Pending Interest Table Sizing in Named Data Networking

Luca Muscariello
Orange Labs Networks / IRT SystemX
G. Carofiglio (Cisco), M. Gallo, D. Perino (Bell Labs)

2nd ACM Conference on Information-Centric Networking
San Francisco 1st of October

motivation

- the pending interest table is responsible for maintaining the data path in NDN
- it is a key data structure that requires careful dimensioning
- when the PIT is full it is not obvious how to manage it
- we want to compute the distribution of the PIT size under realistic traffic assumptions
- PIT size as a function of the offered traffic load

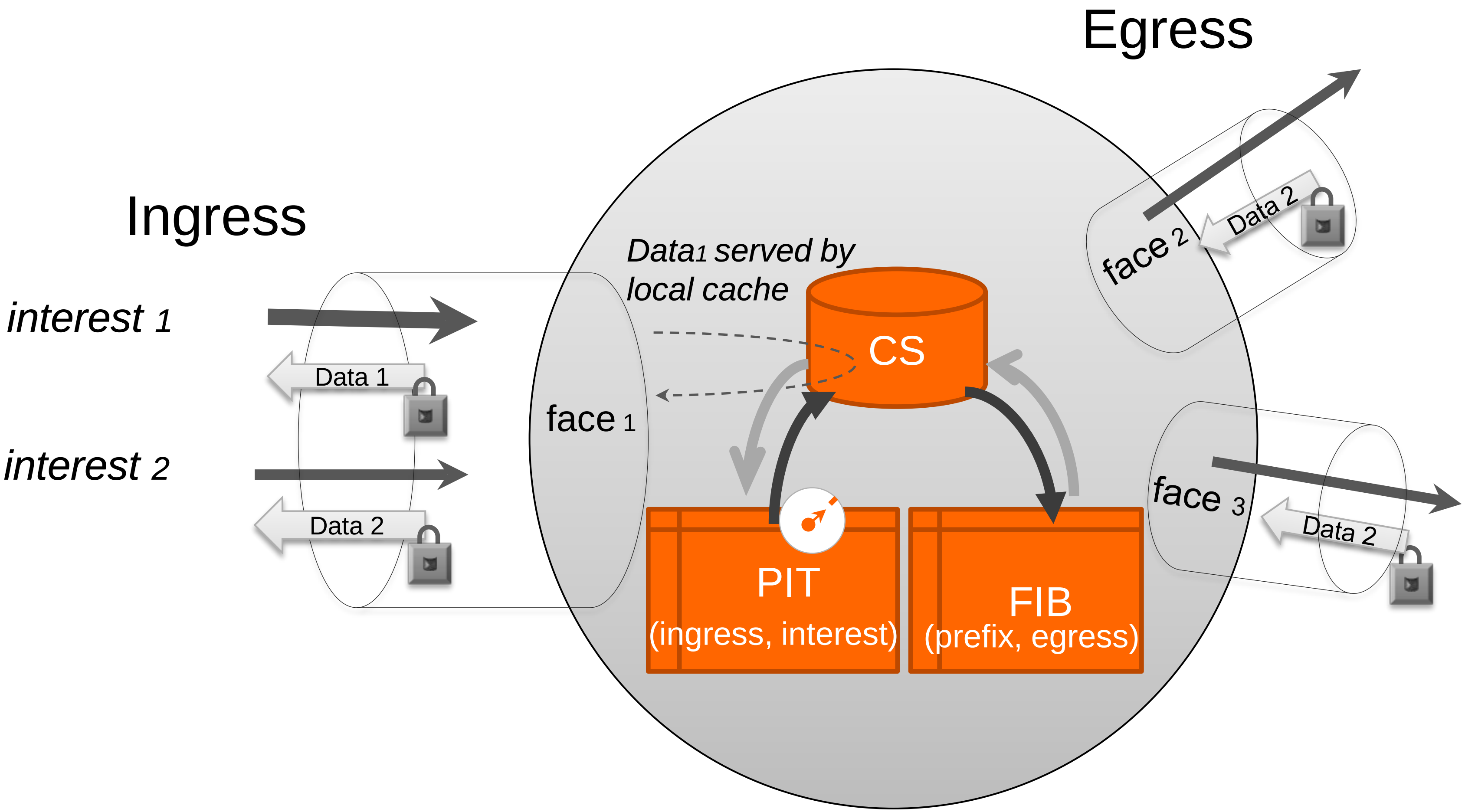
outline

1 system dynamics

2 mathematical modeling

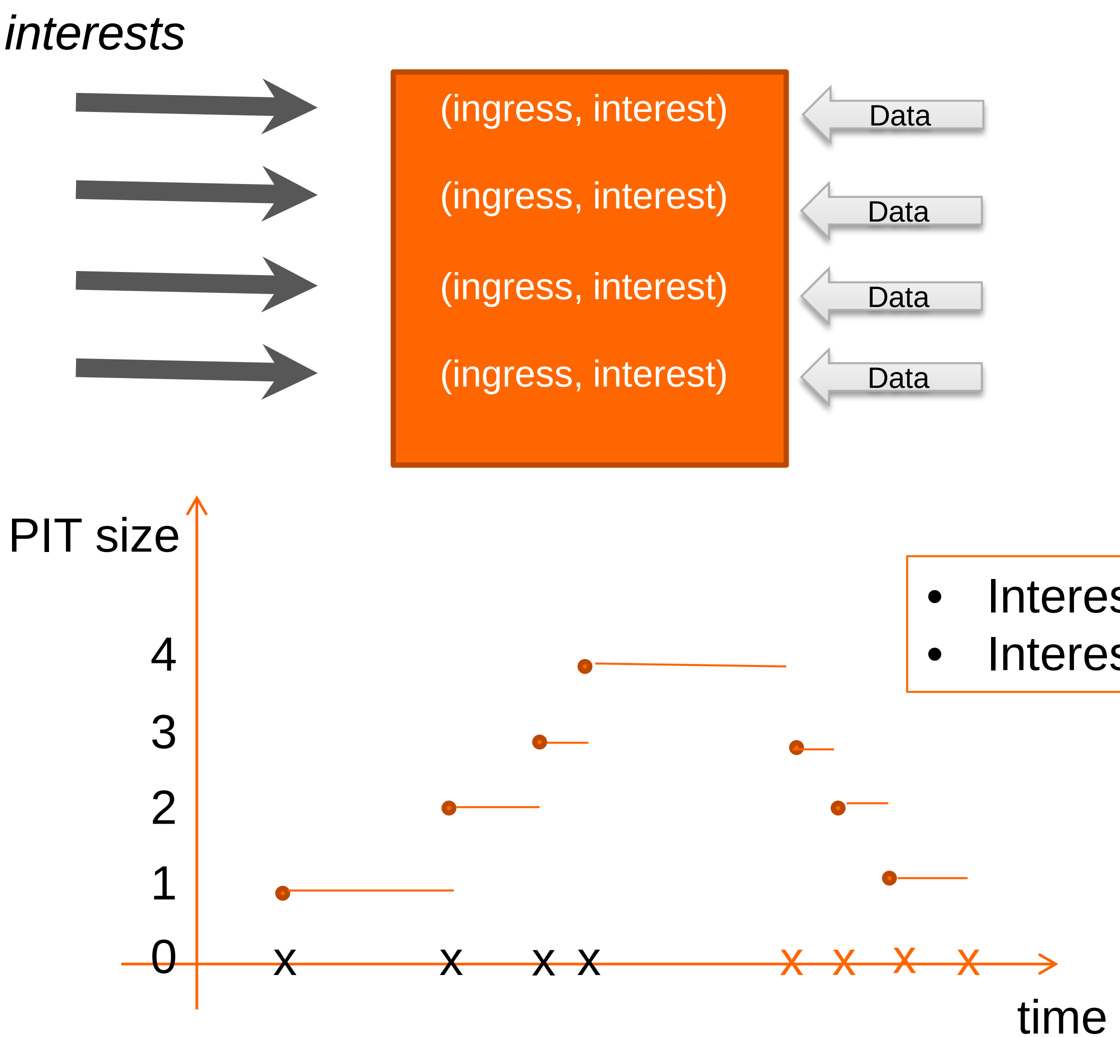
3 sizing

dynamics (1/2)



