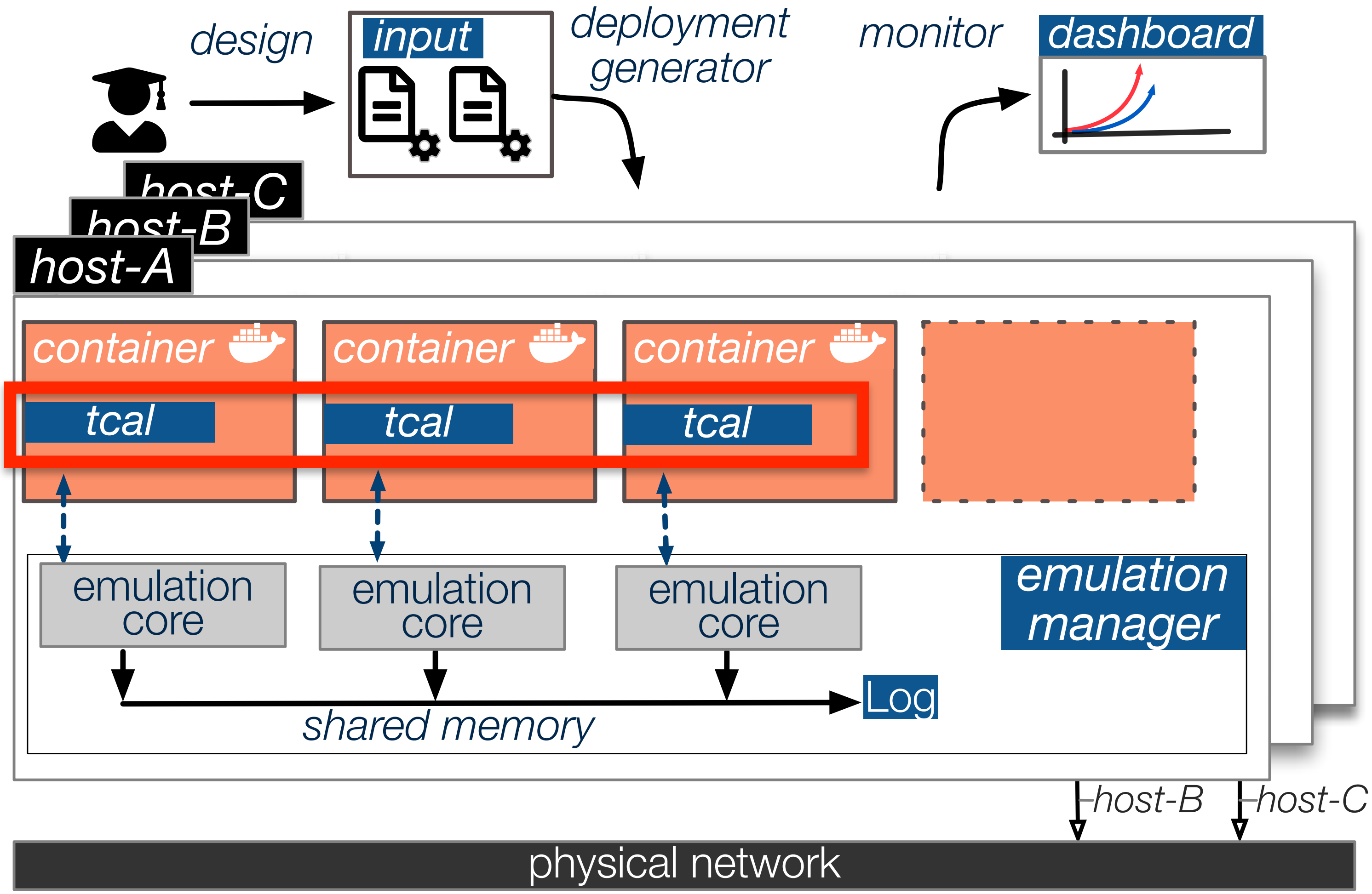
ARCHITECTURE



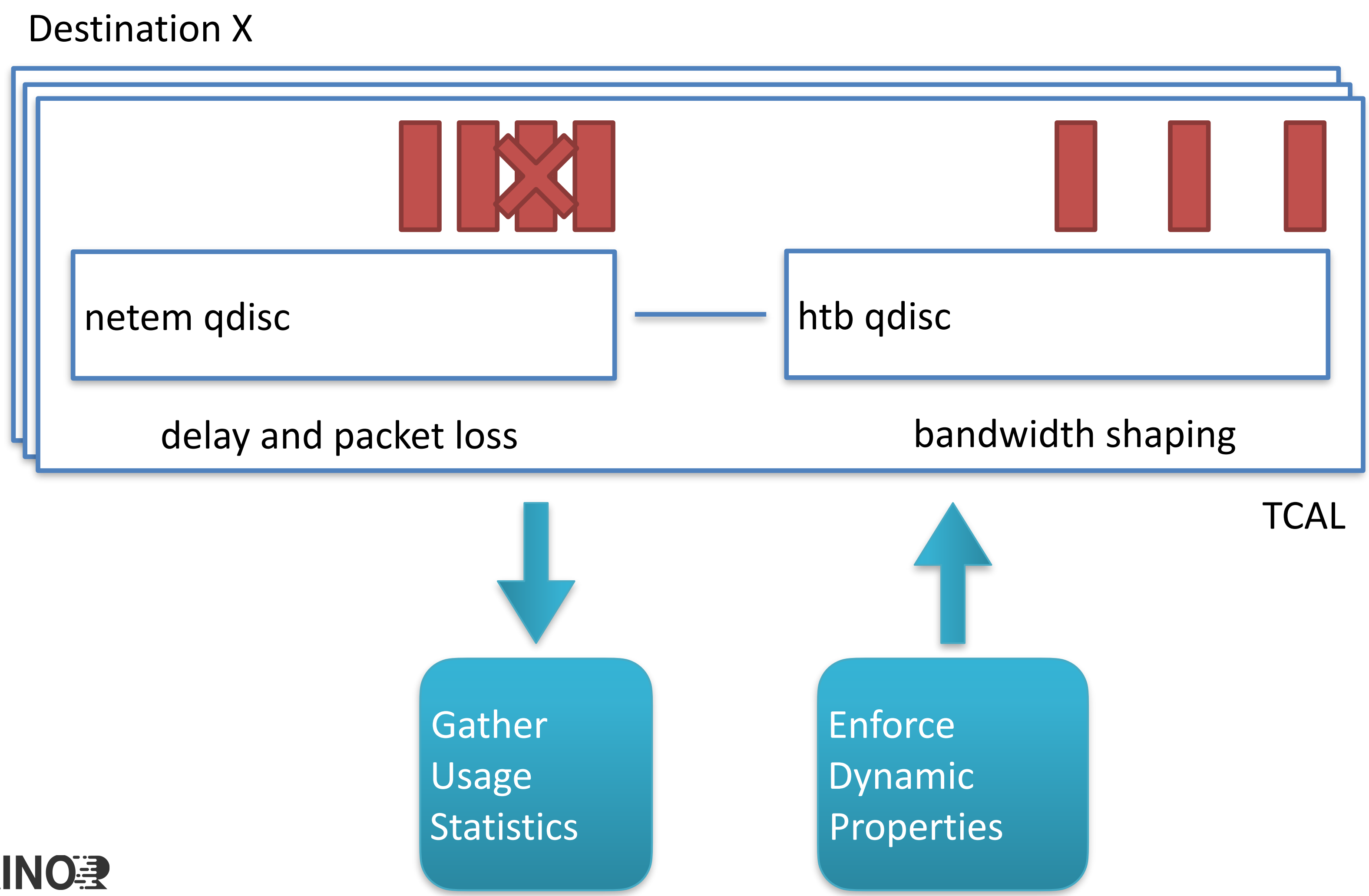
EMULATION CORE

- Spawned by the Emulation Manager
- One instance per container
- Collect's container's usage
- Exchanges metadata with EC through shared memory
  - no bandwidth overhead for local containers

