# Cost-aware caching

A. Araldo
araldo@lri.fr

F. Martignon
fabio.martignon@lri.fr

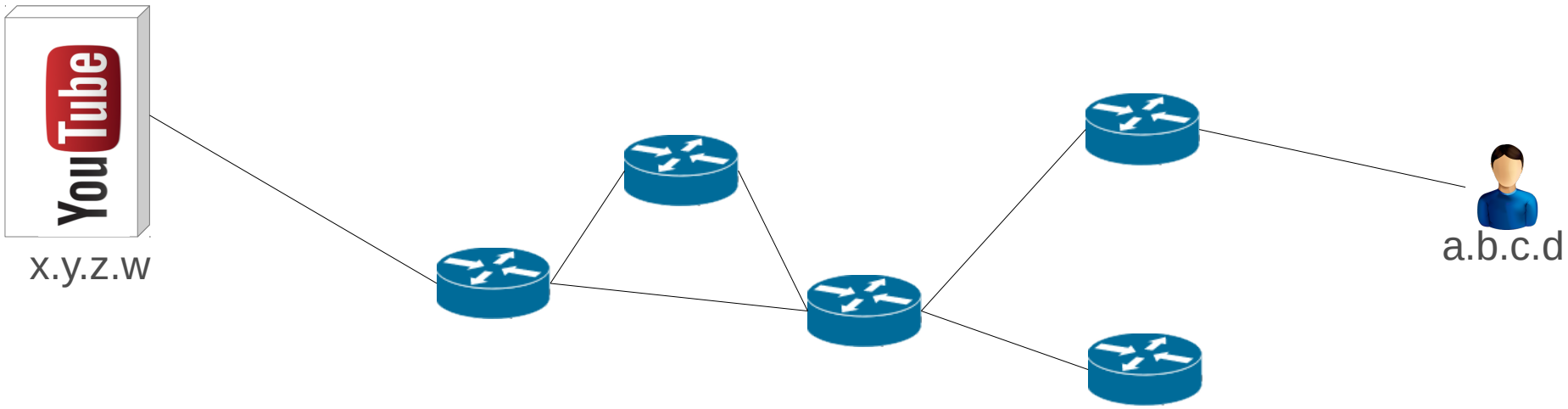
M. Mangili
mangili@lri.fr

D. Rossi
dario.rossi@enst.fr

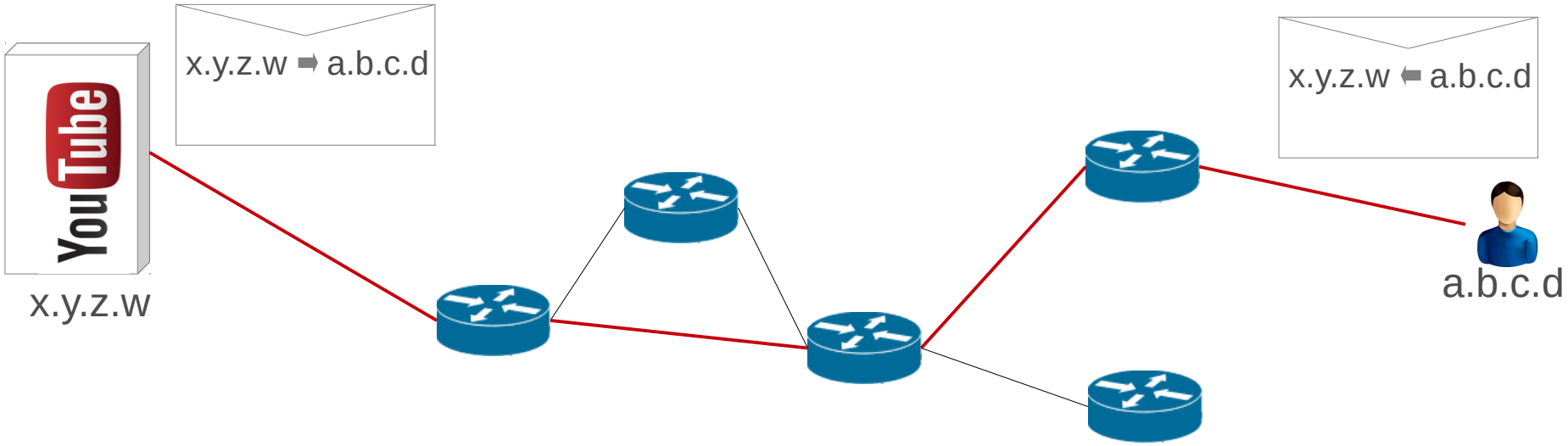
# Outline

- This talk about:
  - Operational costs of ISPs
  - Reduction through caching
  - Information Centric Networks
- Our contribution
  - Classic caching is cost-ineffective
  - Optimal caching for ISP cost reduction (in theory)
  - Cost-aware cache policy (real implementation)