Egress
Ingress
interest 1
Data 1
interest 2
Data 2
Data1 served by
local cache
CS
face 1
PIT
(ingress, interest)
FIB
(prefix, egress)
face 2
Data 2
face 3
Data 2

dynamics (2/2)



- Interest arrival process
- Interest lifetime

traffic model

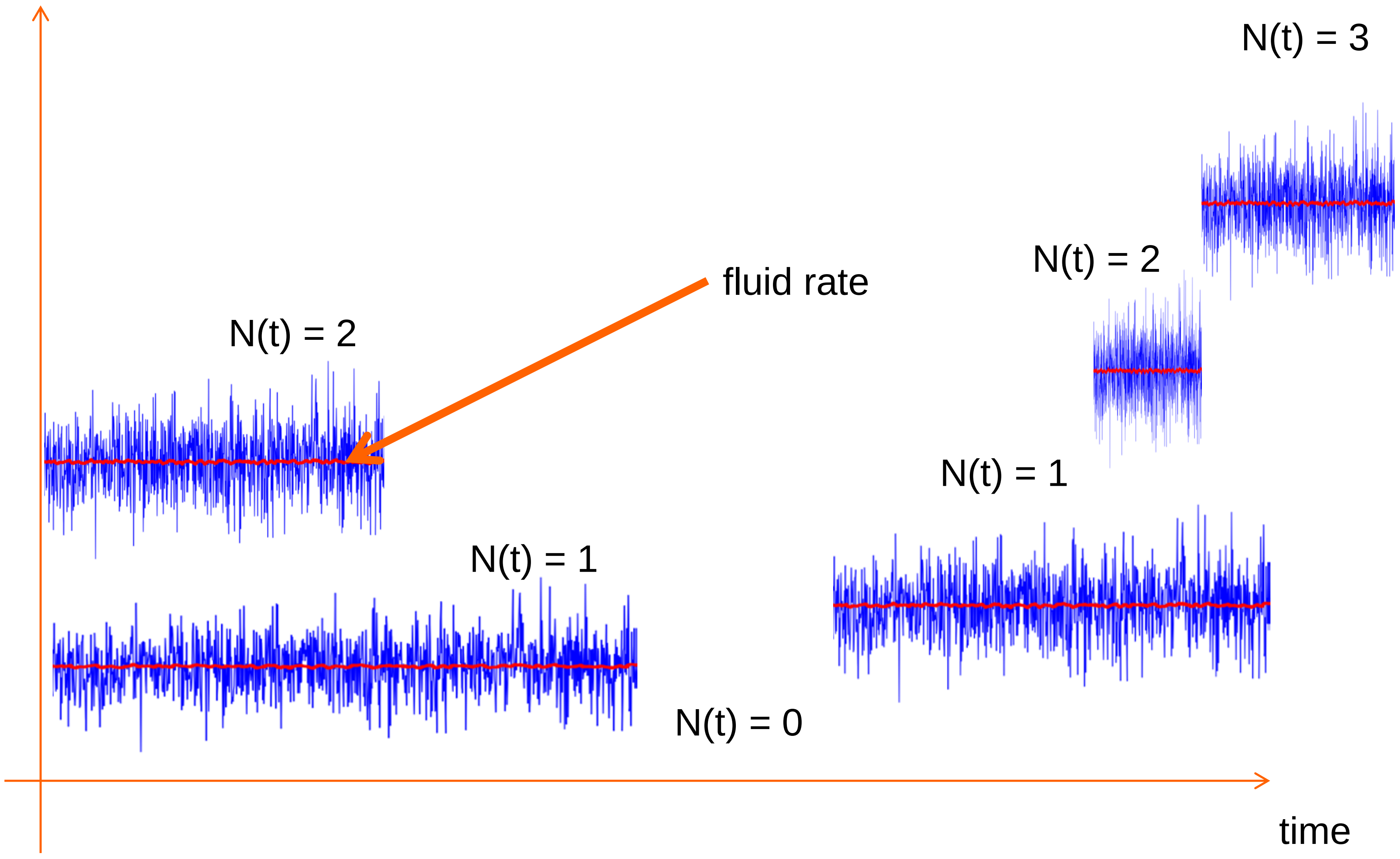
- we want to compute the size of the PIT as a function of the offered traffic
- for sizing purposes we want the quantiles
- under some general assumptions:

