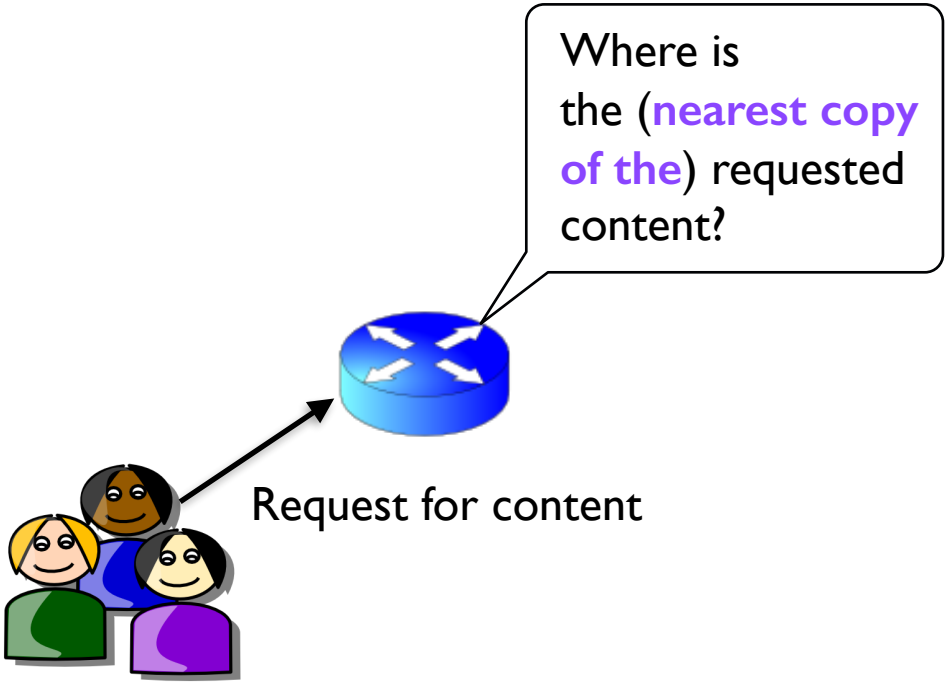
On Content Indexing for Off-Path Caching in Information-Centric Networks

Suzan Bayhan, Liang Wang, Jörg Ott, Jussi Kangasharju, Arjuna Sathiaseelan, Jon Crowcroft

University of Helsinki (Finland), TU Munich (Germany), University of Cambridge (UK)

UNIVERSITY OF CAMBRIDGE

ACM ICN, Kyoto, Japan, September 26-28, 2016


Where is
the (nearest copy
of the) requested
content?
Request for content

Name resolution service in ICN

- **Standalone service (look-up by name)**
  - Directory service, *Name Resolution Server (NRS)*
  - NRS maps names to locators and routing is done using locators
  - 🙂 Nearest-copy routing
  - ☹️ Scalability (temporary copies, update, storage, lookup)
- **Name-based routing**
  - Route on names
  - 🙂 No need for a particular infrastructure support
  - ☹️ Nearest-copy routing?

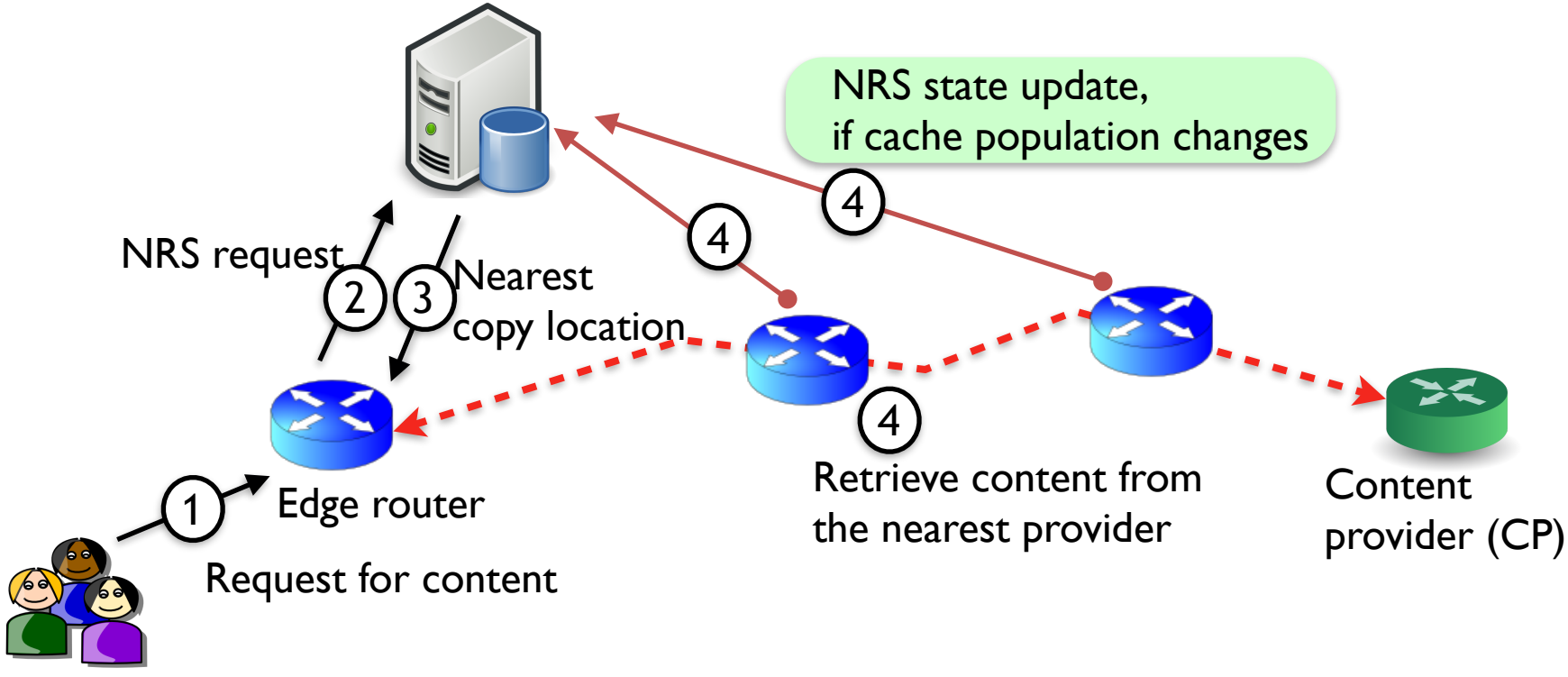
Our contribution

- Can *we have an NRS that indexes only some of the content but brings most of the benefits?*
  - Partial NRS to balance the tradeoff between **scalability** and **nearest copy routing**