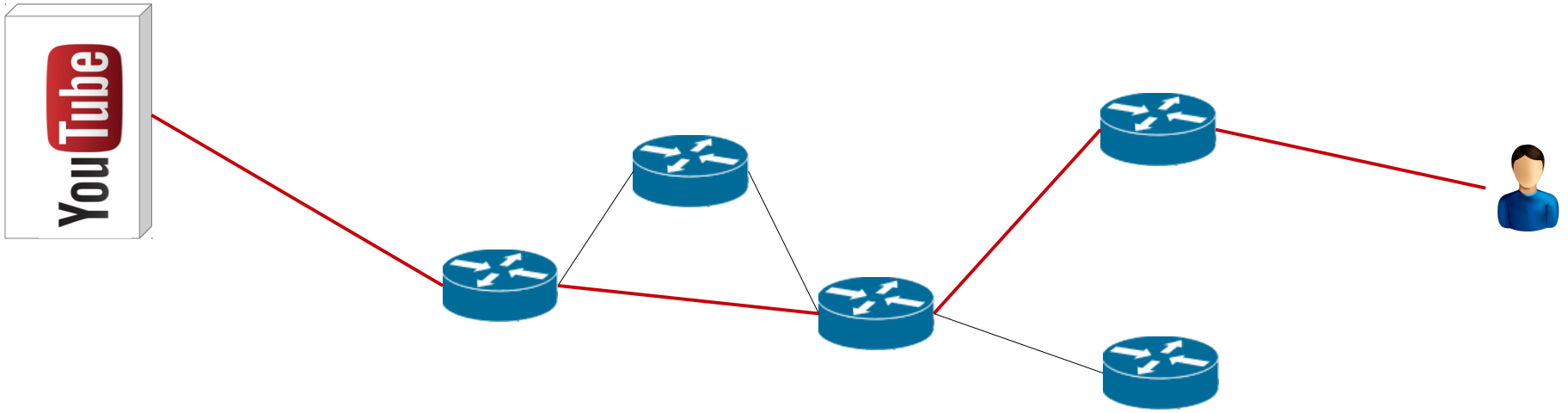
# Current Host Centric Network



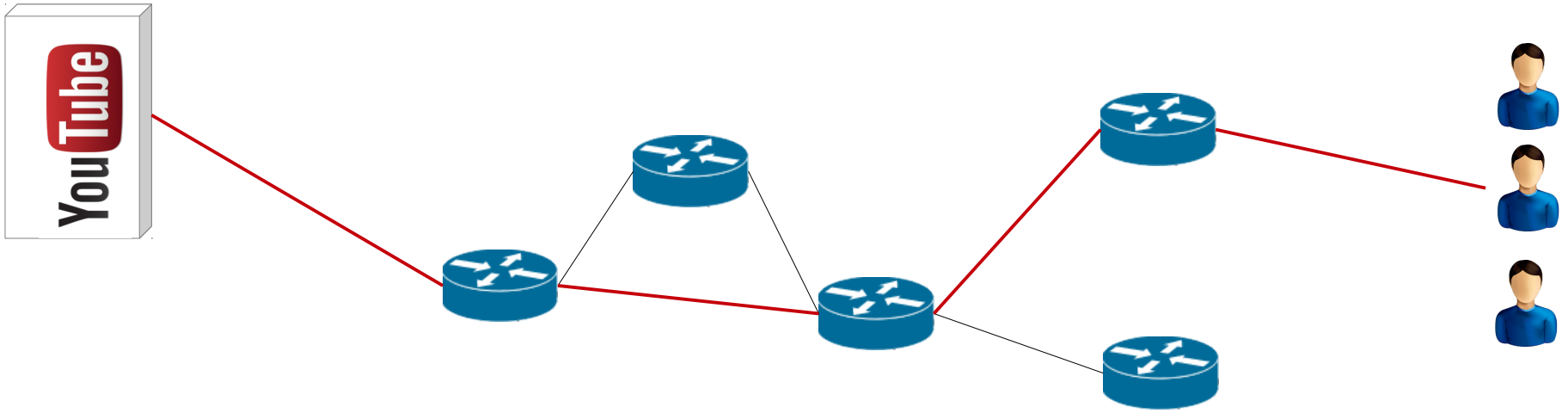
# Current Host Centric Network



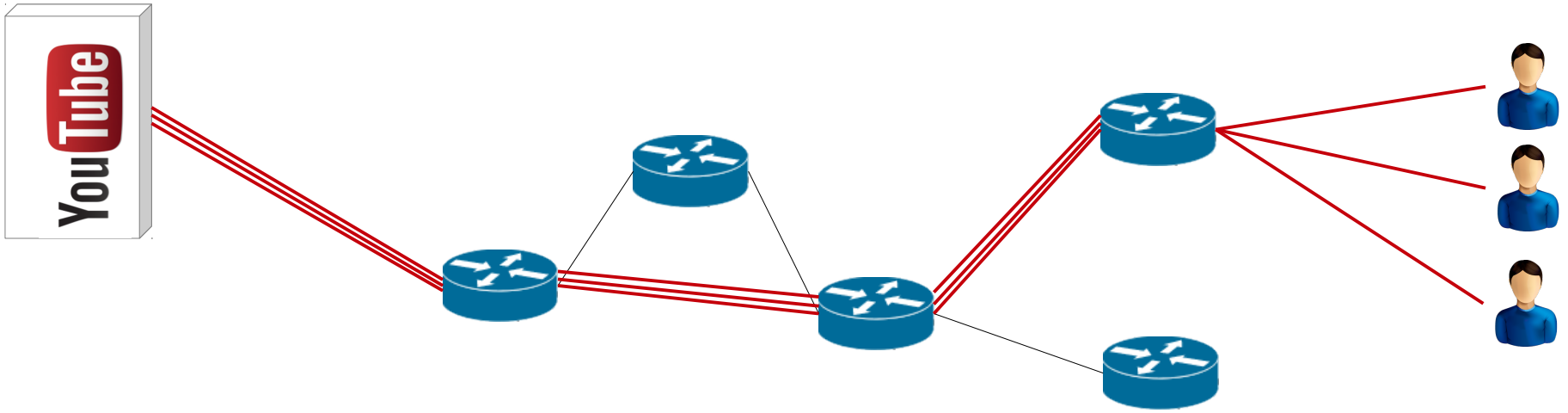
# Current Host Centric Network



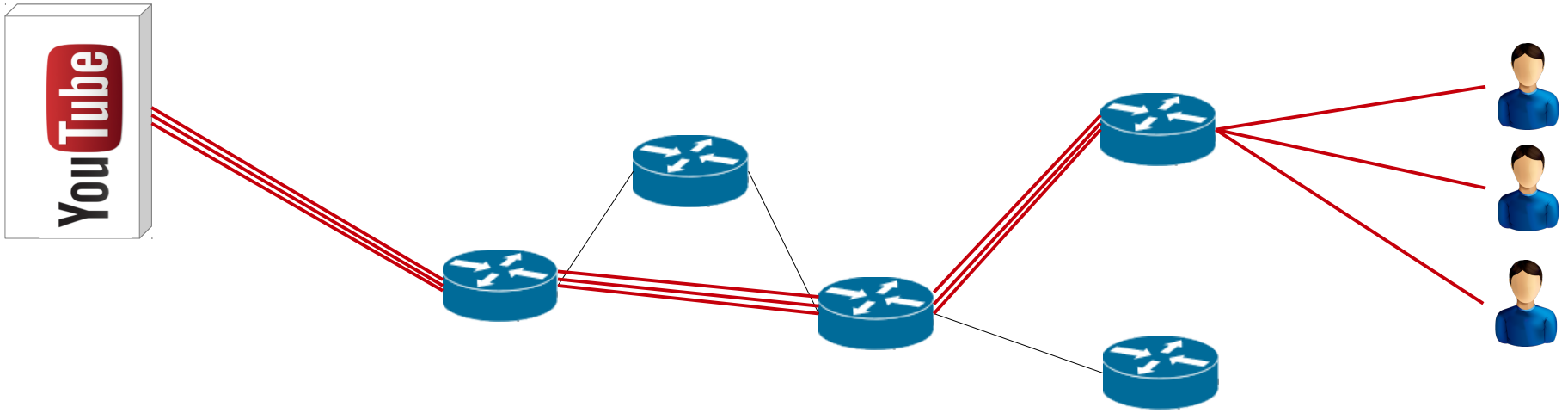
# Current Host Centric Network



# Current Host Centric Network



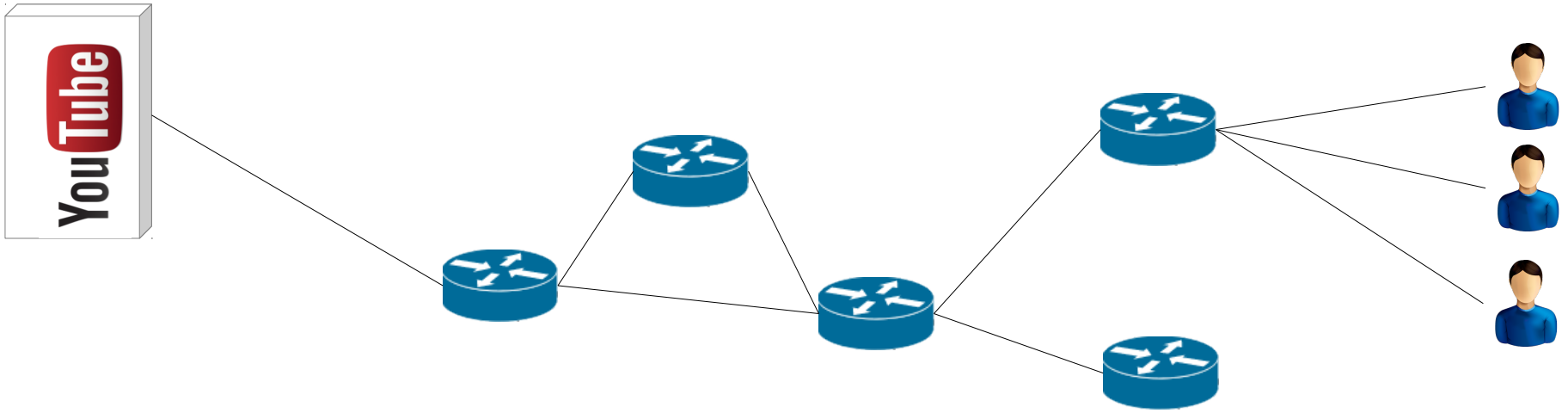
# Current Host Centric Network



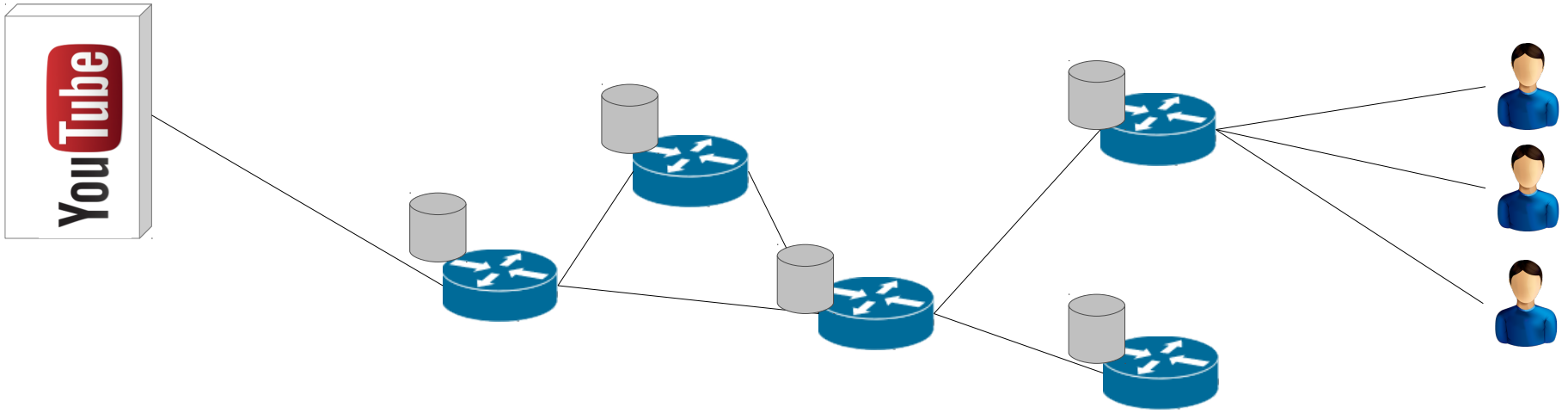

- Redundancy
- Wasted bandwidth

Van Jacobson

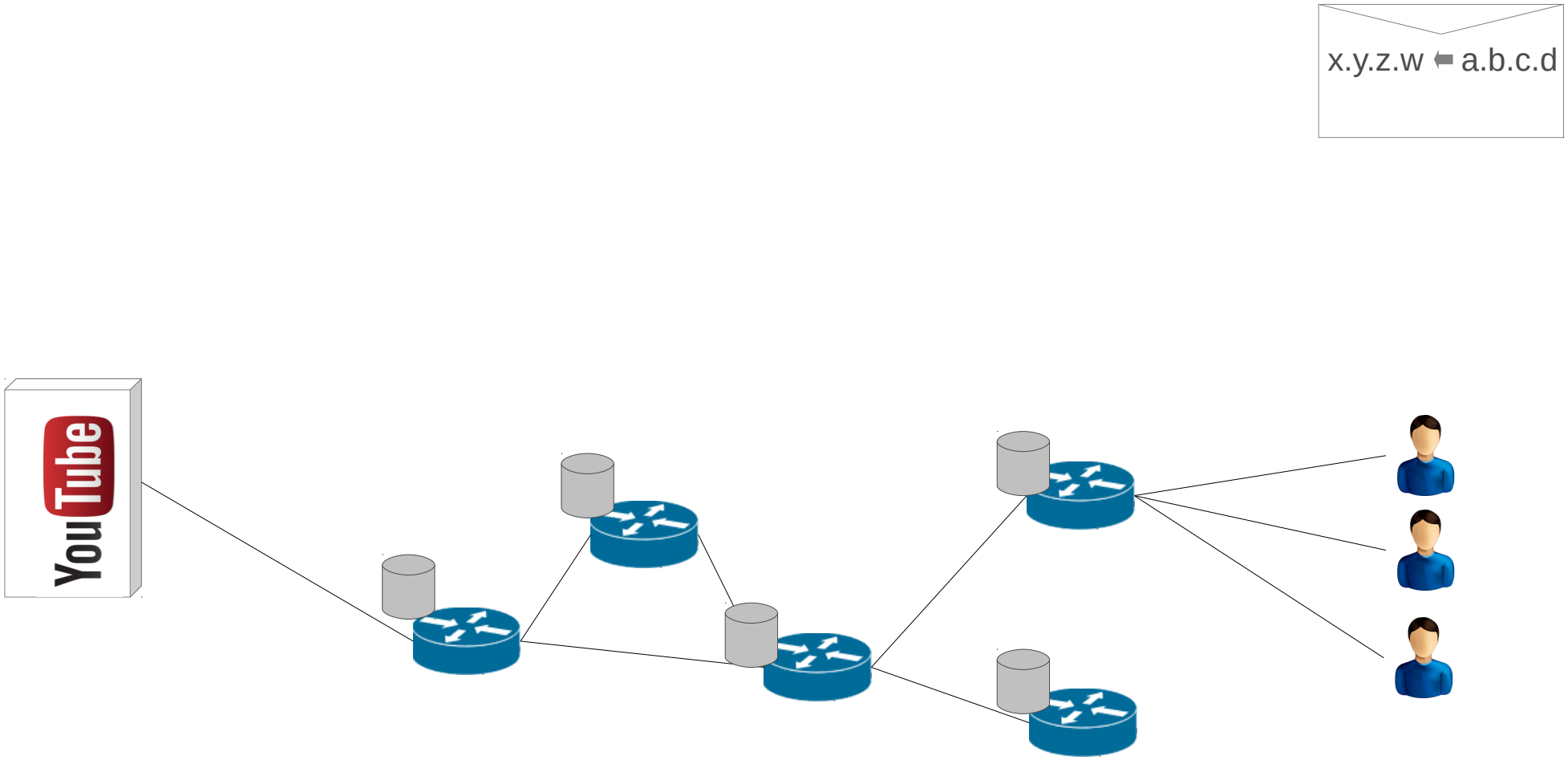
# Information Centric Network (ICN)