- objects are requested following a random process
  - we chose an object **Poisson arrival** process with rate λ

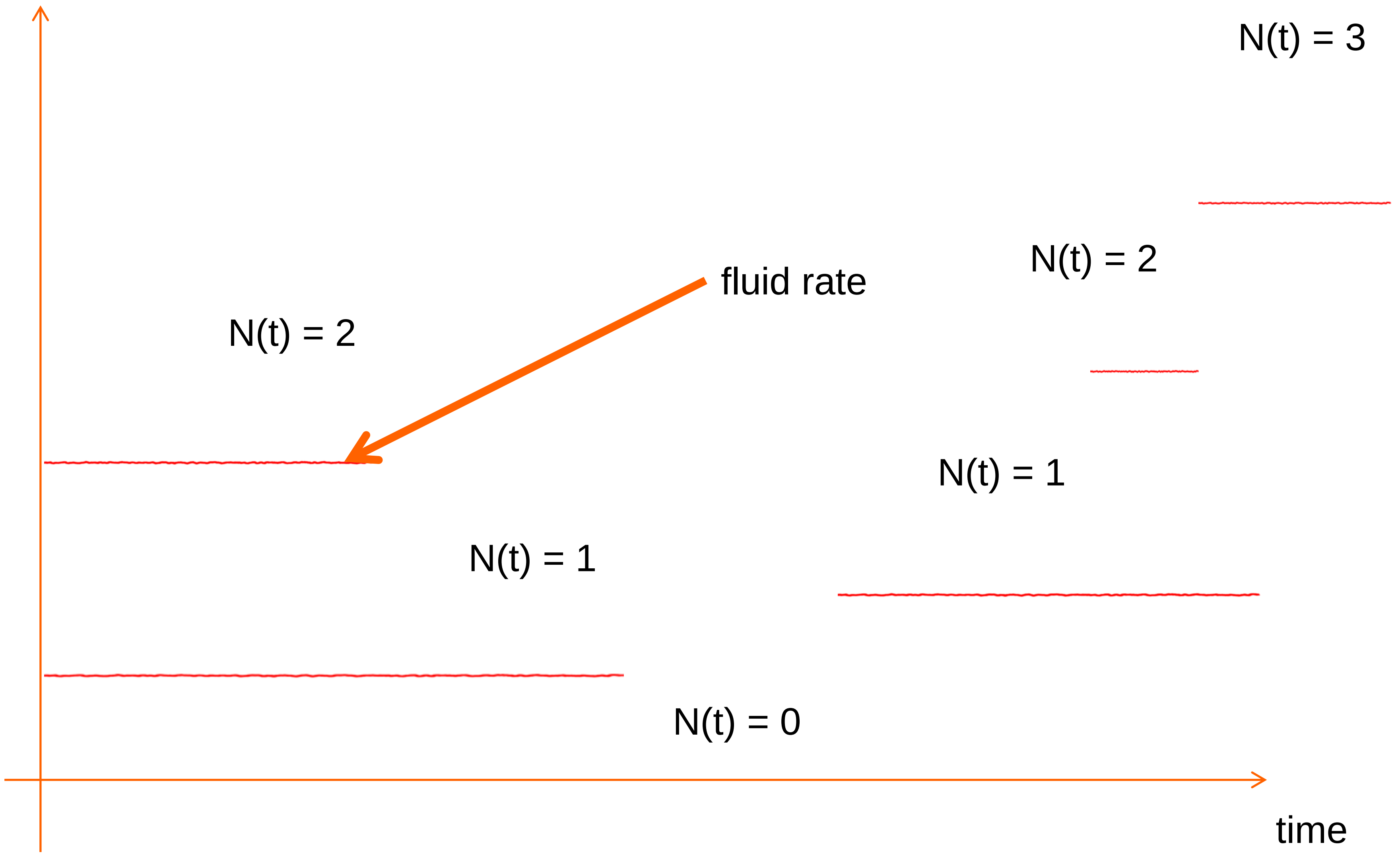
- an object has distributed size S with finite average