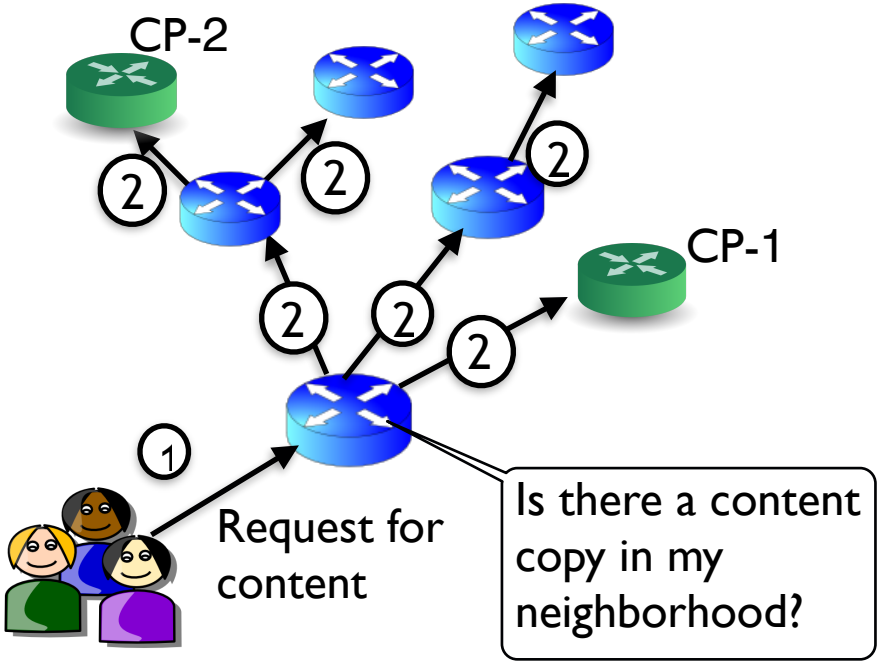
- We model the content delivery with and without NRS, and formulate an optimization problem to identify **which items to index**

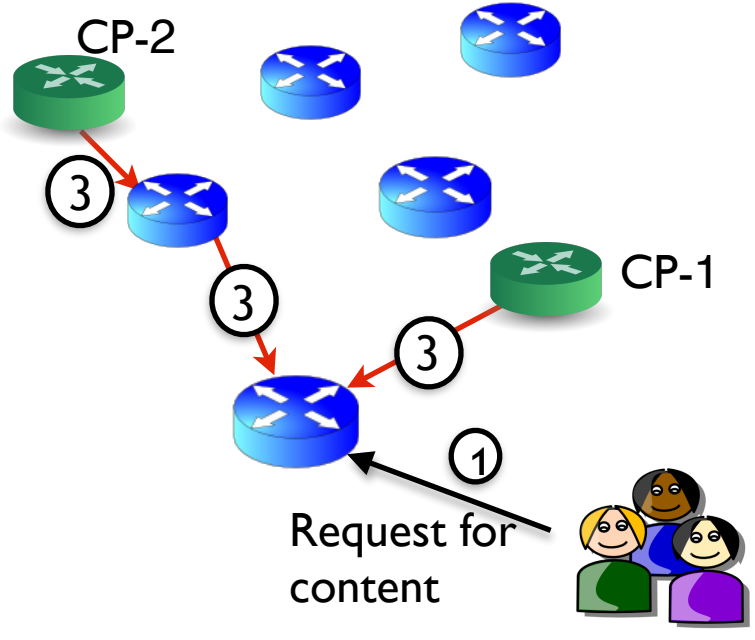
- Our **optimal NRS setting** aims to minimize the content delivery cost under a limit on maximum number of items to be indexed

NRS-based operation in ICN



Operation without an NRS in ICN

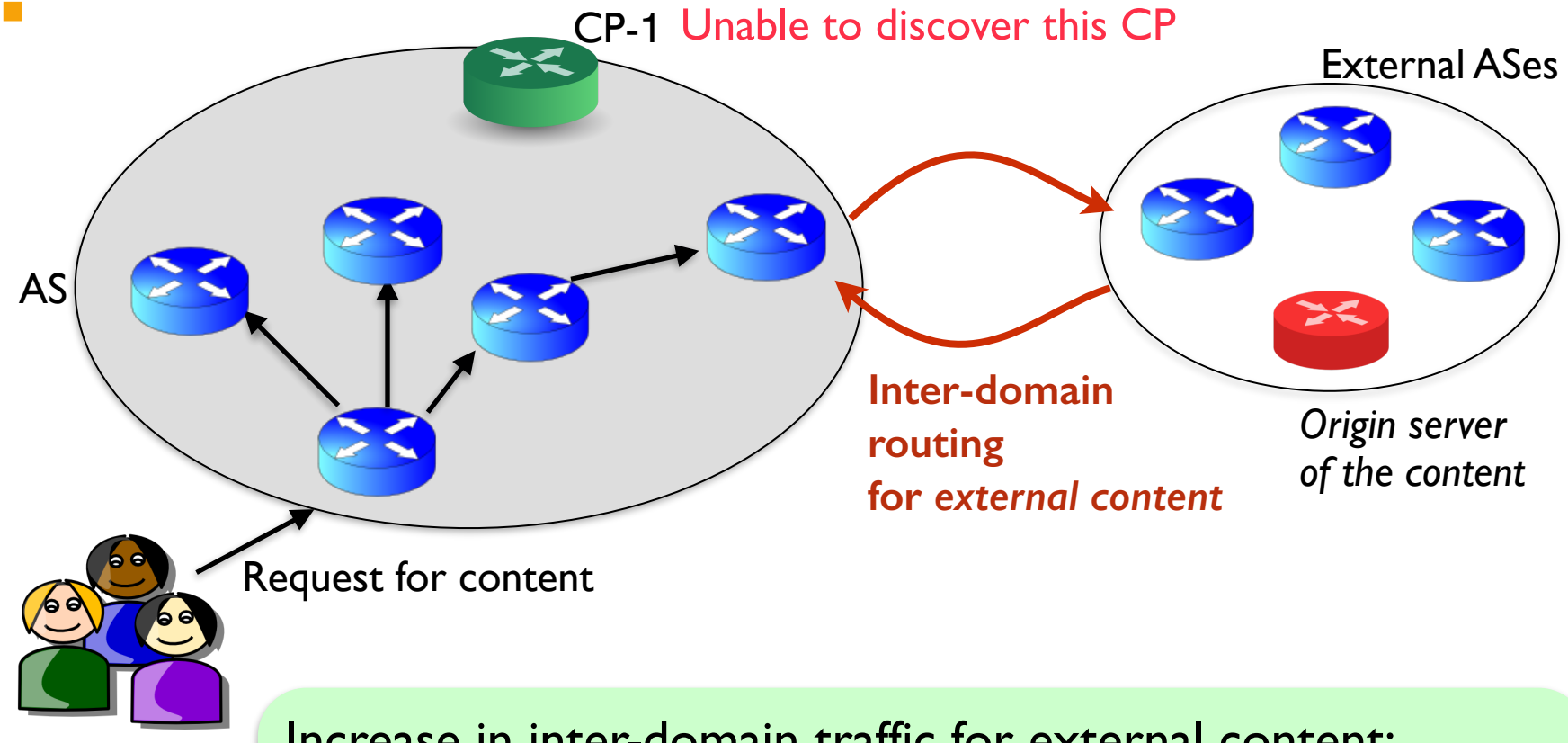




- Content discovery:
  - Search for content in the neighborhood
  - Flooding-based search
  - Time waste if content is NOT in the nw.