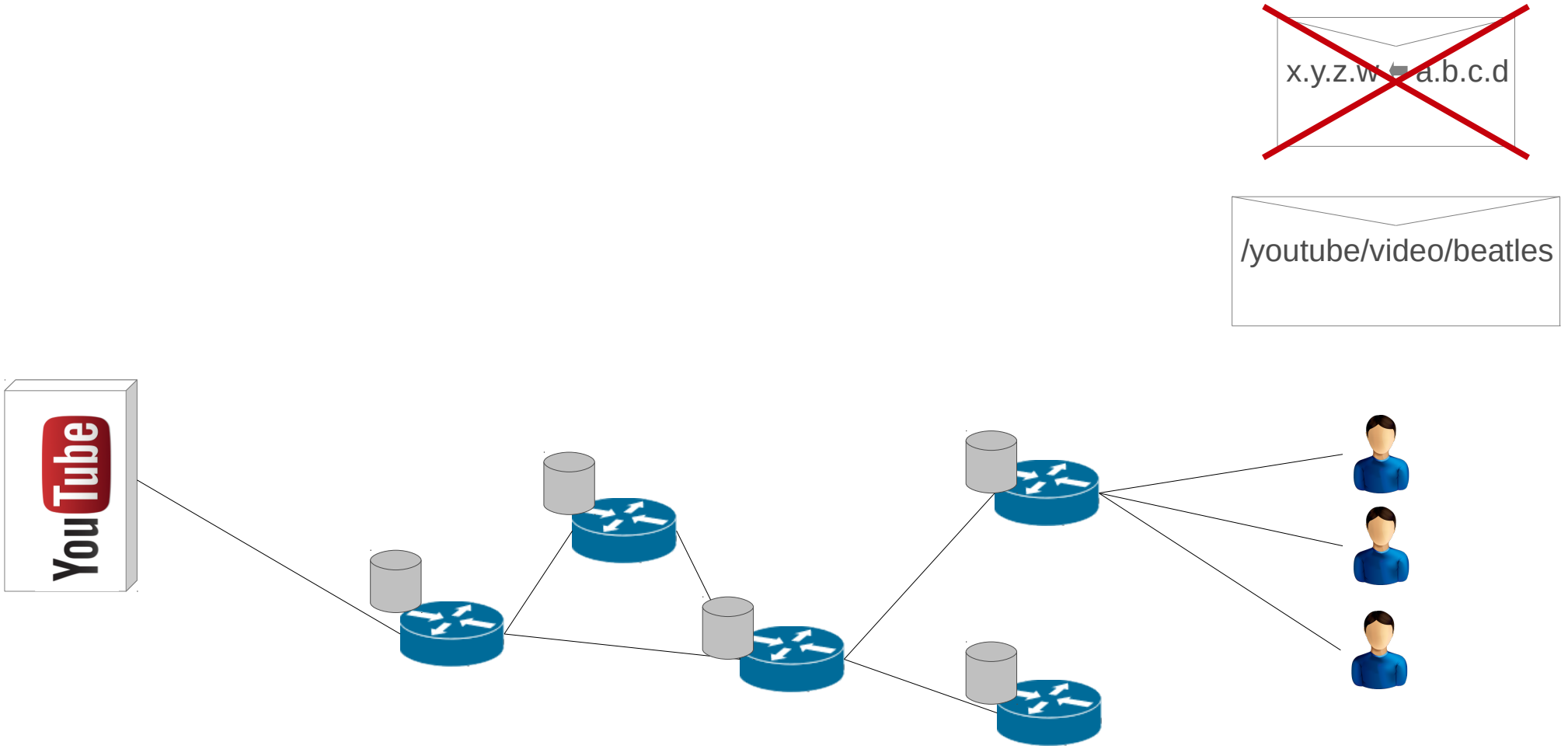
# Information Centric Network (ICN)



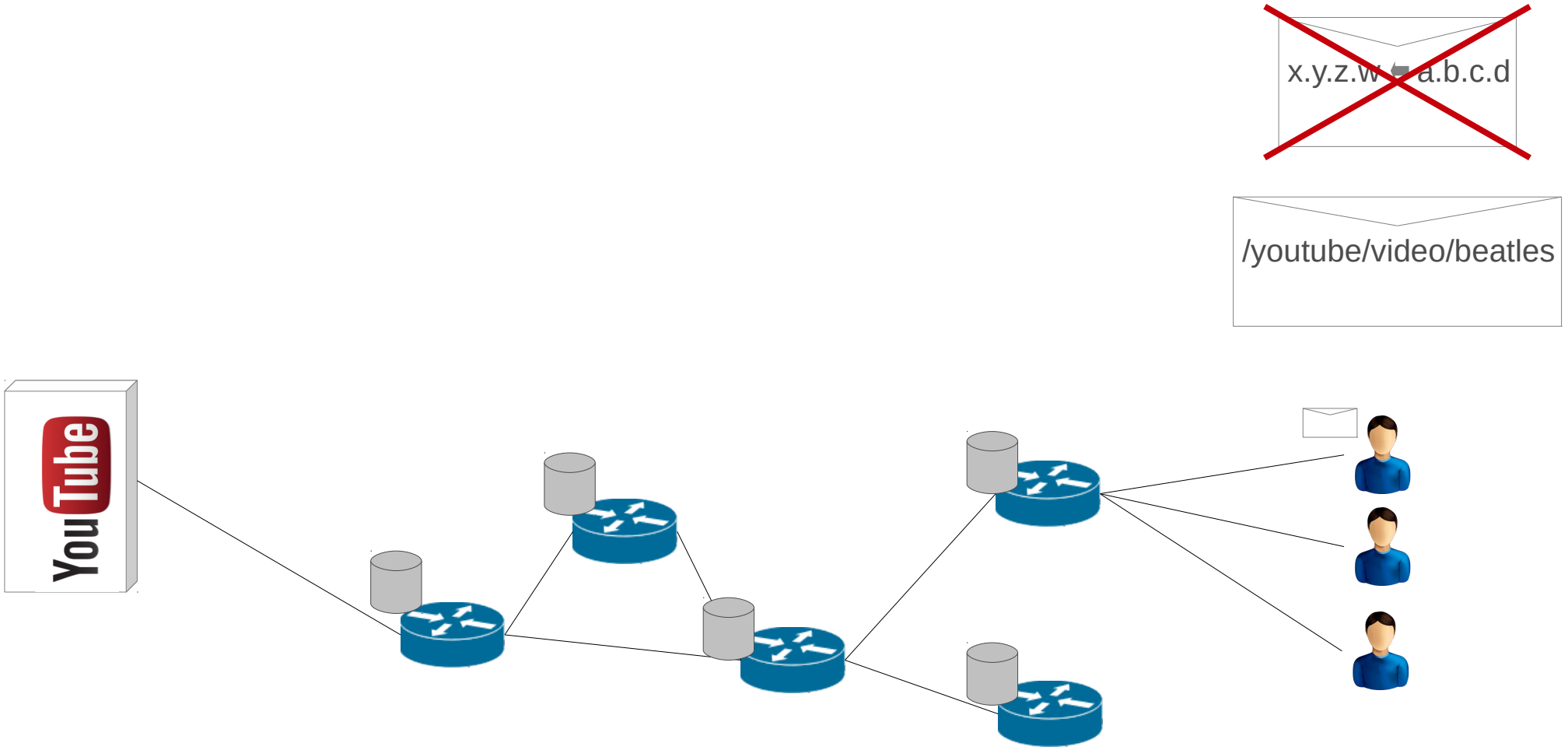
# Information Centric Network (ICN)



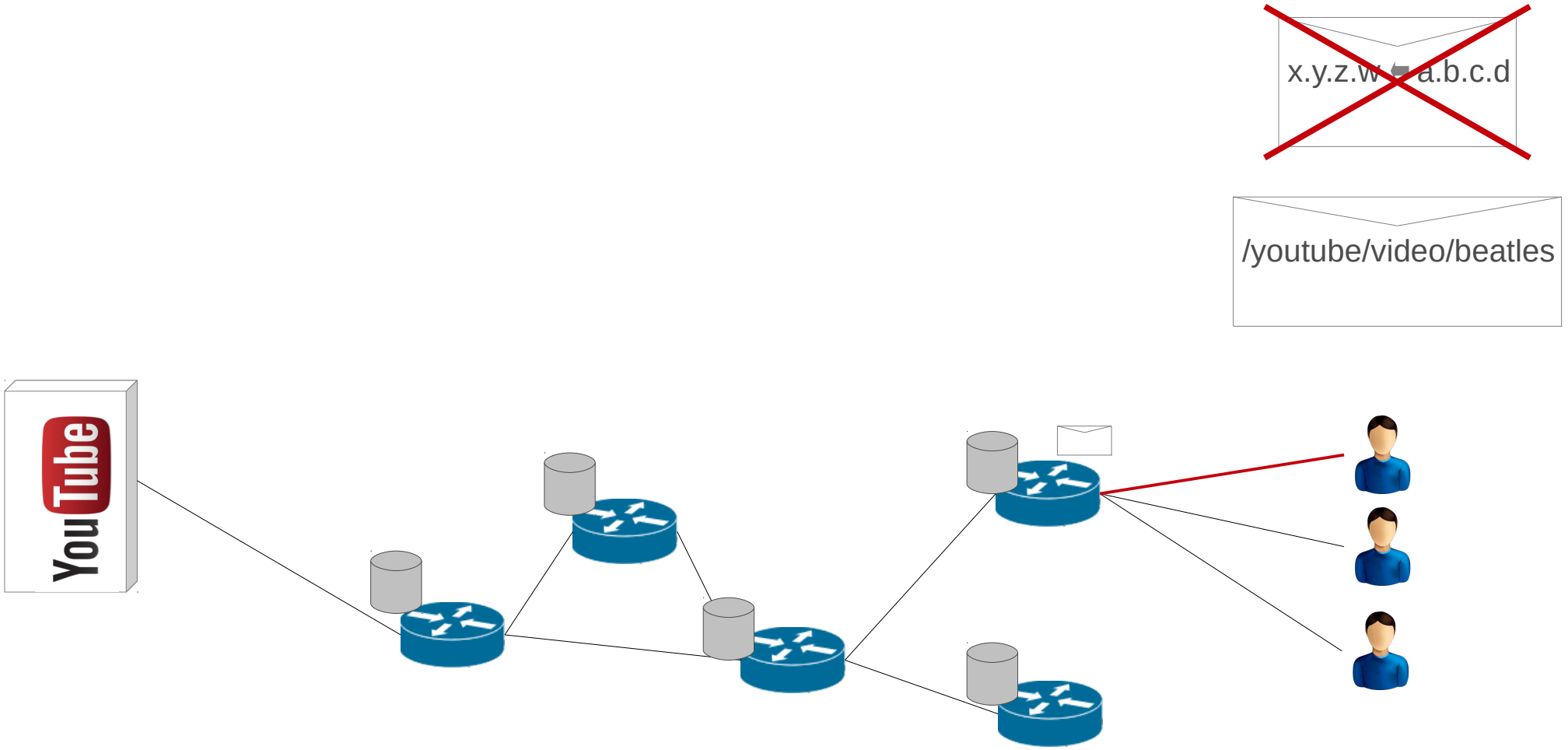
# Information Centric Network (ICN)