ARCHITECTURE

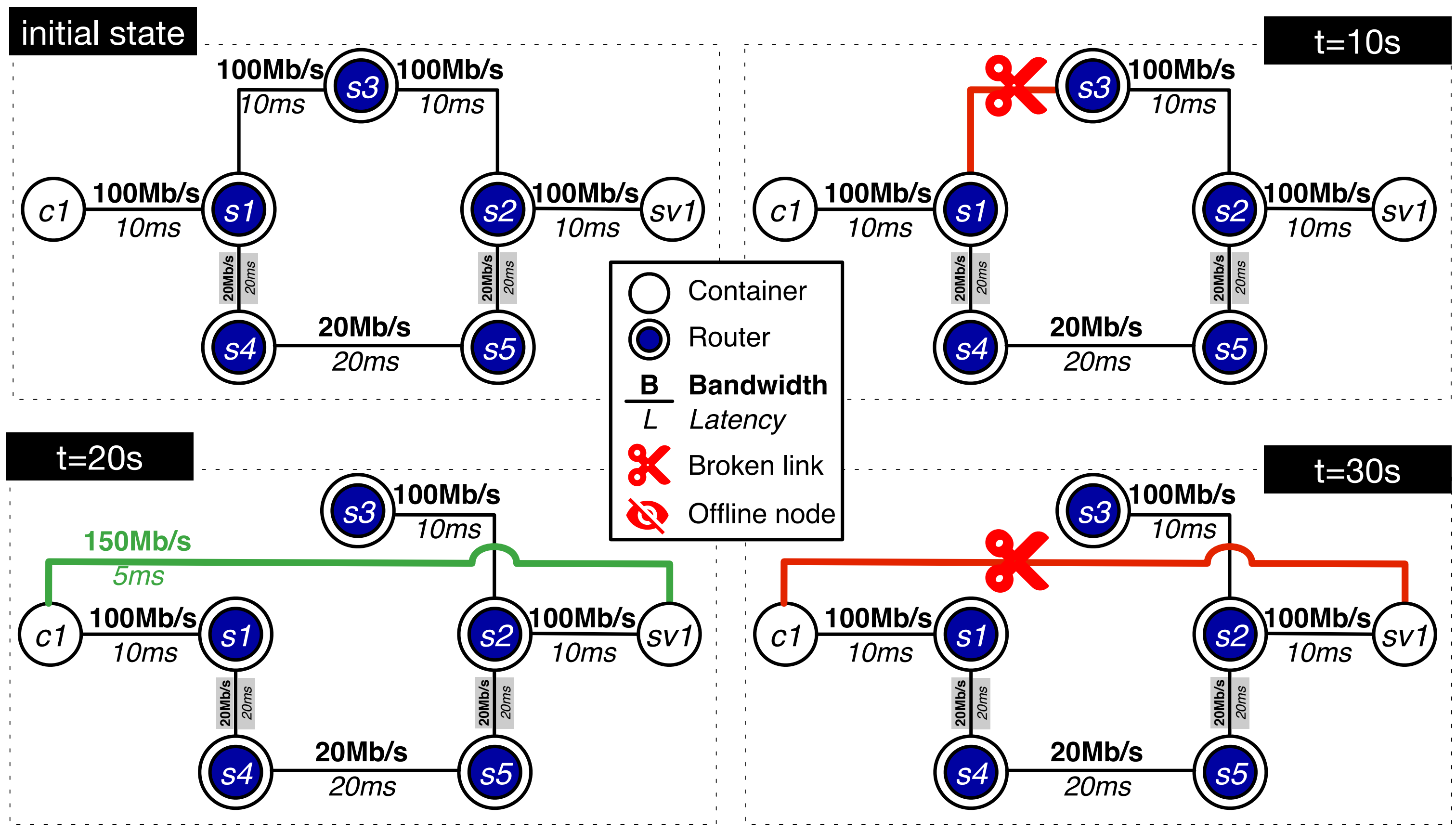


LINUX TC ABSTRACTION LAYER



DESCRIBE DYNAMIC EXPERIMENTS

- Emulation of network dynamics



ANGAINOR

inesc id lisboa

THUNDERSTORM DESCRIPTION LANGUAGE

```yaml
experiment:
  services:
    name: c1
      image: "iperf"
    name: sv
      image: "nginx"
        replicas: 2
  bridges:
    name: s1
    name: s2
  links:
    orig: c1
      dest: s1
      latency: 10
      up: 10Mbps
      down: 10Mbps
      jitter: 0.25
```

```yaml
dynamic:
  orig: c1
    dest: s1
    jitter: 0.5
    time: 120
  action: leave
    name: s1
    time: 200
  action: join
    orig: c1
    dest: s2
    up: 100Mbps
    down: 100Mbps
    latency: 10
    time: 210
  action: leave
    name: sv
```

EVALUATION

- Link-level emulation
- Scalability and metadata overhead
- Short- and long-lived connections
- Cubic and Reno congestion control algorithms
- Dynamic behavior
- Large-scale topologies
- Reproducing published results
- Geo-replicated Systems
- What-if use cases

EVALUATION

- Link-level emulation
- Scalability and metadata overhead
- Short- and long-lived connections
- Cubic and Reno congestion control algorithms