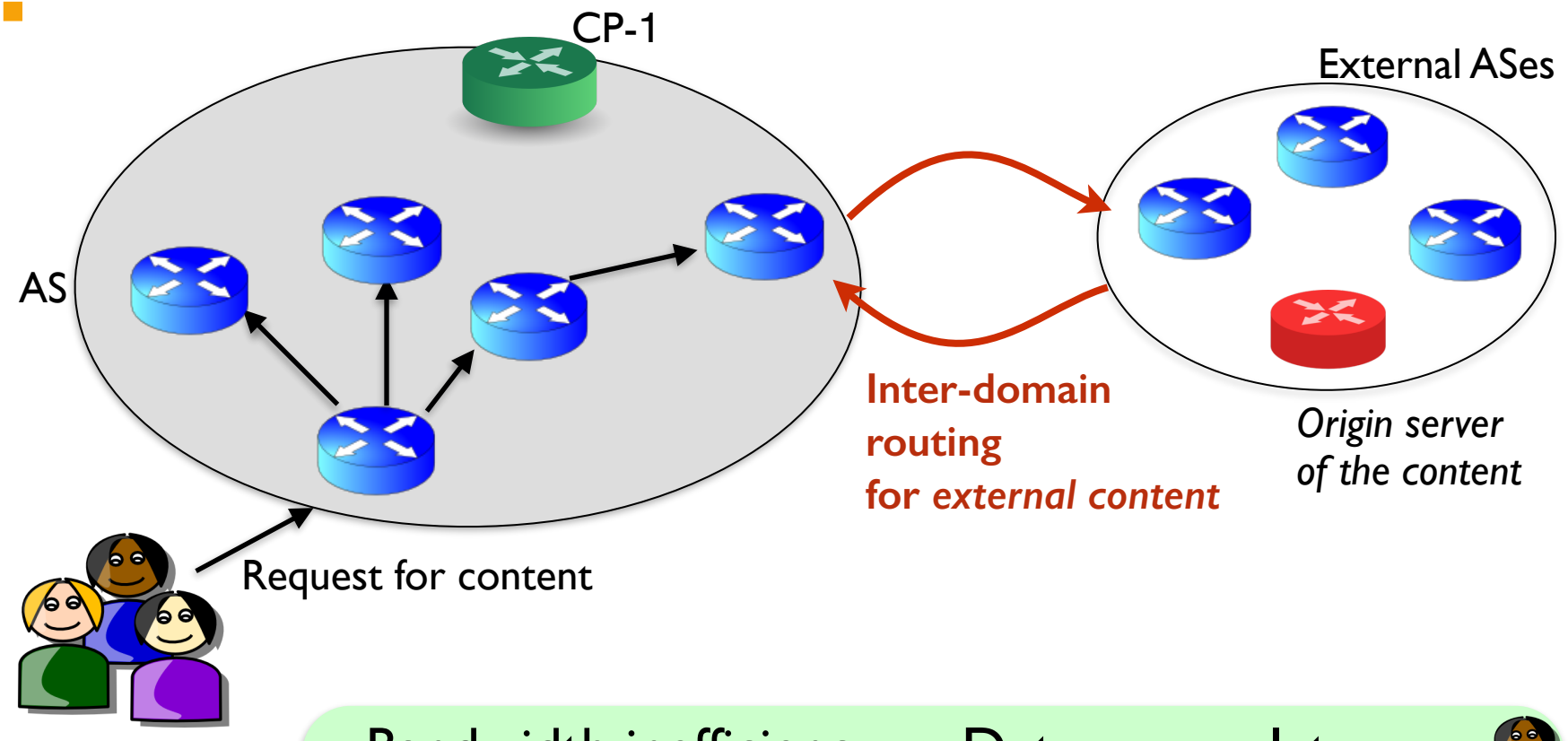
- Content retrieval: Redundant transmissions due to multiple content providers in the search scope

Operation without an NRS in ICN



Increase in inter-domain traffic for external content:
monetary inefficiency

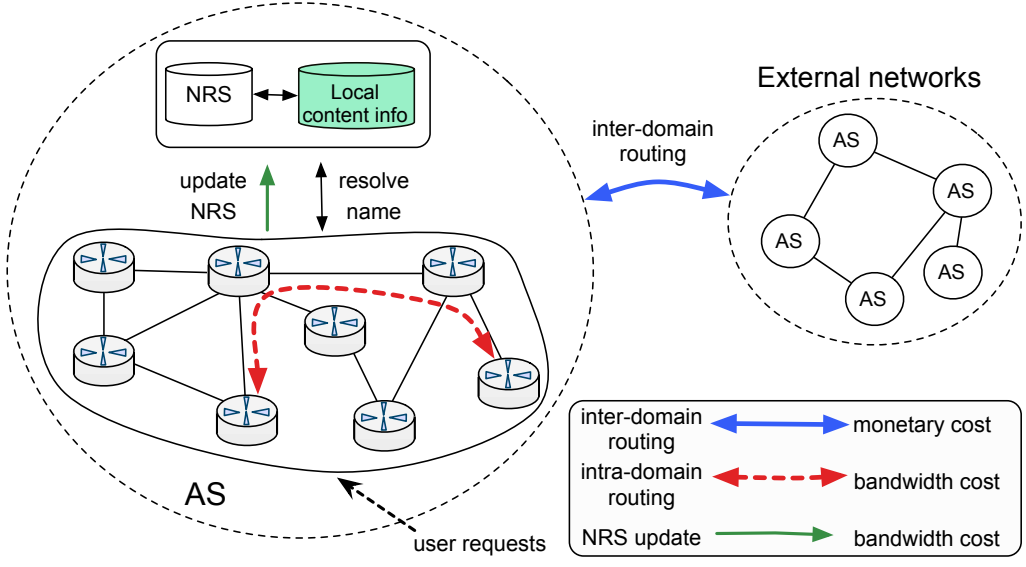
Operation without an NRS in ICN



- Bandwidth inefficiency
- Data access latency
- Monetary inefficiency

A closer look from an AS' viewpoint

- **Local content:** origin server in the AS,
- **External content**: origin server outside AS
- To improve scalability:
  - NRS indexes only some fraction(w) of all content catalogue (K)