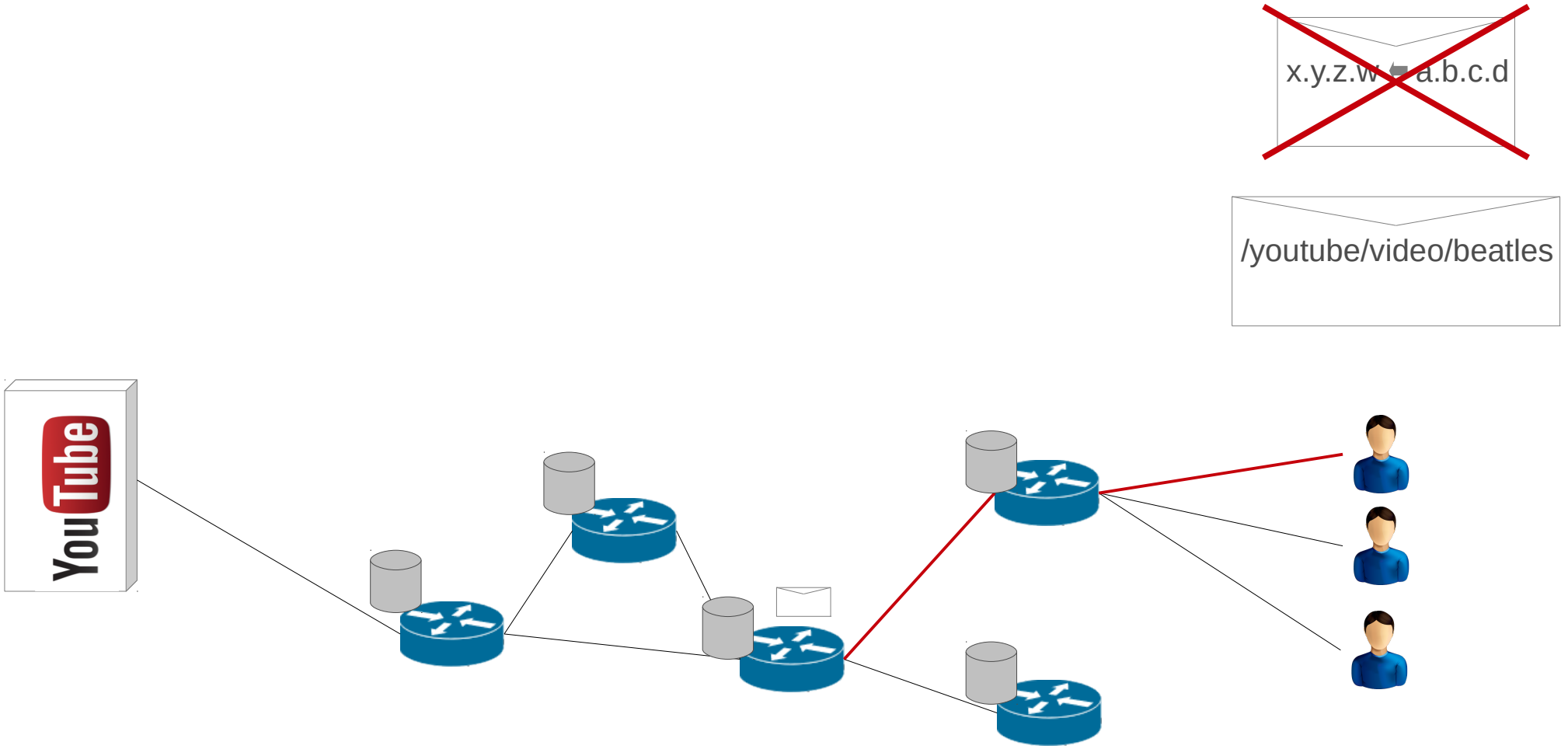
# Information Centric Network (ICN)



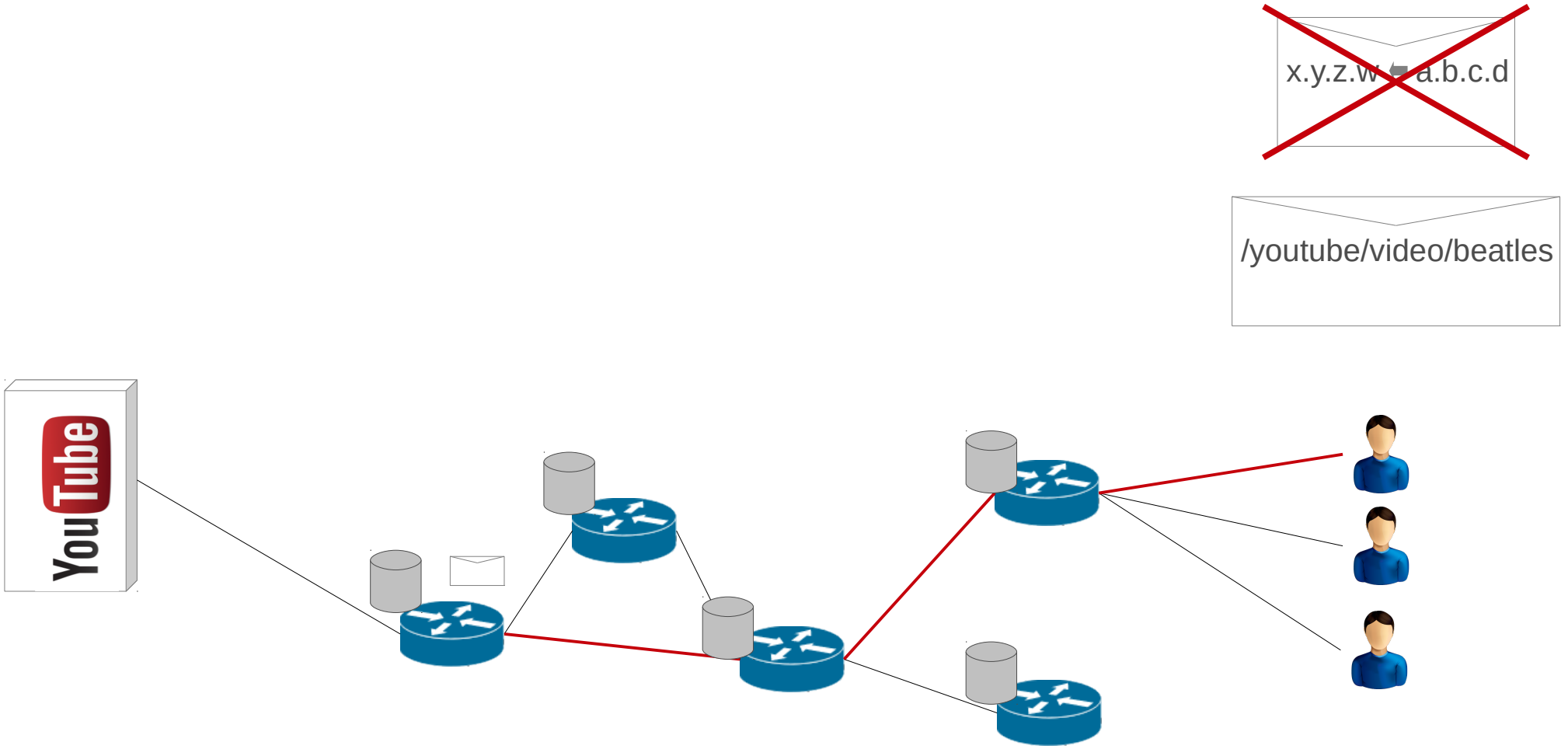
# Information Centric Network (ICN)



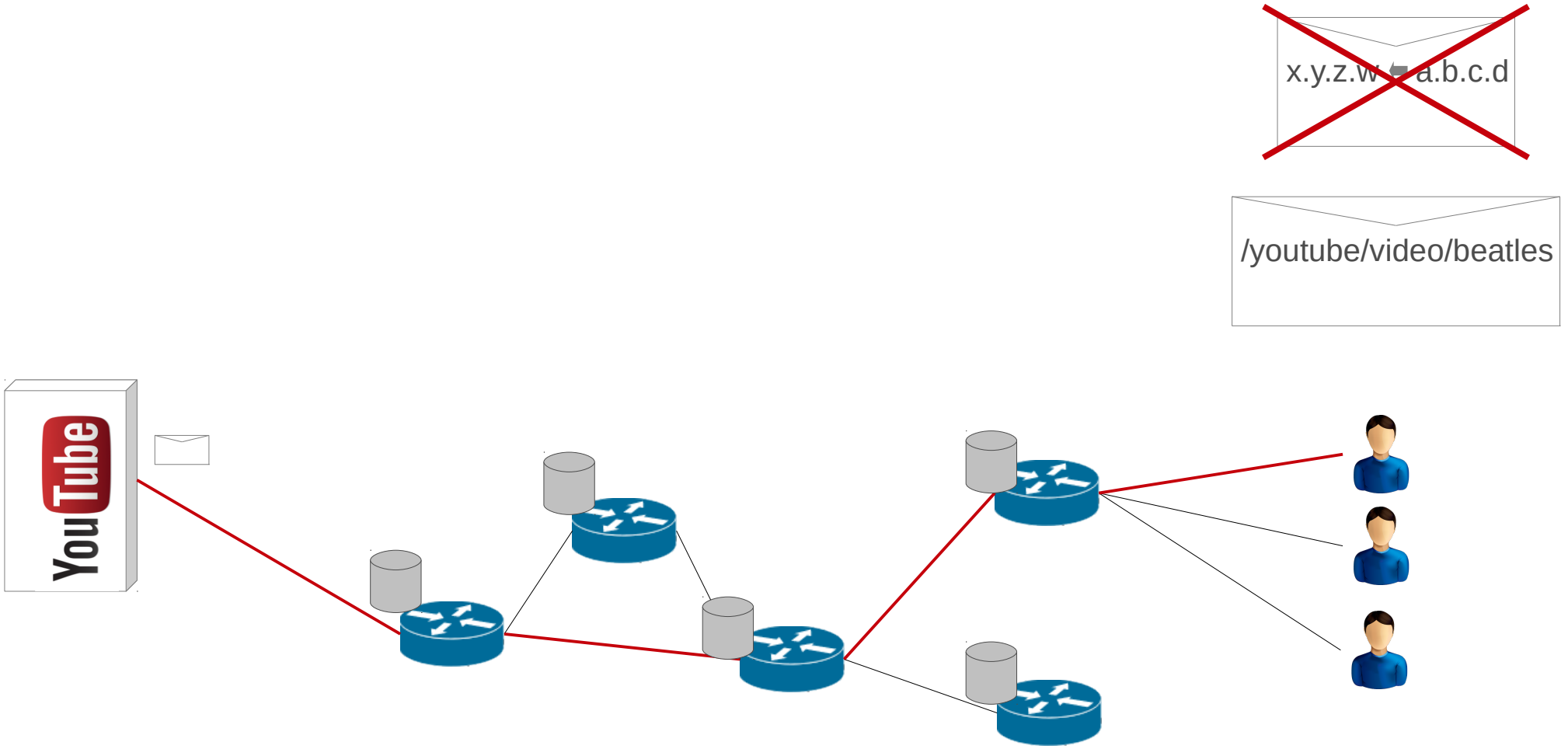
# Information Centric Network (ICN)