- Dynamic behavior
- Large-scale topologies
- Reproducing published results
- Geo-replicated Systems
- What-if use cases

LARGE-SCALE TOPOLOGIES

- Scale-free networks with random ping requests
- Mean-square error w.r.t. theoretical RTT:

Size (# nodes + # switches)	KOLLAPS	Mininet	Maxinet
1000	0.0261	0.0079	28.0779
2000	0.0384	N/A	347.5303
4000	0.0721	N/A	N/A

GEO-REPLICATED SYSTEM

- Cassandra on EC2 (replication factor: 2)
  - 4 replicas in Frankfurt
  - 4 replicas in Sydney
  - 4 YCSB clients in Frankfurt

WHAT-IF SCENARIO

- Cassandra on EC2 (replication factor: 2)
  - 4 replicas in Frankfurt
  - 4 replicas in ~~Sydney~~ (Seoul)
  - 4 YCSB clients in Frankfurt

CONCLUSION AND FUTURE WORK

- KOLLAPS: a decentralized topology emulator
  - Focuses on end-to-end properties
  - Relies on network collapsing techniques
  - Leverages Linux tc and Container Technologies
- Thunderstorm
  - Language to concisely write dynamic experiments
  - Precise description of experiments
    - Key to reproducibility
- Team
  - Miguel Matos, Valerio Schiavoni, Shady Issa, Paulo Gouveia, João Neves, Carlos Segarra, Luca Liechti