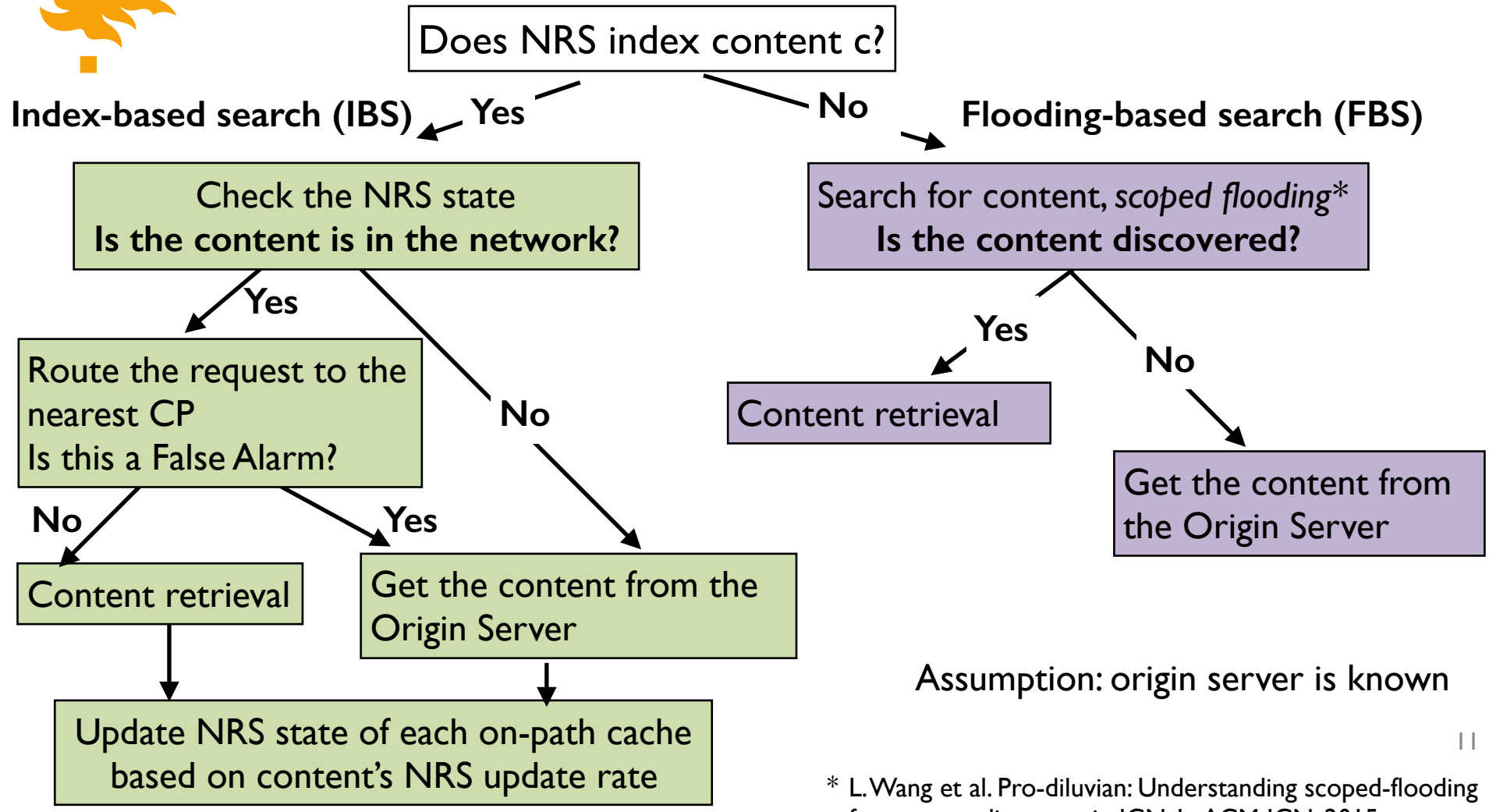
- Routers update the NRS **not upon every single change in their cache**, but based on calculated rates under certain *false positive* and *negative* probabilities according to rate-distortion theory proposed in Azimdoost et al.

B. Azimdoost, C. Westphal, and H. R. Sadjadpour. The price of updating the control plane in ICNs. *CoRR*, abs/1406.1284, 2014.

Which items to index by NRS?

- Calculate the **cost of content delivery** with NRS and without NRS
- **Indexing gain** is the difference between the cost
- Index the ones with the highest indexing gain

Cost of content delivery: discovery and retrieval



Assumption: origin server is known

* L. Wang et al. Pro-diluvian: Understanding scoped-flooding for content discovery in ICN. In ACM ICN, 2015.

NRS-based content delivery: more formally

System state: <Content state, NRS state>

		NRS state, $S_{NRS}(k)$			Content is retrieved from
		NRS indexes c_k		NRS does not index c_k	
		0	1	NA	
Content state $S(k)$	0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
	1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

- Above steps are taken if the content is NOT in the edge cache
- Content is expected to be in the cache with probability α: Che's approximation

NRS-based content delivery: more formally

Content state $S(k)$	NRS state, $S_{NRS}(k)$: NRS indexes c_k: 0	NRS indexes c_k: 1	NRS does not index c_k: NA	Content is retrieved from
0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

False negative False positive

NRS-based content delivery: more formally

		NRS state, $S_{NRS}(k)$			Content is retrieved from
		NRS indexes c_k		NRS does not index c_k	
		0	1	NA	
Content state $S(k)$	0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
	1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

Rely on IBS

NRS-based content delivery: more formally

		NRS state, $S_{NRS}(k)$			Content is retrieved from
		NRS indexes c_k		NRS does not index c_k	
		0	1	NA	
Content state $S(k)$	0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
	1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

Rely on FBS

NRS-based content delivery: more formally

		NRS state, $S_{NRS}(k)$			Content is retrieved from
		NRS indexes c_k		NRS does not index c_k	
		0	1	NA	
Content state $S(k)$	0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
	1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

Cost of IBS, FBS, and fetching from the origin server:

of transmissions of the message x the message size x cost of transmission/bit

- IBS: expected hop distance between the nearest content provider and a randomly selected requesting node
- FBS: # of nodes receiving the message for a certain search scope (2 or 3 hops)
- Origin server:
  - expected hop distance between a randomly-selected router and a content provider
  - inter-AS routing cost

NRS-based content delivery: more formally

		NRS state, $S_{NRS}(k)$			Content is retrieved from
		NRS indexes c_k		NRS does not index c_k	
		0	1	NA	
Content state $S(k)$	0	Fetch from the origin $p_{00}=(1-\alpha_k)(1-P_k)(1-\varepsilon^1)$	IBS, fetch from the origin $p_{01}=(1-\alpha_k)(1-P_k)\varepsilon^1$	FBS, fetch from the origin $p_{0na}=(1-\alpha_k)(1-P_k)$	AS, if local content External AS, if external content
	1	Fetch from the origin $p_{10}=(1-\alpha_k)P_k(1-\varepsilon^0)$	IBS $p_{11}=(1-\alpha_k)P_k\varepsilon^0$	FBS, fetch from the origin $p_{1na}=(1-\alpha_k)P_k$	AS, if discovered or local External AS, if external and not discovered

Cost for discovery, retrieval and NRS update

Content discovery cost: $\phi_k = \begin{cases} l^{req}(\alpha_k\phi^c + (p_{11}+p_{01})\phi_k^{\text{IBS}} + (1-p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (2) \\ l^{req}(\alpha_k\phi^c + \phi^{\text{FBS}} + (p_{0na}+p_{1na}(1-\gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (3) \end{cases}$

Content retrieval cost: $\beta_k = \begin{cases} s_k(\alpha_k\phi^c + p_{11}\phi_k^{\text{IBS}} + (1-p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (4) \\ s_k(\alpha_k\phi^c + p_{1na}\gamma_k^{\text{FBS}} n_k \phi_k^{\text{IBS}} + (p_{0na}+p_{1na}(1-\gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (5) \end{cases}$

NRS update cost: $\psi_k = \begin{cases} R_k(\varepsilon^1,\varepsilon^0) l^{up}\phi^{\text{up}}, & \text{where } l^{up} = \log K_\omega + \log N + 1, \text{ if } x_k = 1 \quad (6) \\ 0, & \text{if } x_k = 0 \quad (7) \end{cases}$

NRS-based content delivery: more formally

NRS indexes content k

NRS does not index

Content discovery cost: $\phi_k = \begin{cases} l^{req}(\alpha_k \phi^c + (p_{11} + p_{01})\phi_k^{\text{IBS}} + (1 - p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (2) \\ l^{req}(\alpha_k \phi^c + \phi^{\text{FBS}} + (p_{0na} + p_{1na}(1 - \gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (3) \end{cases}$

Content retrieval cost: $\beta_k = \begin{cases} s_k(\alpha_k \phi^c + p_{11}\phi_k^{\text{IBS}} + (1 - p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (4) \\ s_k(\alpha_k \phi^c + p_{1na}\gamma_k^{\text{FBS}} n_k \phi_k^{\text{IBS}} + (p_{0na} + p_{1na}(1 - \gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (5) \end{cases}$