- an object is retrieved by variable rate interest requests
  - the rate is congestion controlled
  - the congestion control protocol is receiver driven
  - is also delay based
  - cf Carofiglio et al IEEE ICNP 2013

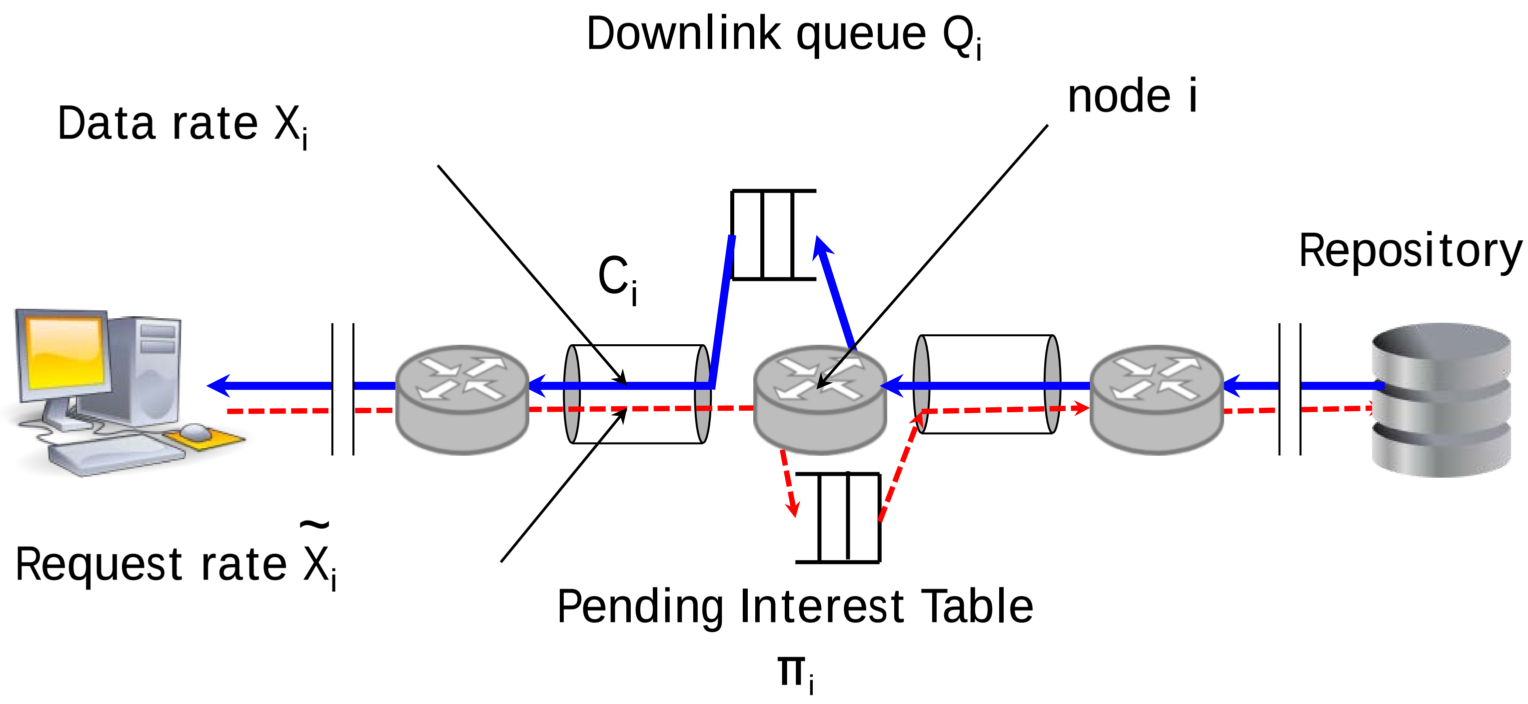
two levels model of the interest rate



two levels model of the interest rate



single transfer PIT occupancy (line network)



state equations (1/2)