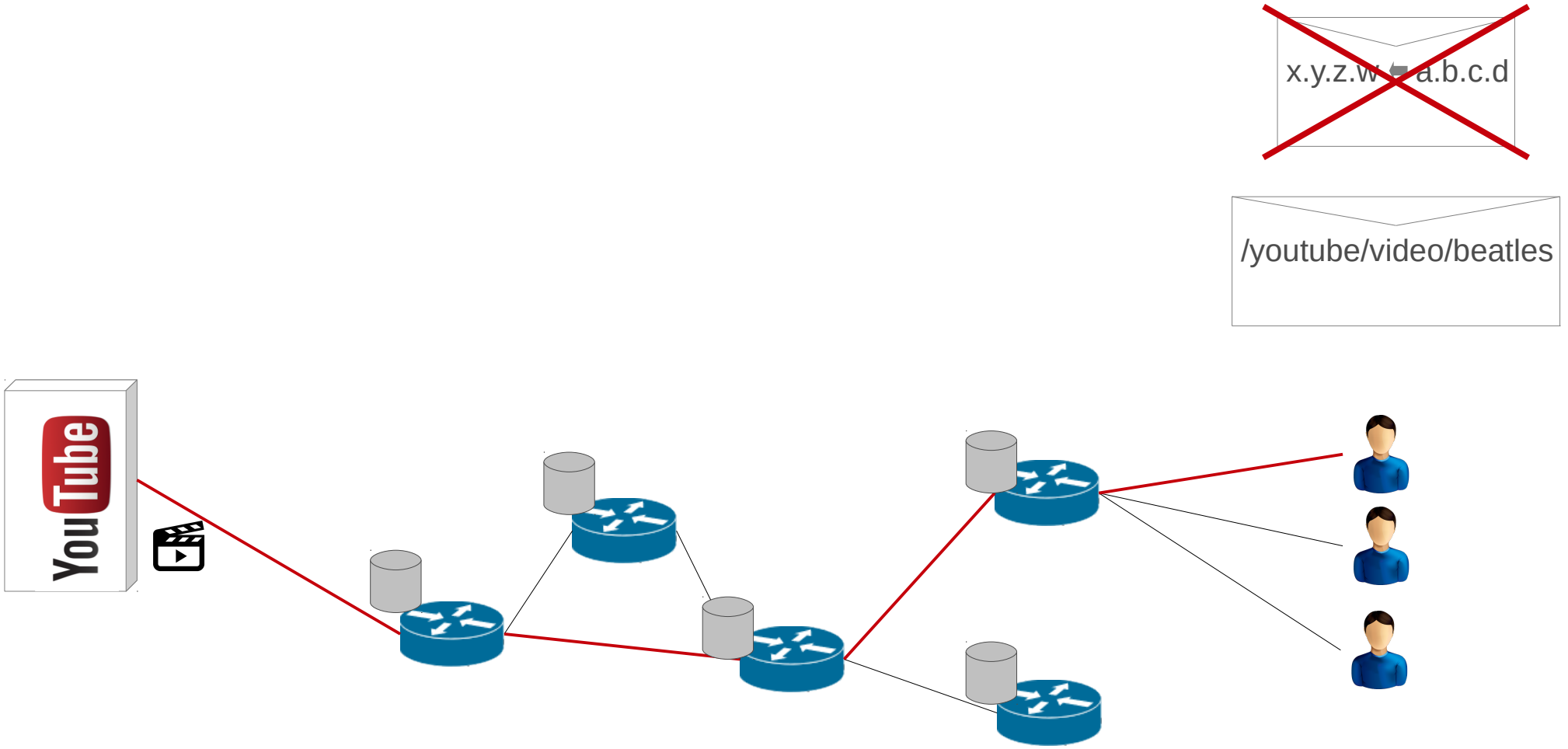
# Information Centric Network (ICN)



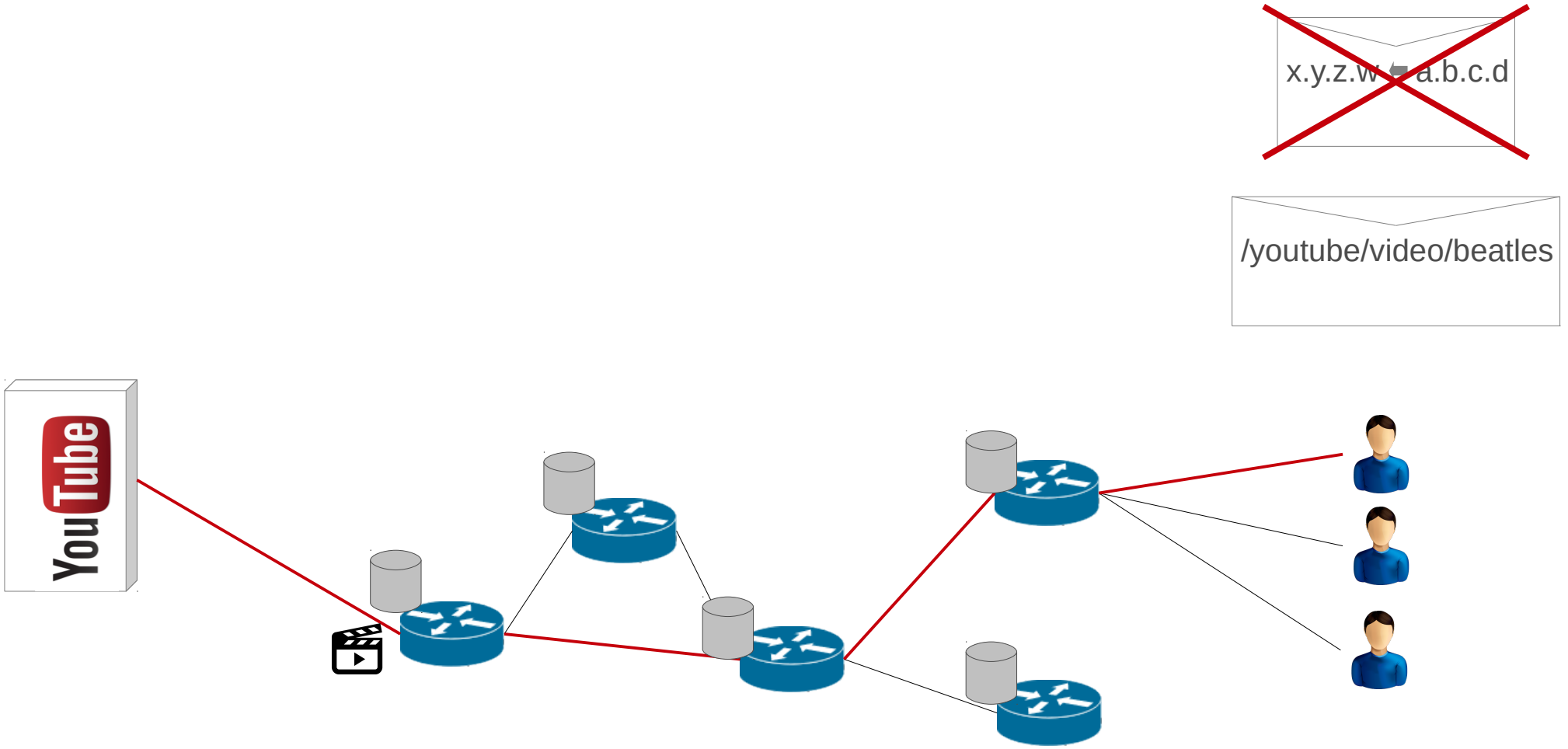
# Information Centric Network (ICN)



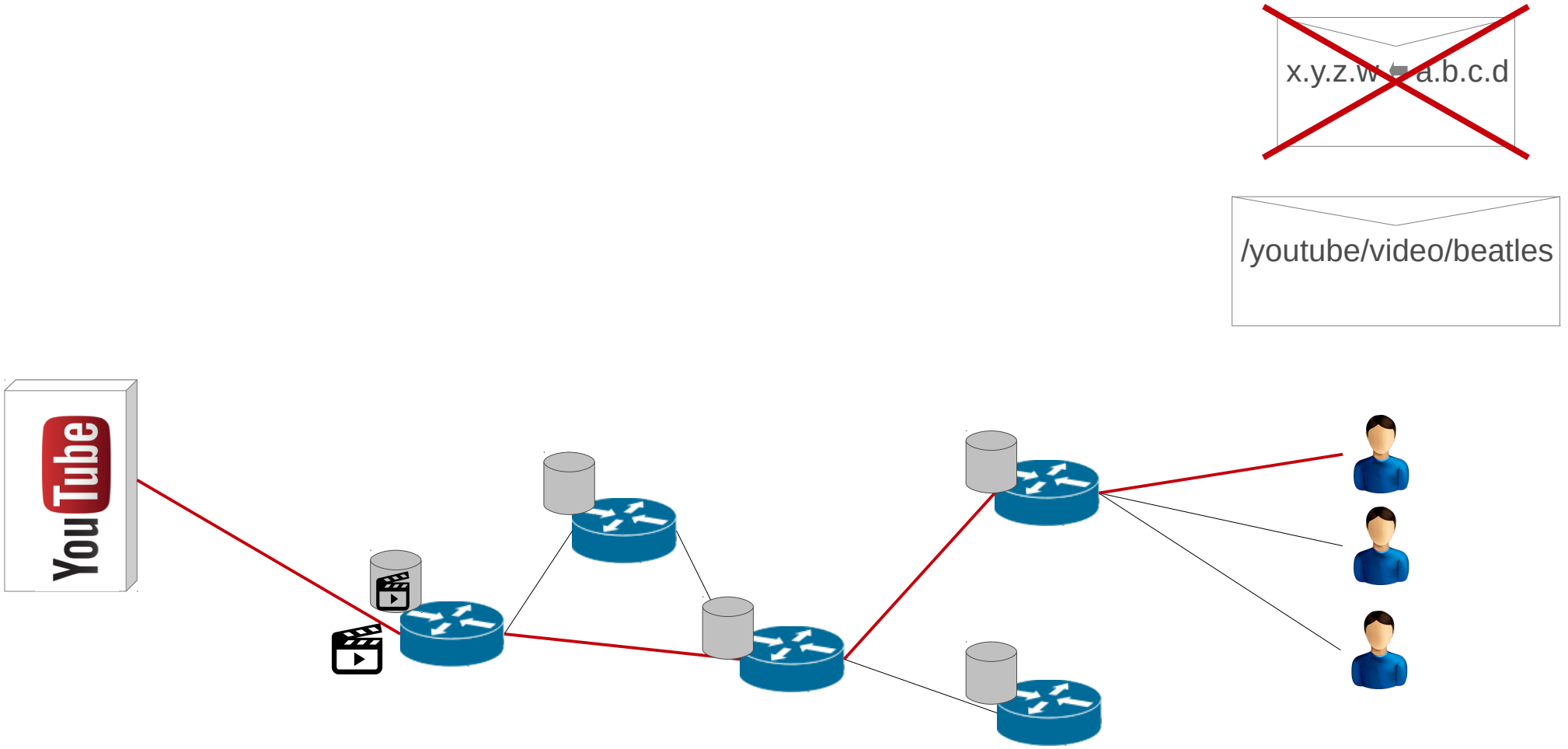
# Information Centric Network (ICN)