https://github.com/miguelammatos/Kollaps

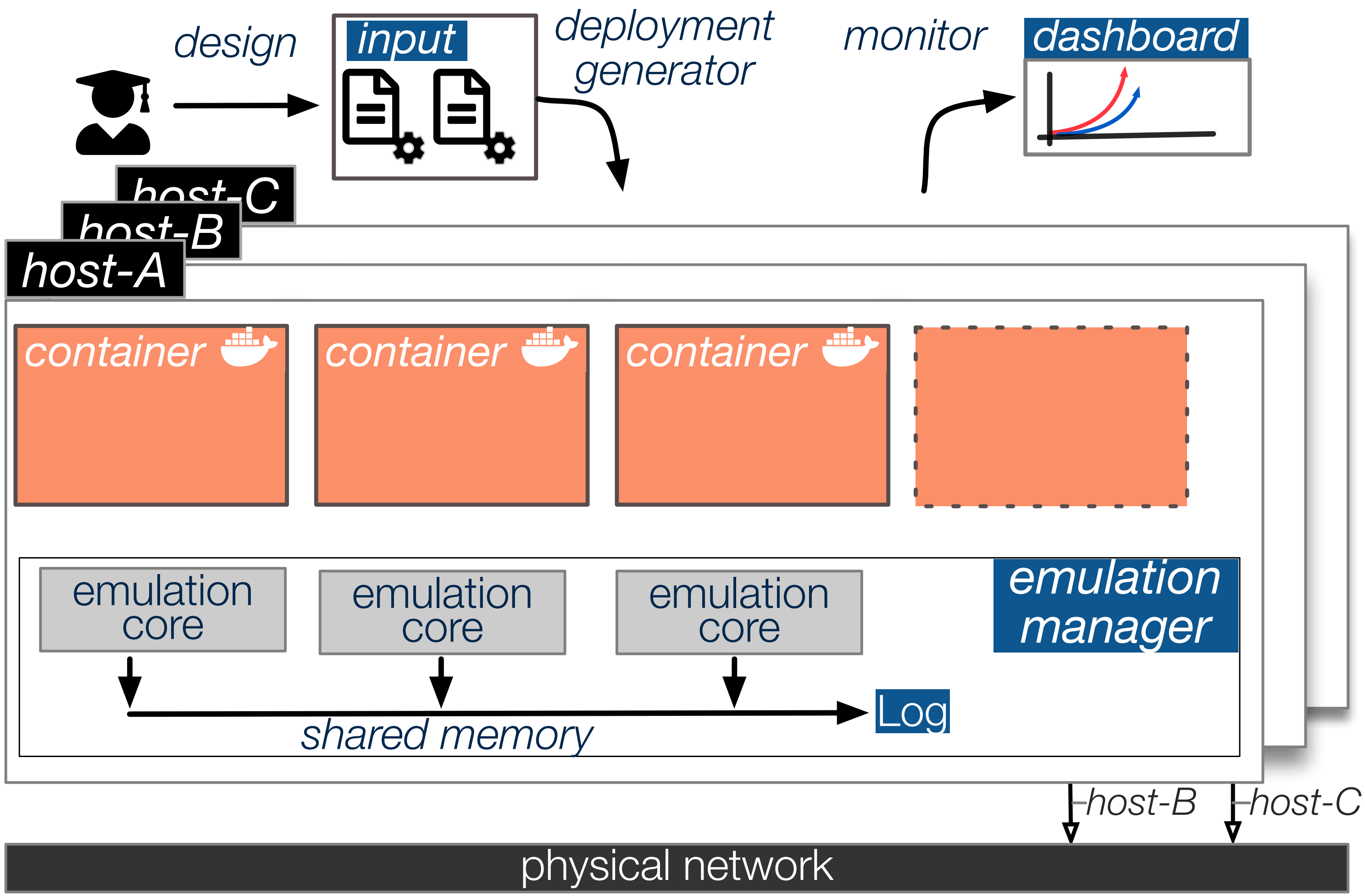
PART II: HANDS-ON TUTORIAL

- Overview of Kollaps & Thunderstorm
- **Hands-on tutorial: basics**
- Hands-on tutorial: advanced features