Interest rate

Data rate

Interest rate decrease ratio

receiver interest rate

$$\frac{d\tilde{X}^n(t)}{dt} = \frac{\eta}{R(t)^2} - \beta \tilde{X}^n(t) X^n(t) p(t - R(t))$$

link input/output rates rate

$$X^n_{i-1}(t) = X^n_i(t) \mathbb{1}_{\{Q_i(t)=0\}} + \left(C_i - \sum_{k \neq n} X^k_i(t) \right) \mathbb{1}_{\{Q_i(t)>0\}}$$

PIT size

$$\frac{d\pi_i(t)}{dt} = \sum_n \left(\tilde{X}^n(t) - X^n_{i+1}(t) \mathbb{1}_{\{\pi_i(t)>0\}} \right)$$

state equations (2/2)

congestion function

$$p(t) = p_{\min} + \Delta p \min\left(\frac{R(t) - R_{\min}}{\Delta R}, 1\right)$$

Round trip time

$$R(t) \equiv R(t) = R_{min} + \sum_{i=1}^{L} Q_i(t)/C_i$$

$$\frac{dQ_i(t)}{dt} = \sum_n X_i^n(t) - C_i \mathbb{1}_{\{Q_i(t)>0\}}$$

Link queue evolution

network model

- the network is a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{L})$
- data object retrievals sharing the same route r in the network are grouped in classes
- be $\mathcal{R}_{ij}$ the set of routes flowing through link ij
- in case of link congestion capacity C_{ij} is shared assuming max-min fairness (approximation or fair queueing assumed) with fair rate X_r

$$\sum_{r \in \mathcal{R}_{ij}} X_r N_r = C_{ij} \text{ and } X_r = \max_{r' \in \mathcal{R}_{ij}} X_{r'}$$

- the number of data transfers in progress on route N_r is a Markov process
- stability is guaranteed by the condition

$$\rho_{ij} = \sum_{r \in \mathcal{R}_{ij}} \lambda_r \sigma_r / C_{ij} < 1$$

- being ρ the offered load on link ij

main results (1/2)

- N flows, single routing class

  After a transient phase, PIT sizes $\pi_i(t)$ are empty above the bottleneck ($\forall i \geq i^*$)
  And equal to the bottleneck queue length below:

$$\pi_i(t) = \begin{cases} Q_{i^*}(t), & \forall i < i^*, \\ 0, & \forall i \geq i^*, \end{cases}$$

- It means that for sizing purposes we need to only focus on the routes that are bottlenecked upstream a given node.

main results (2/2)

- average values

$$\bar{\tilde{X}}^n = \bar{X}^n = C_{i^*}/N, \bar{\pi}_i = \bar{Q}_{i^*} = \left(\frac{N^2 \eta C \Delta R}{\beta \Delta p}\right)^{1/3}, i < i^*.$$

- maximum PIT size in steady state (variance estimation)
  based on the analysis of the modulus of the **Laplace Transform** of the bottleneck queue function:

$$\max_{t>0} Q(t) = \max_{t>0} \pi_i(t) \leq \bar{Q} + \frac{CR}{2\sqrt{3}} = \bar{Q}\frac{1+2\sqrt{3}}{2\sqrt{3}}, i < i^*$$