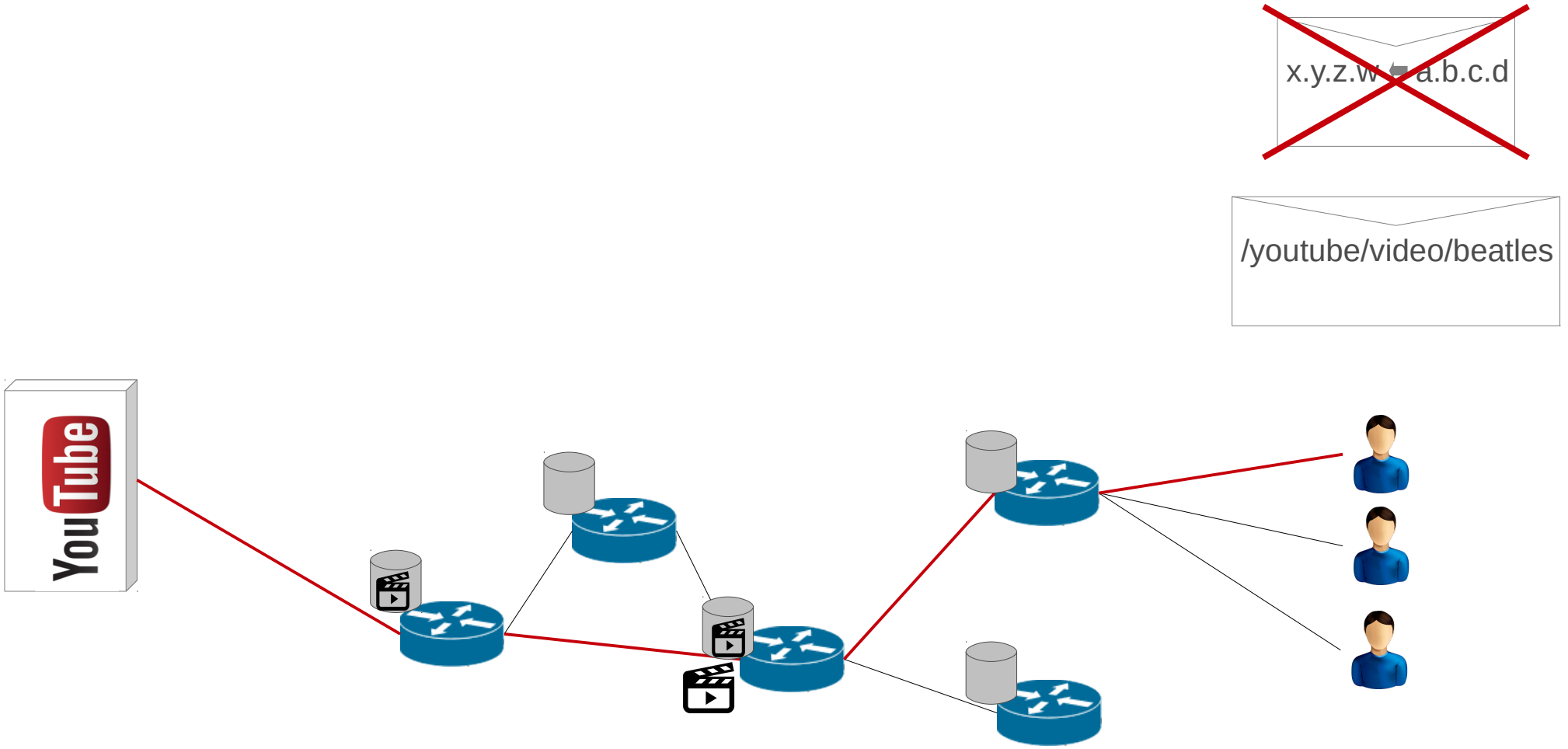
# Information Centric Network (ICN)



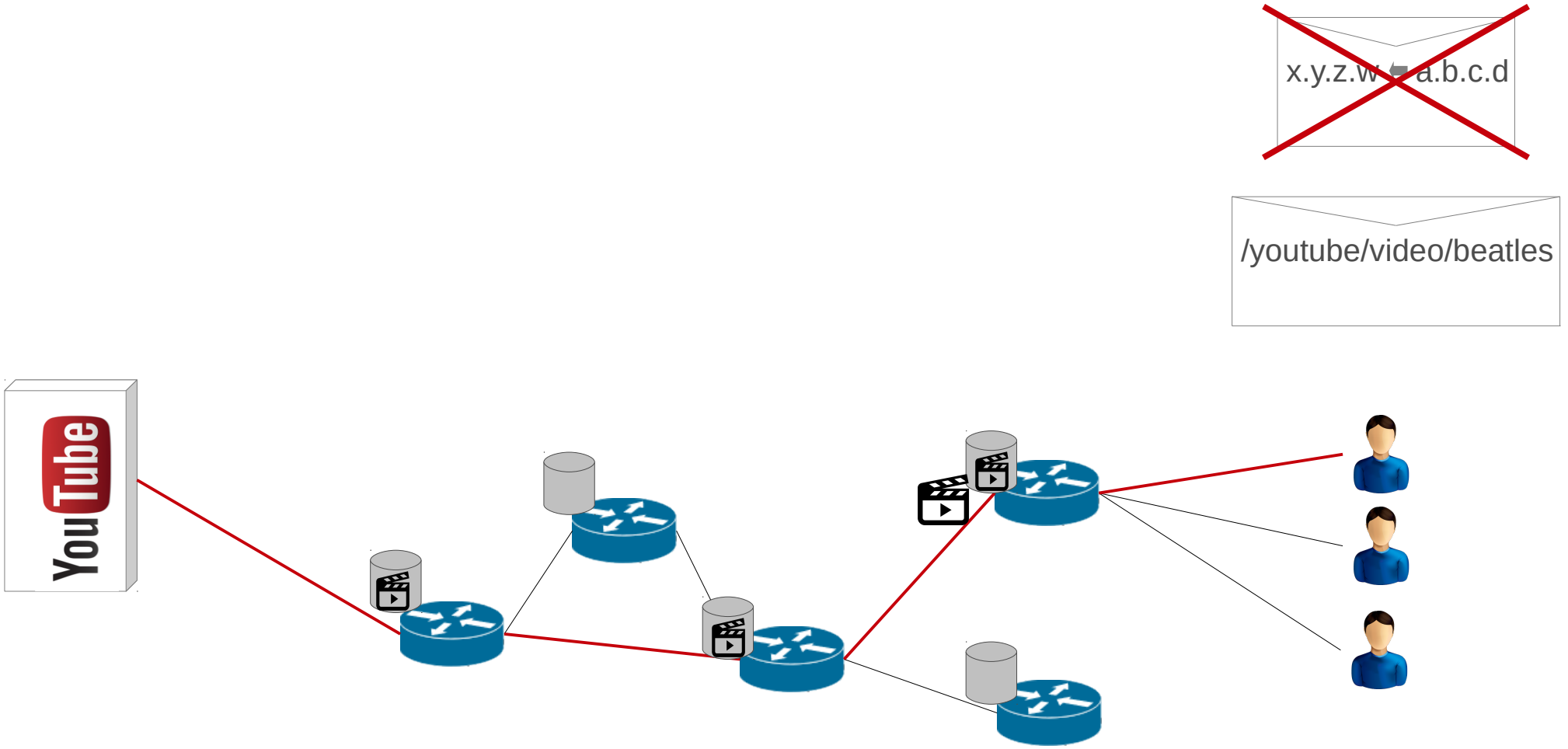
# Information Centric Network (ICN)



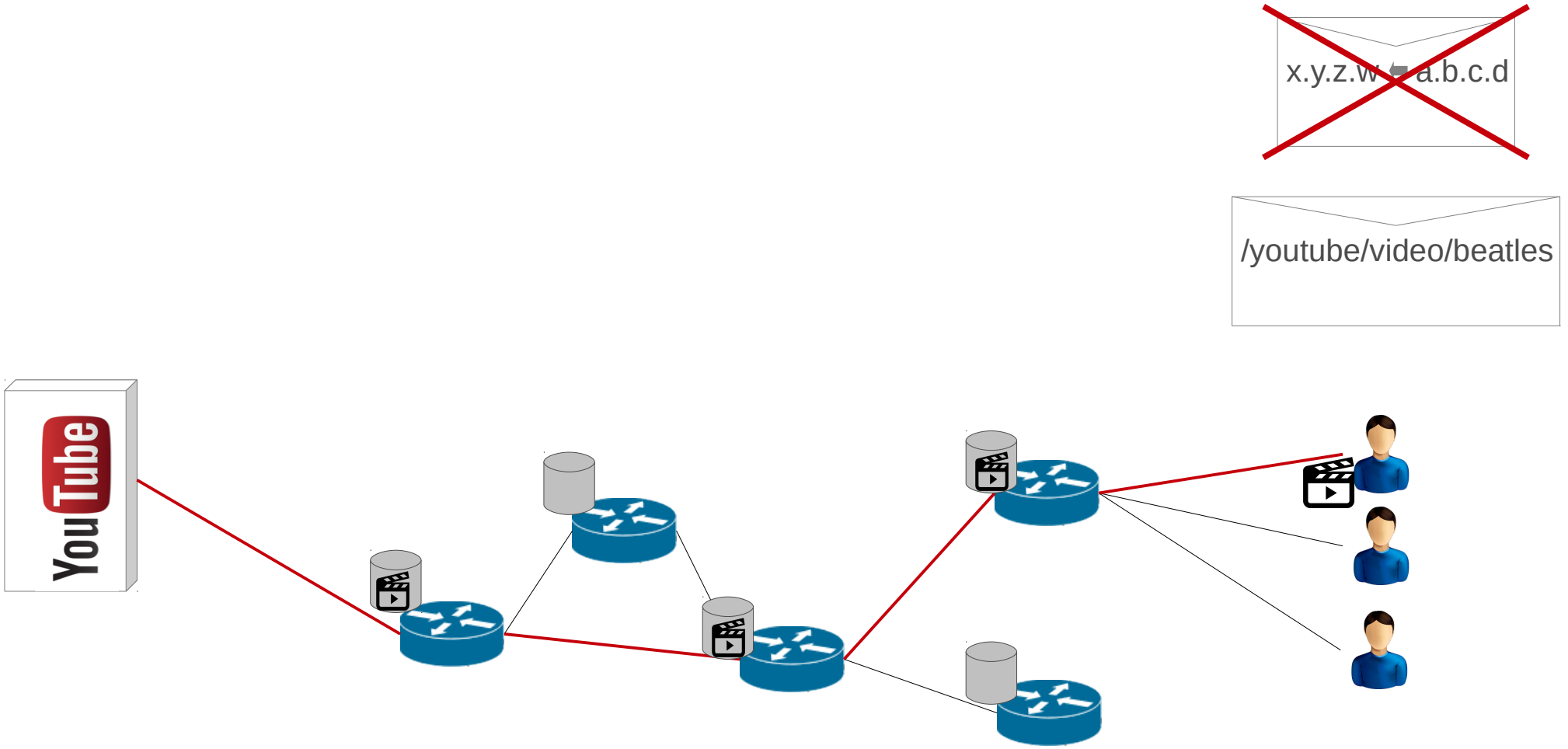
# Information Centric Network (ICN)