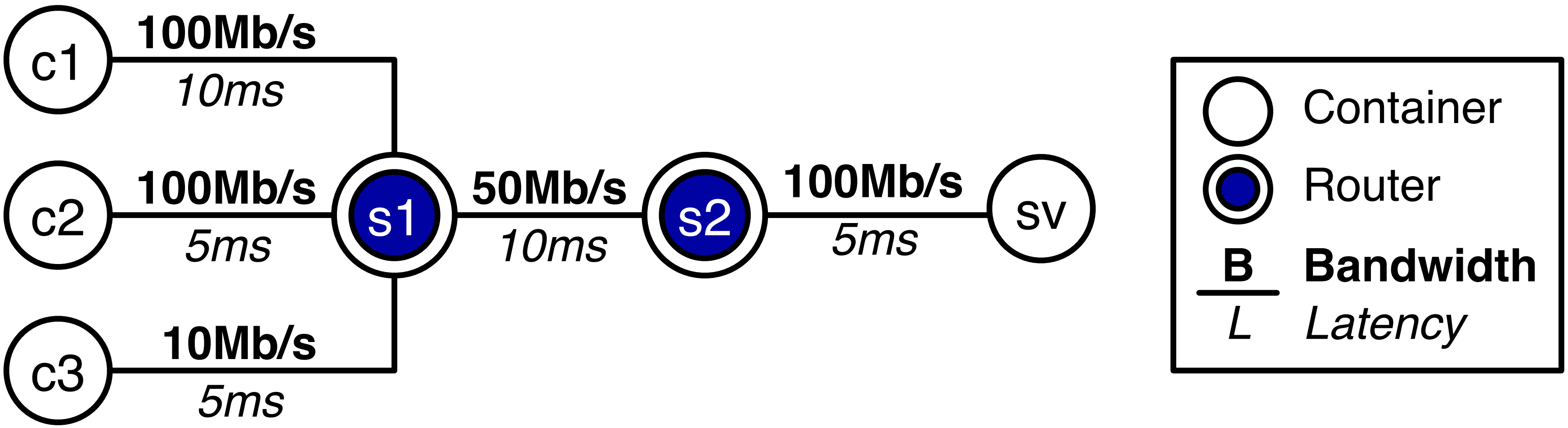
PART II: HANDS-ON TUTORIAL

- Goal
  - Install Kollaps/Thunderstorm
    - Assumptions
      - *Linux*
      - *Docker and Docker Swarm*
  - Run a simple experiment
    - iPerf3 server
    - iPerf3 client
    - measure bandwidth
    - measure ping

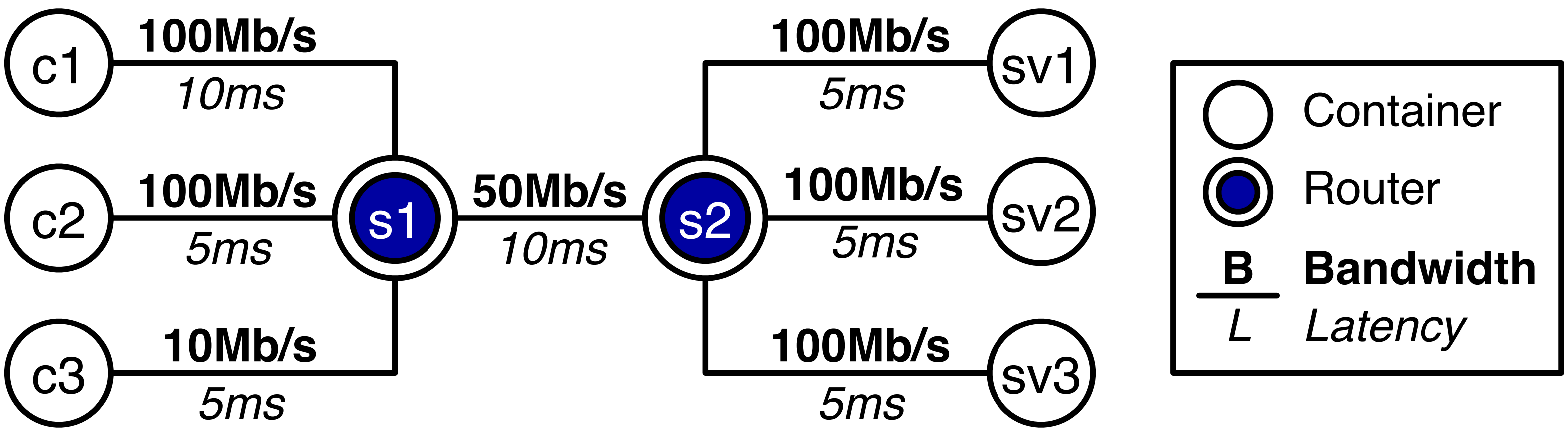
ARCHITECTURE



IPERF3 TOPOLOGY



IPERF3 TOPOLOGY



PART II: HANDS-ON TUTORIAL

- Goal
  - Install Kollaps/Thunderstorm
    - Assumptions
      - *Linux*
      - *Docker and Docker Swarm*
  - Run a simple experiment
    - iPerf3 server
    - iPerf3 client

https://github.com/miguelammatos/Kollaps

https://github.com/miguelammatos/Kollaps/wiki