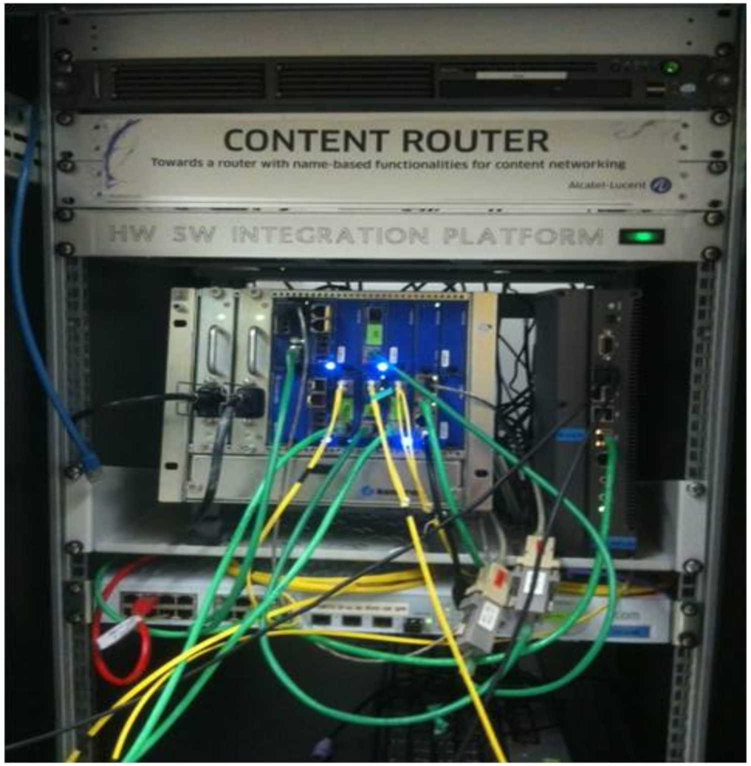
where $Q(t) \equiv Q_{i^*}(t)$, $C \equiv C_{i^*}$.

- taking into account the variable number of transfers in progress

$$\mathbb{E}[\pi_v] = \sum_{ij \in \mathcal{L}_v} Q_{ij}(1) \sum_{r \in \mathcal{R}_{ij}^u} \mathbb{E}[N_r] \leq \sum_{ij \in \mathcal{L}_v} Q_{ij}(1) \sum_{r \in \mathcal{R}_{ij}^u} \frac{\rho_r}{1-\rho_r}.$$

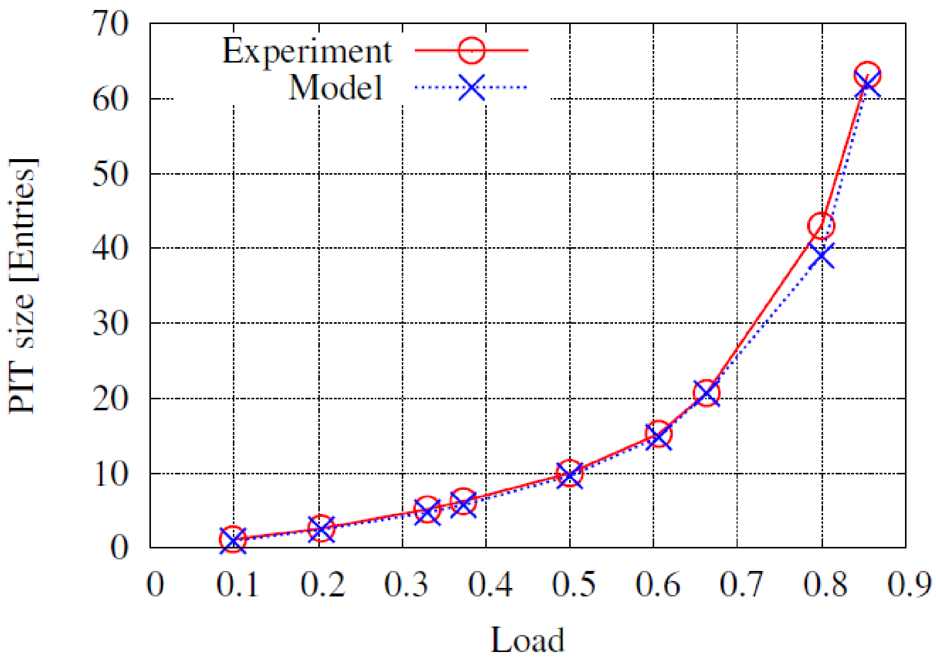
experimental analysis (the platform)