NRS update cost: $\psi_k = \begin{cases} R_k(\varepsilon^1, \varepsilon^0) l^{up} \phi^{\text{up}}, & \text{where } l^{up} = \log K_\omega + \log N + 1, \text{ if } x_k = 1 \quad (6) \\ 0, & \text{if } x_k = 0 \quad (7) \end{cases}$

NRS-based content delivery: more formally

Cost of IBS for content k

Search success for FBS for content k

Cost of getting the content from cache

Cost of fetching content k from origin server

Content discovery cost: $\phi_k = \begin{cases} l^{req}(\alpha_k \phi^c + (p_{11} + p_{01})\phi_k^{\text{IBS}} + (1 - p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (2) \\ l^{req}(\alpha_k \phi^c + \phi^{\text{FBS}} + (p_{0na} + p_{1na}(1 - \gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (3) \end{cases}$

Content retrieval cost: $\beta_k = \begin{cases} s_k(\alpha_k \phi^c + p_{11}\phi_k^{\text{IBS}} + (1 - p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (4) \\ s_k(\alpha_k \phi^c + p_{1na}\gamma_k^{\text{FBS}} n_k \phi_k^{\text{IBS}} + (p_{0na} + p_{1na}(1 - \gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (5) \end{cases}$

NRS update cost: $\psi_k = \begin{cases} R_k(\varepsilon^1, \varepsilon^0) l^{up} \phi^{\text{up}}, & \text{where } l^{up} = \log K_\omega + \log N + 1, \text{ if } x_k = 1 \quad (6) \\ 0, & \text{if } x_k = 0 \quad (7) \end{cases}$

NRS-based content delivery: more formally

$$n_k = 1 + \rho(N^{\mathrm{FBS}}\alpha_k)$$

n_k: overhead factor

$\rho \in [0,1]$

FBS redundancy coefficient

Content discovery cost: $\phi_k = \begin{cases} l^{req}(\alpha_k\phi^c + (p_1 + p_{01})\phi_k^{\mathrm{IBS}} + (1-p_{11})\phi^{\mathrm{ori}}), & \text{if } x_k = 1 \quad (2) \\ l^{req}(\alpha_k\phi^c + \phi^{\mathrm{FBS}} + (p_{0na} + p_{1na}(1-\gamma_k^{\mathrm{FBS}}))\phi^{\mathrm{ori}}), & \text{if } x_k = 0 \quad (3) \end{cases}$

Content retrieval cost: $\beta_k = \begin{cases} s_k(\alpha_k\phi^c + p_{11}\phi_k^{\mathrm{IBS}} + (1-p_{11})\phi^{\mathrm{ori}}), & \text{if } x_k = 1 \quad (4) \\ s_k(\alpha_k\phi^c + p_{1na}\gamma_k^{\mathrm{FBS}} n_k\phi_k^{\mathrm{IBS}} + (p_{0na} + p_{1na}(1-\gamma_k^{\mathrm{FBS}}))\phi^{\mathrm{ori}}), & \text{if } x_k = 0 \quad (5) \end{cases}$

NRS update cost: $\psi_k = \begin{cases} R_k(\varepsilon^1, \varepsilon^0) l^{up}\phi^{\mathrm{up}}, & \text{where } l^{up} = \log K_\omega + \log N + 1, \text{ if } x_k = 1 \quad (6) \\ 0, & \text{if } x_k = 0 \quad (7) \end{cases}$

Cost for discovery, retrieval, and NRS update

$$\text{Content discovery cost: } \phi_k = \begin{cases} l^{req}(\alpha_k \phi^c + (p_{11}+p_{01})\phi_k^{\text{IBS}} + (1-p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (2) \\ l^{req}(\alpha_k \phi^c + \phi^{\text{FBS}} + (p_{0na} + p_{1na}(1-\gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (3) \end{cases}$$

$$\text{Content retrieval cost: } \beta_k = \begin{cases} s_k(\alpha_k \phi^c + p_{11}\phi_k^{\text{IBS}} + (1-p_{11})\phi^{\text{ori}}), & \text{if } x_k = 1 \quad (4) \\ s_k(\alpha_k \phi^c + p_{1na}\gamma_k^{\text{FBS}} n_k \phi_k^{\text{IBS}} + (p_{0na} + p_{1na}(1-\gamma_k^{\text{FBS}}))\phi^{\text{ori}}), & \text{if } x_k = 0 \quad (5) \end{cases}$$

$$\text{NRS update cost: } \psi_k = \begin{cases} R_k(\varepsilon^1, \varepsilon^0) l^{up} \phi^{\text{up}}, & \text{where } l^{up} = \log K_\omega + \log N + 1, \text{ if } x_k = 1 \quad (6) \\ 0, & \text{if } x_k = 0 \quad (7) \end{cases}$$

Rate of update for meeting certain false positive and negative rates *Azimdoost et al.*

l^{up}: NRS update message (payload) size

Indexing for minimum content delivery cost

Utility for item *k* (cost of delivery)

$$
\begin{aligned}
U_k &= q_k(\phi_k + \beta_k) + \psi_k \\
\phi_k &= l^{req}(\alpha_k \phi^c + (p_{11} + p_{01})\phi_k^{\mathrm{IBS}}) + x_k((1-p_{11})\phi^{\mathrm{ori}}) \\
&\quad + (1-x_k)(\phi^{\mathrm{FBS}} + (p_{0na} + p_{1na}(1-\gamma_k^{\mathrm{FBS}}))\phi^{\mathrm{ori}})) \\
\beta_k &= \alpha_k s_k \phi^c + s_k(p_{11}\phi_k^{\mathrm{IBS}} + (x_k((1-p_{11})\phi^{\mathrm{ori}}) \\
&+ (1-x_k)(p_{1na}\gamma_k^{\mathrm{FBS}} n_k \phi_k^{\mathrm{IBS}}) + (p_{0na} + p_{1na}(1-\gamma_k^{\mathrm{FBS}}))\phi^{\mathrm{ori}})) \\
\psi_k &= x_k R_k(\varepsilon^1, \varepsilon^0) l^{up} \phi^{\mathrm{up}}
\end{aligned}
$$

Optimization problem

$$
\begin{aligned}
&\min \quad \sum_{k=1}^{K} U_k \\
&s.t. \sum_{k=1}^{K} x_k \leqslant K_\omega .
\end{aligned}
$$

Indexing gain for item *k*

$$\Delta U_k = U_k(x_k = 0) - U_k(x_k = 1)$$

Select the first *Kw* items

Performance analysis

- Which items to index?
- How much can we benefit with increasing NRS size?
- How is ICN performance affected?
  - cache hits
  - inter-domain traffic
  - data access latency