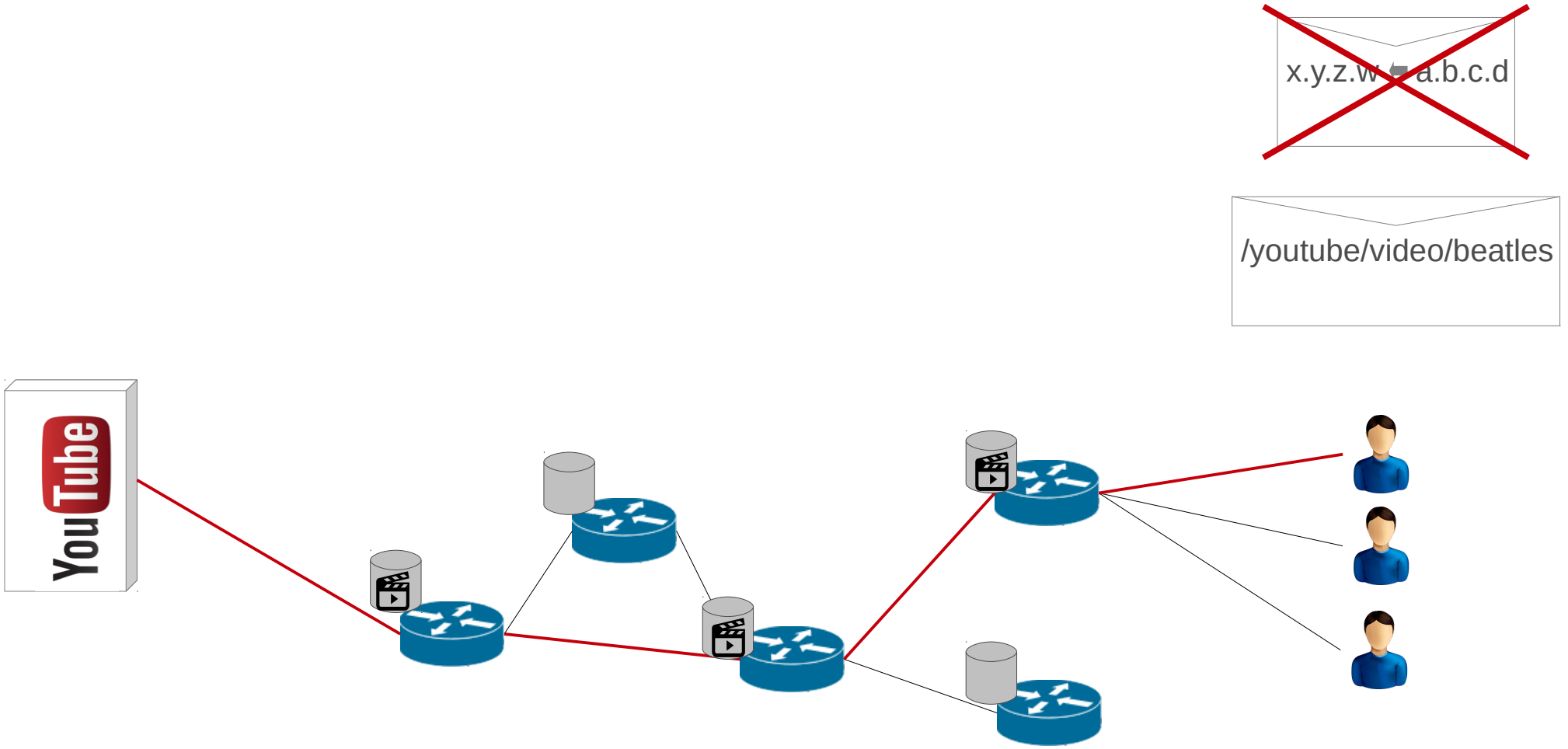
# Information Centric Network (ICN)



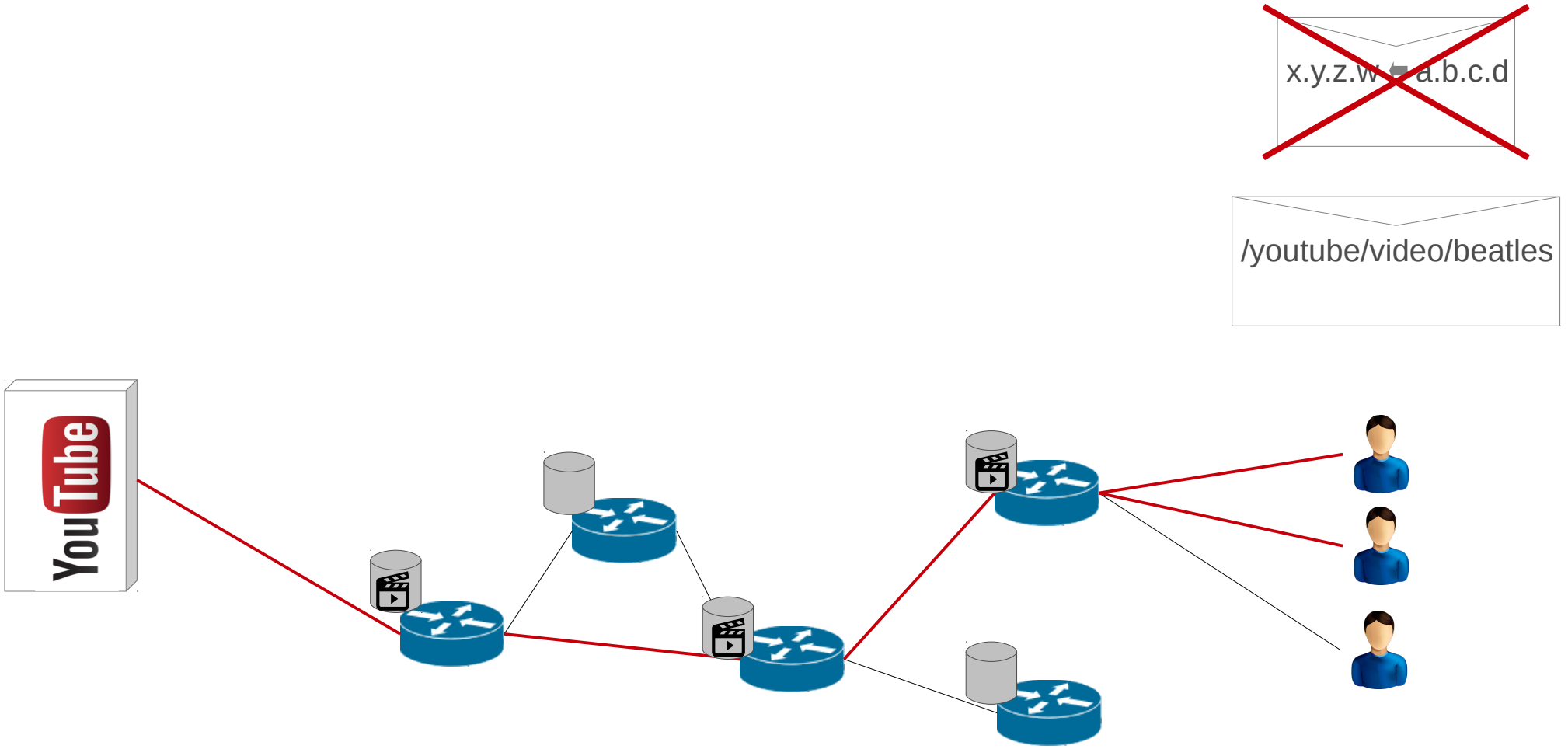
# Information Centric Network (ICN)



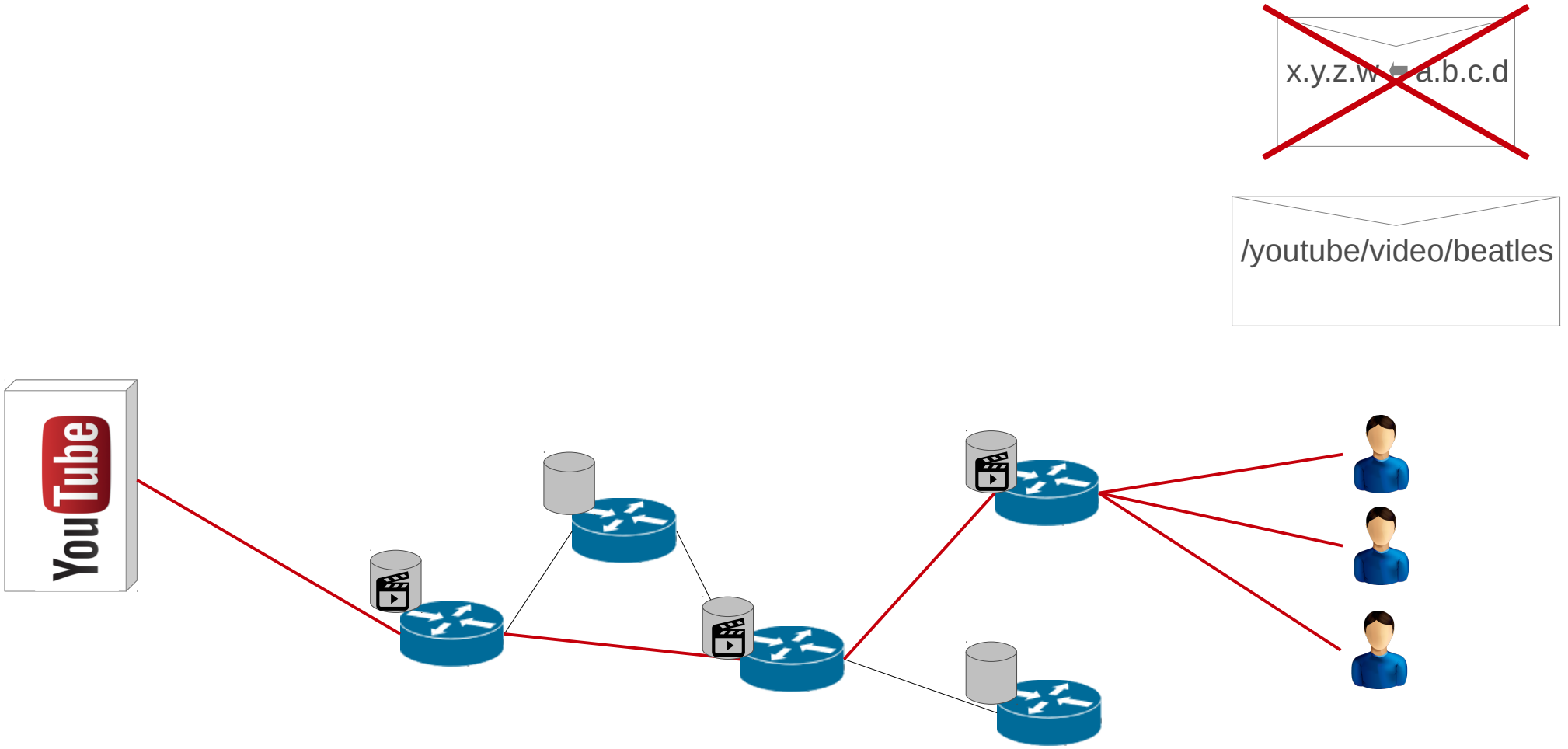
# Information Centric Network (ICN)