- The platform:
  - 4 AMC boards in a microTCA
  - NPU with 4GB off chip DRAM
  - a set of 10GbE
  - 12 cores per NPU
  - 800MHz 64bits MIPS
    16kB L1 cache , 2MB L2 cache
  - an NDN node per card
- the forwarder:
  - PIT optimized open-addressed hash table
  - hardware timers for PIT timeouts
  - data collection by a platform controller not to affect forwarding
  - sample are processed offline
  - faces over UDP
- **1422B-92B data/interest** packets
- traffic generation on client/repo servers



comparison model/experiments

- line network
- single bottlenecked link upstream at 100Mbps (all others at 5Gbps)
- relation PIT size/offered load is correctly measured by the model
- experiments are run form 100Mbps to 1Gbps



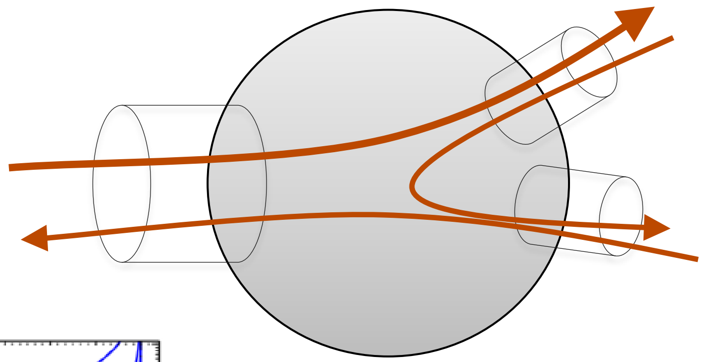
PIT sizing

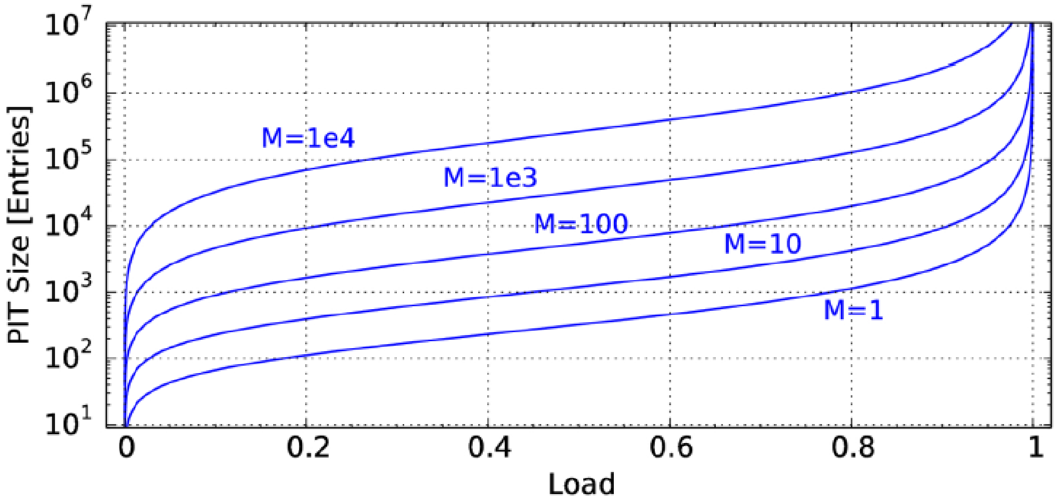
- PIT sizing is made by using the 95% percentile assuming a Gaussian approximation $\mu(1 + z_\alpha c_v)$

- M routes with the same offered load bottlenecked upstream

$$\mu = M\frac{\rho}{1-\rho}\sqrt{\frac{\eta}{\beta\Delta p}(1+2\sqrt{3})}$$

$$\sigma = \mu(1+1/2\sqrt{3})\sqrt{1+1/\rho}/\sqrt{M}$$





conclusions

- the model catches the essential properties of realistic traffic assumptions
  - congestion controlled sources with **delay based congestion control**
  - the knowledge of traffic that is bottleneck upstream is important to compute this size
  - fluid models turn out to be tractable to obtain simple closed formulas
- the PIT stores information about congestion level downstream/upstream
- under congestion controlled traffic the PIT size does not constitute a barrier for high speed implementations
- for non controlled (poorly controlled) traffic the PIT size requires active (local) management

Thank you