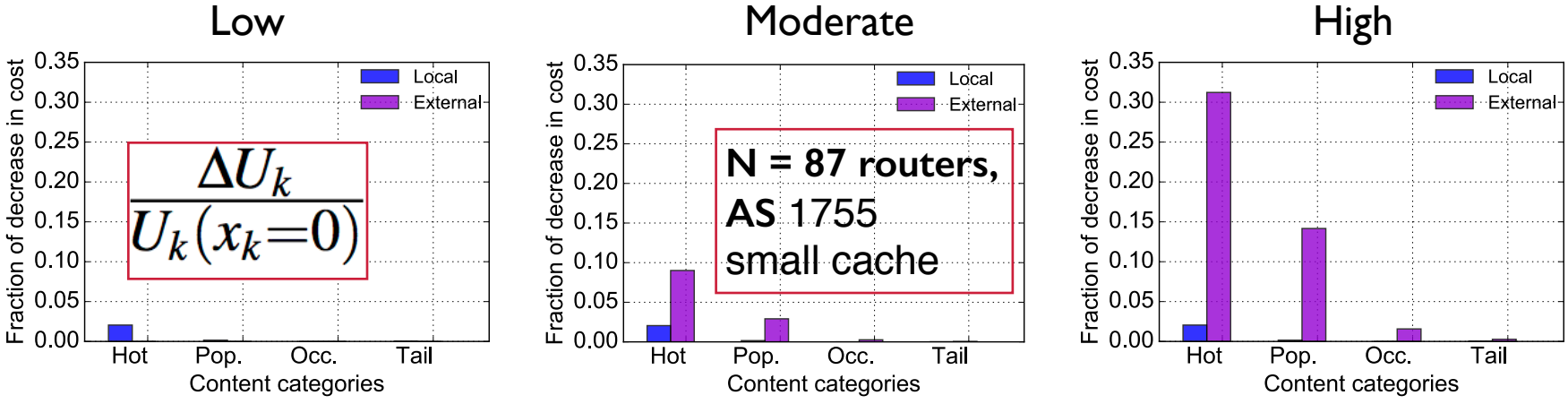
Setting

- ICARUS simulator, available at https://icarus-sim.github.io/
- Realistic AS topologies from Rocketful project (N: # routers 79-315 nodes)
- Content
  - Local content: 30% of the requested contents (K: # content)
  - Popularity: Zipf with parameter 0.8
  - Size: 1 Mb chunks
  - Categories: h*ot (0.1%), popular(1%), occasional(10%), far tail* as in *K. Mokhtarian et al.*
- Network's cache capacity
  - small cache: 10^{-2} K/N
  - large cache 10^{-1} K/N
- NRS updates: 0.01 false alarm, 0.01 false negative

- Impact of inter-AS traffic cost
- Impact of NRS size

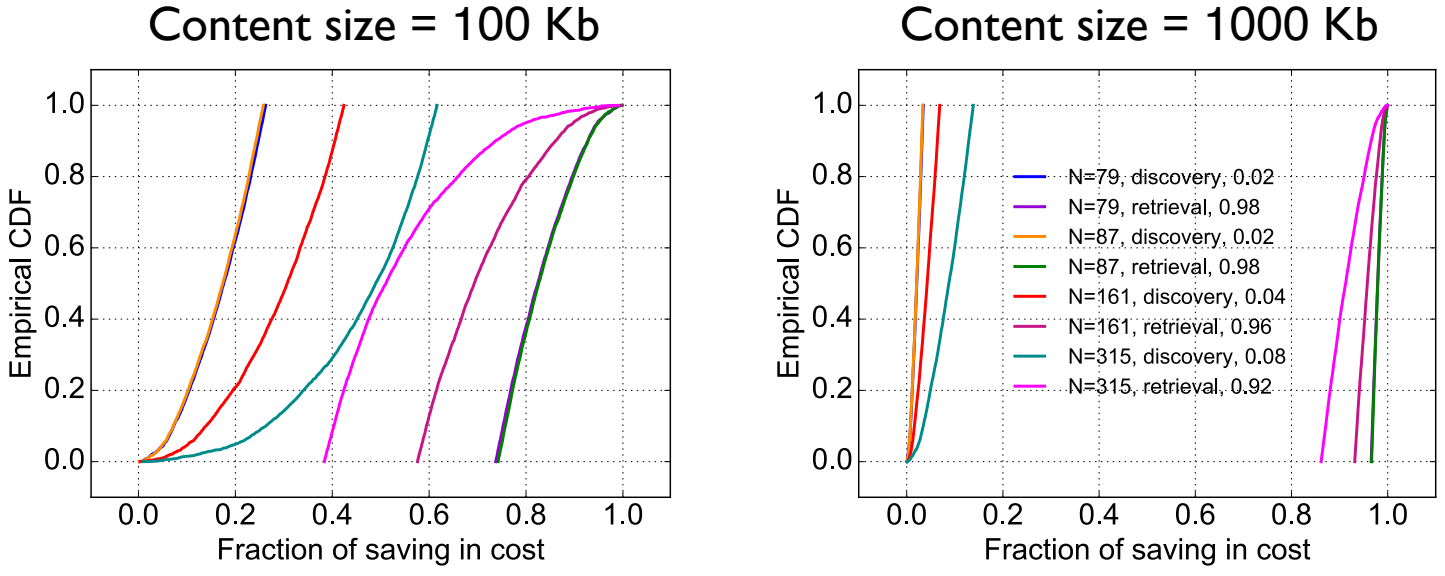
Impact of inter-AS traffic cost: fraction of *decrease in cost*

- Let's assume all routers are synchronized, i.e. no duplicate responses
- *Low (3x), moderate(5x), high (10x)* inter-AS cost compared to intra-AS tx cost



- Higher decrease in cost for more popular content,
- Higher decrease in cost for external content under high inter-AS cost

Fraction of savings in discovery and retrieval

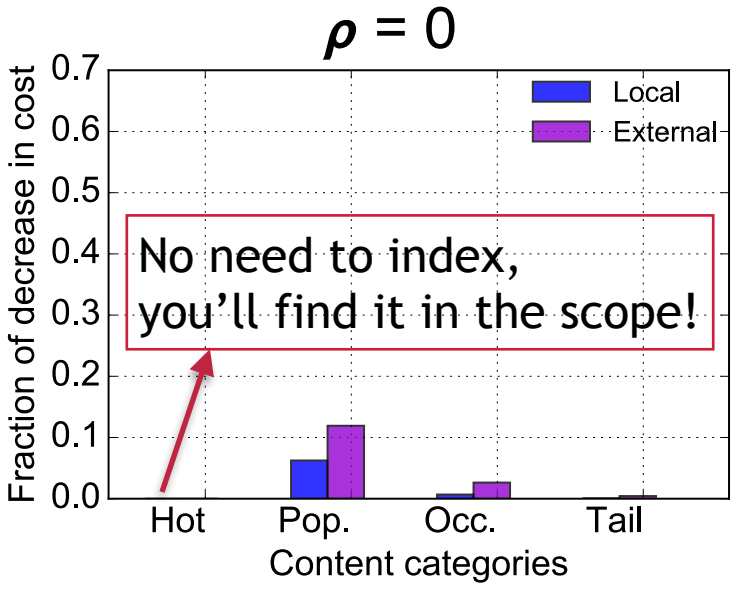


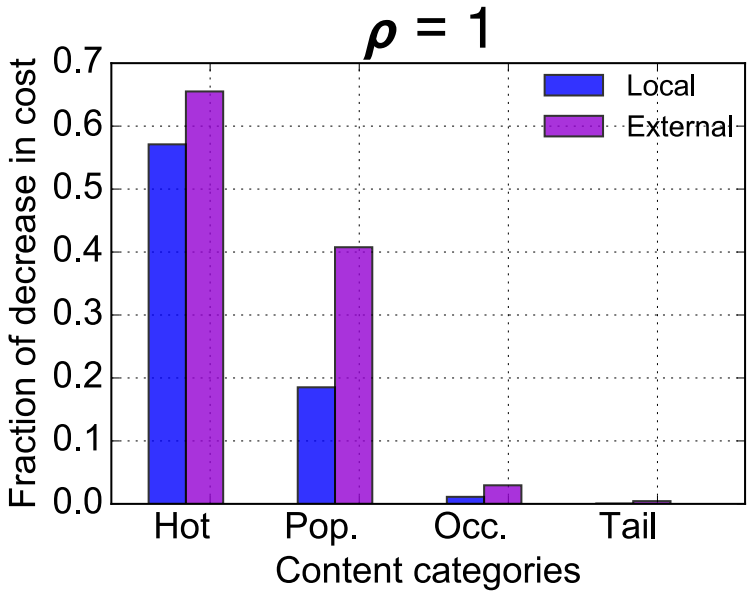
Fraction of saving due to discovery and retrieval :

- comparable for small content size
- retrieval cost dominates for larger content size

More savings for tightly-connected topologies for content discovery as costly FBS is skipped

What if FBS redundancy coefficient ρ is higher?





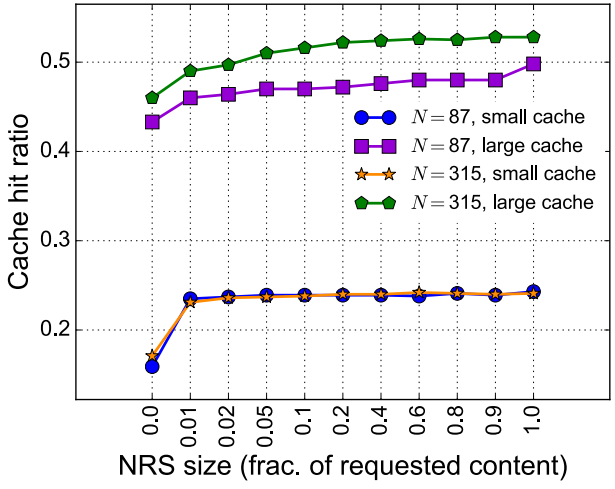
Under *large cache regime*, bandwidth inefficiency due to multiple content transmissions becomes significant, **so does the importance of NRS** (up to 65% savings for external, and 55% for local content)!

Impact of NRS size

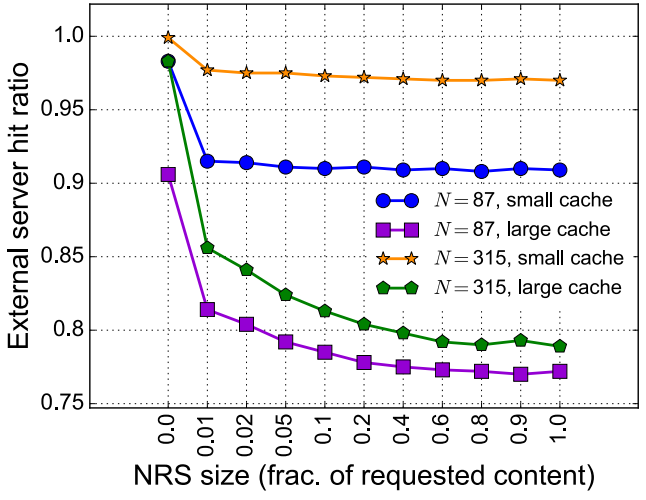
- Content catalogue size: 4×10^4
- Requests: 5×10^5 (warmup period: 10^5 requests)

Impact of NRS size

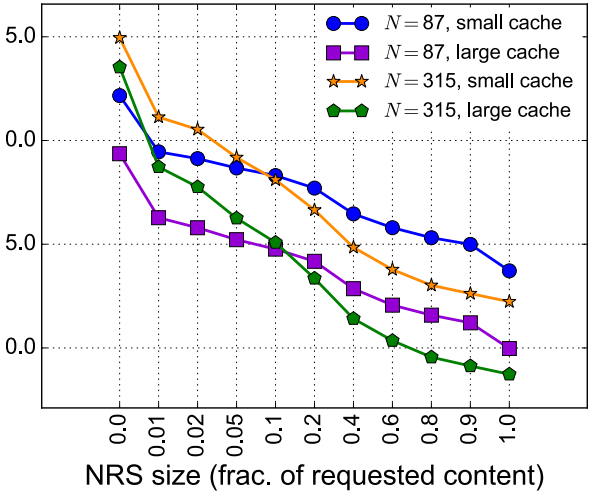
Cache hit



Inter-AS traffic

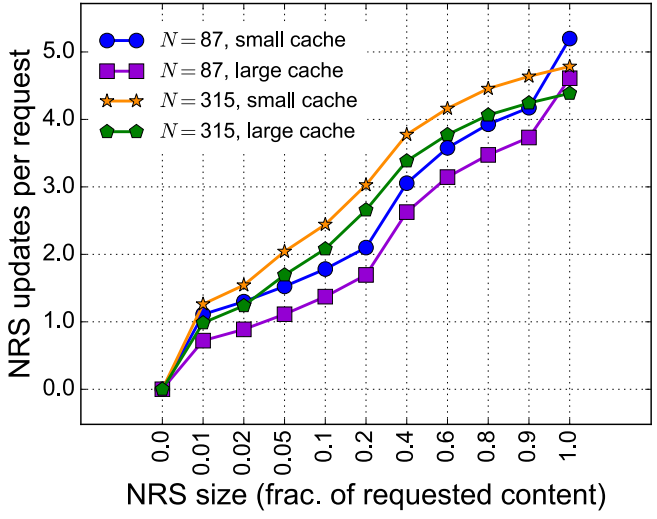


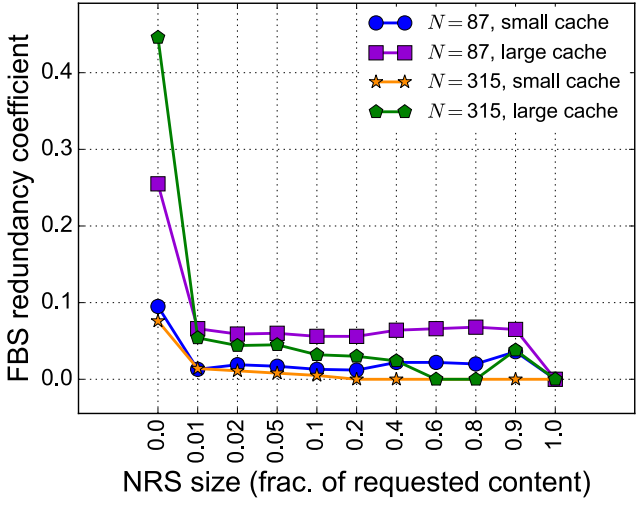
Data access latency



- For small cache, with 1% indexing, cache hit increases from 16-17% to 23%, only marginal increase from 23-24% afterwards
- For larger cache, highest increase in cache hits at 1%, but further increases after 1%
- Highest decrease in inter-domain traffic for the first 1% (intra-AS traffic marginal change)
- Data access latency gets shorter with increasing NRS size due to saving in content discovery time based on FBS

NRS updates and duplicate coefficient





- NRS update cost smaller than average path length due to tolerance to false negatives and positives
- FBS redundancy is about 10% for small cache regime and about 25-45% for larger caches

To sum up

- We proposed a **hybrid name resolution** scheme
  - index the content whose delivery cost decreases the most with NRS,
  - apply content search for other unindexed content
- We determined which items to index based on the calculated cost of content discovery, retrieval, and NRS updates
  - The most significant improvement achieved by **only indexing slight fraction** of content catalogue (~1% for small caches)
  - External content, most popular content
- Future work:
  - Real Internet topology
  - Resolution across AS hierarchy
  - Content discovery can be less aggressive than flooding

Thank you!

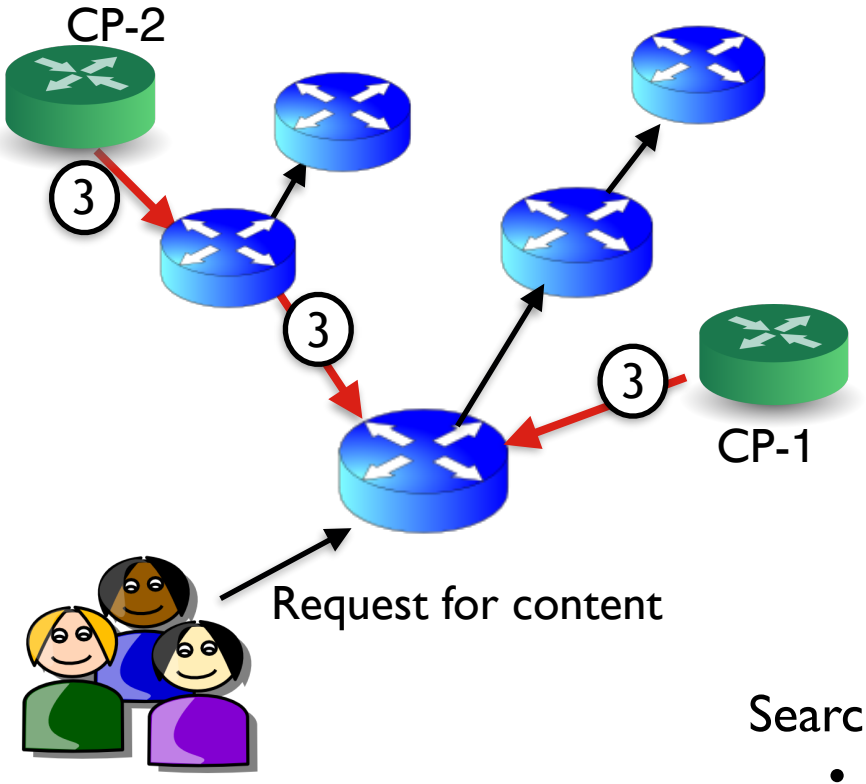
bayhan@hiit.fi

UNIVERSITY OF
CAMBRIDGE

References

- B. Azimdoost, C. Westphal, and H. R. Sadjadpour. The price of updating the control plane in ICNs. *CoRR*, abs/1406.1284, 2014.
- L. Dong, C. Westphal, G. Wang, and J. Wang. Requirements of Name Resolution Service in ICN, work in progress. Internet-Draft draft-dong-icnrg-nrs-requirement-00, IETF Secretariat, July 2016. http://www.ietf.org/internet-drafts/draft-dong-icnrg-nrs-requirement-00.txt
- A. Baid, T. Vu, and D. Raychaudhuri, "Comparing Alternative Approaches for Networking of Named Objects in the Future Internet." in IEEE Workshop on Emerging Design Choices in Name-Oriented Networking (NOMEN), 2012.
- M. F. Bari, S. R. Chowdhury, R. Ahmed, R. Boutaba and B. Mathieuy, "A Survey of Naming and Routing in Information Centric Networks.", IEEE Communications Magazine, Vol. 50, No. 12, P. 44-53.
- K. V. Katsaros, N. Fotiou, X. Vasilakos, C. N. Ververidis, C. Tsilopoulos, G. Xylomenos, and G. C. Polyzos, "On Inter-Domain Name Resolution for Information-Centric Networks," in Proc. IFIP-TC6 Networking Conference, 2012.
- D. Zhang, H. Liu, Routing and Name Resolution in Information-Centric Networks, 22nd International Conference on Computer Communications and Networks (ICCCN), 2013.
- S. Sevilla, P. Mahadevan, J. Garcia-Luna-Aceves, "iDNS: Enabling Information Centric Networking Through The DNS." Name Oriented Mobility, 2014.
- K. Mokhtarian, H. Jacobsen, Coordinated caching in planet-scale CDNs: Analysis of feasibility and benefits, INFORM 2016.

Operation without an NRS in ICN



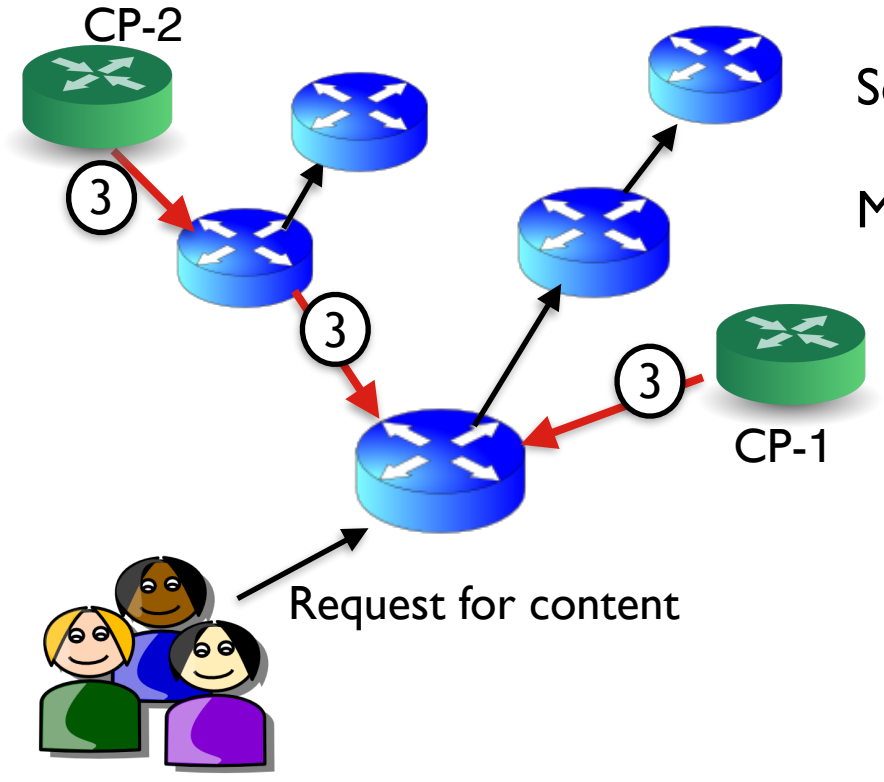
Search for content in the neighborhood

- inefficiency in content discovery

Multiple content providers routing content

- inefficiency in content retrieval

Operation without an NRS in ICN



Search for content in the neighborhood
- inefficiency in content discovery

Multiple content providers routing content
- inefficiency in content retrieval

Bandwidth inefficiency

Indexing for minimum content delivery cost

Decision variable:

- index content k or not, x_k

Objective:

- minimize the expected cost of content delivery over all requests from within this AS

Constraints:

- subject to total number of items to be indexed Kw