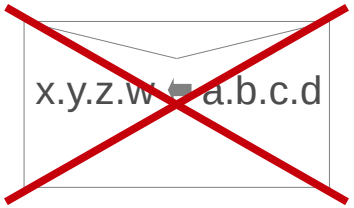
# Information Centric Network (ICN)

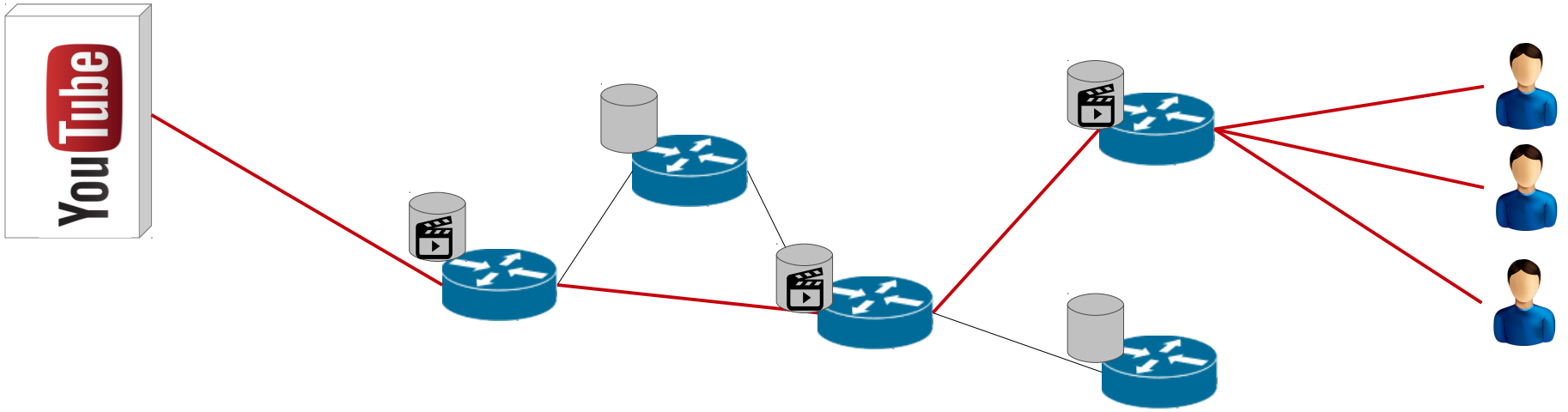


# Information Centric Network (ICN)



# Information Centric Network (ICN)






- Efficient bandwidth usage
- Reduced network load
- Reduced delay
- Better QoE

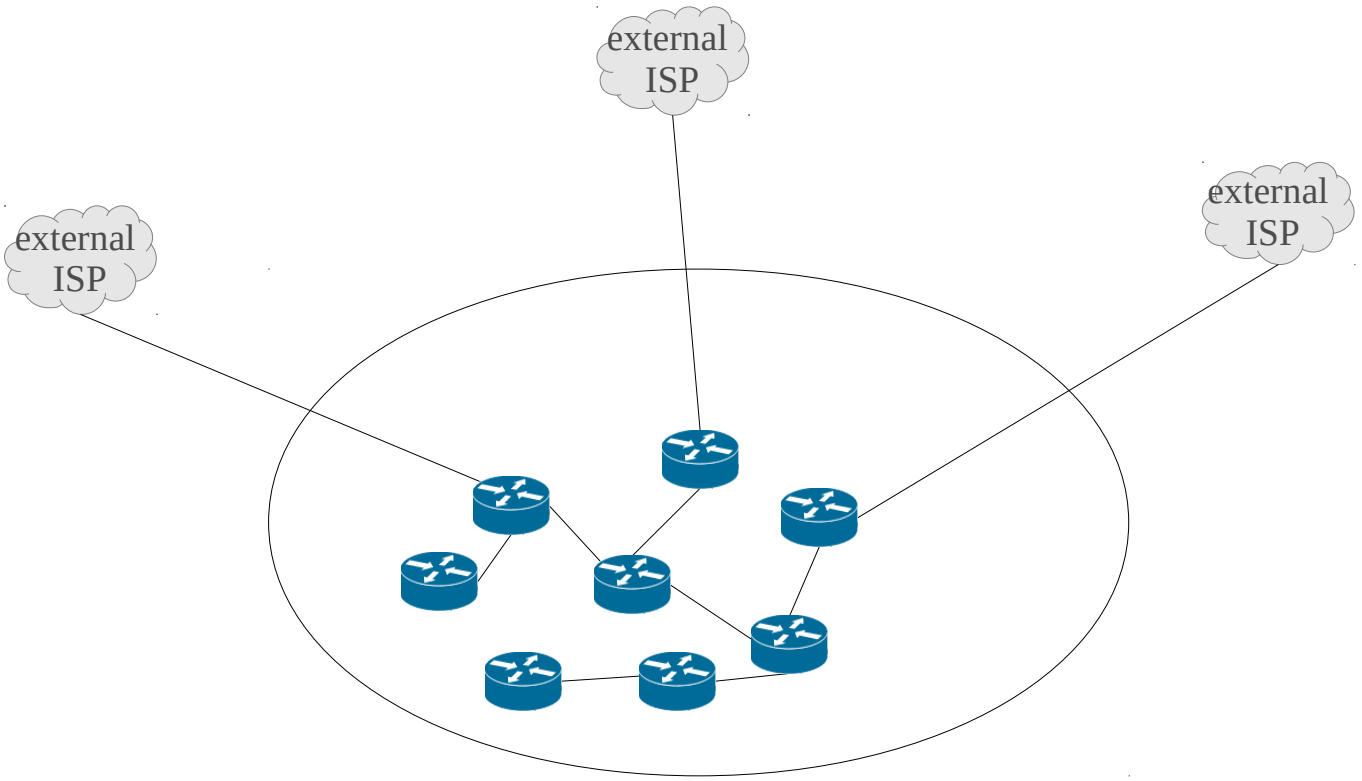
*“Money plays the largest part in determining the course of history”*

Karl Marx

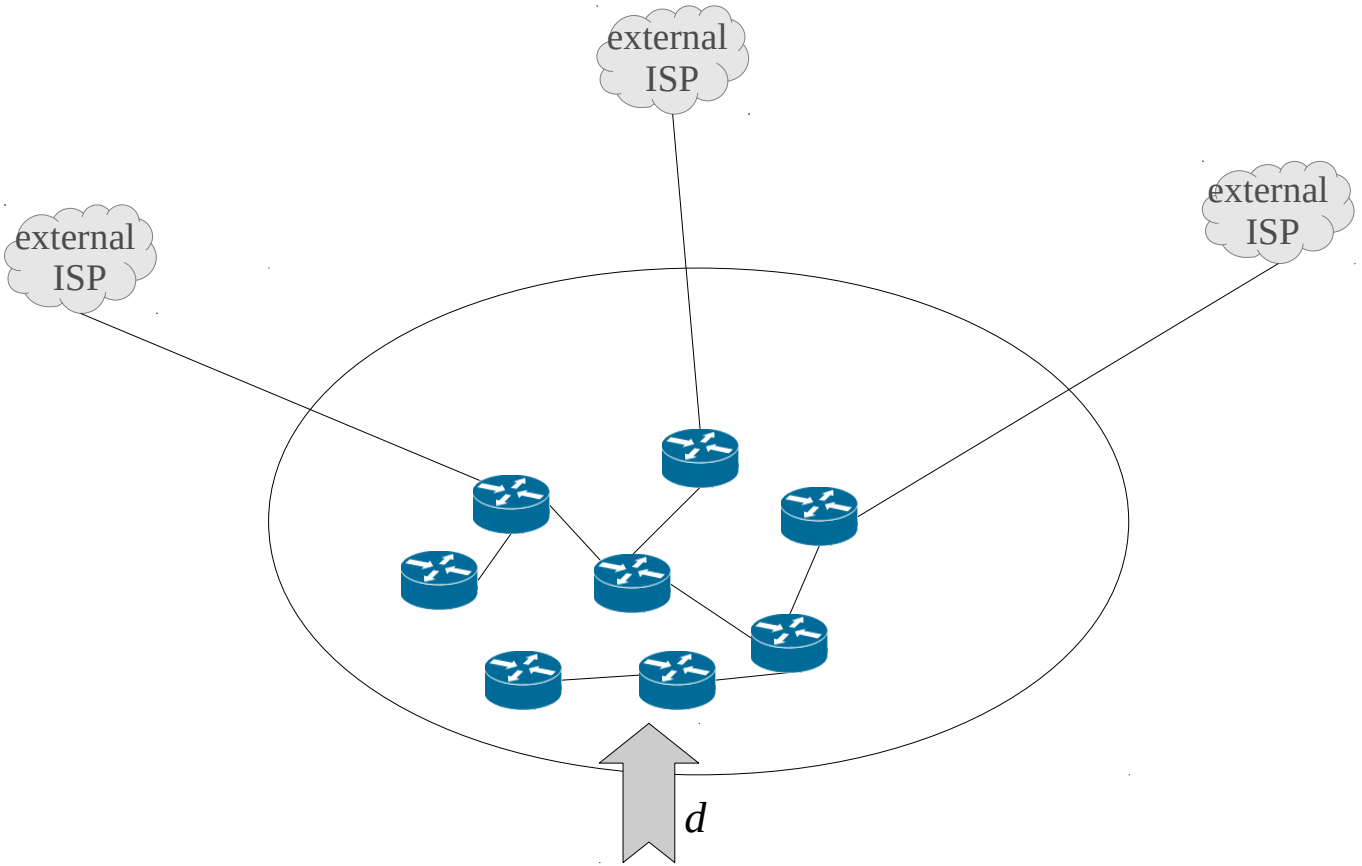
Can caching in ICN reduce ISP operational costs?

A. Araldo, M. Mangili, F. Martignon, D. Rossi, "Cost-aware caching: optimizing cache provisioning and object placement in ICN", in **IEEE Globecom**, Austin, **2014**

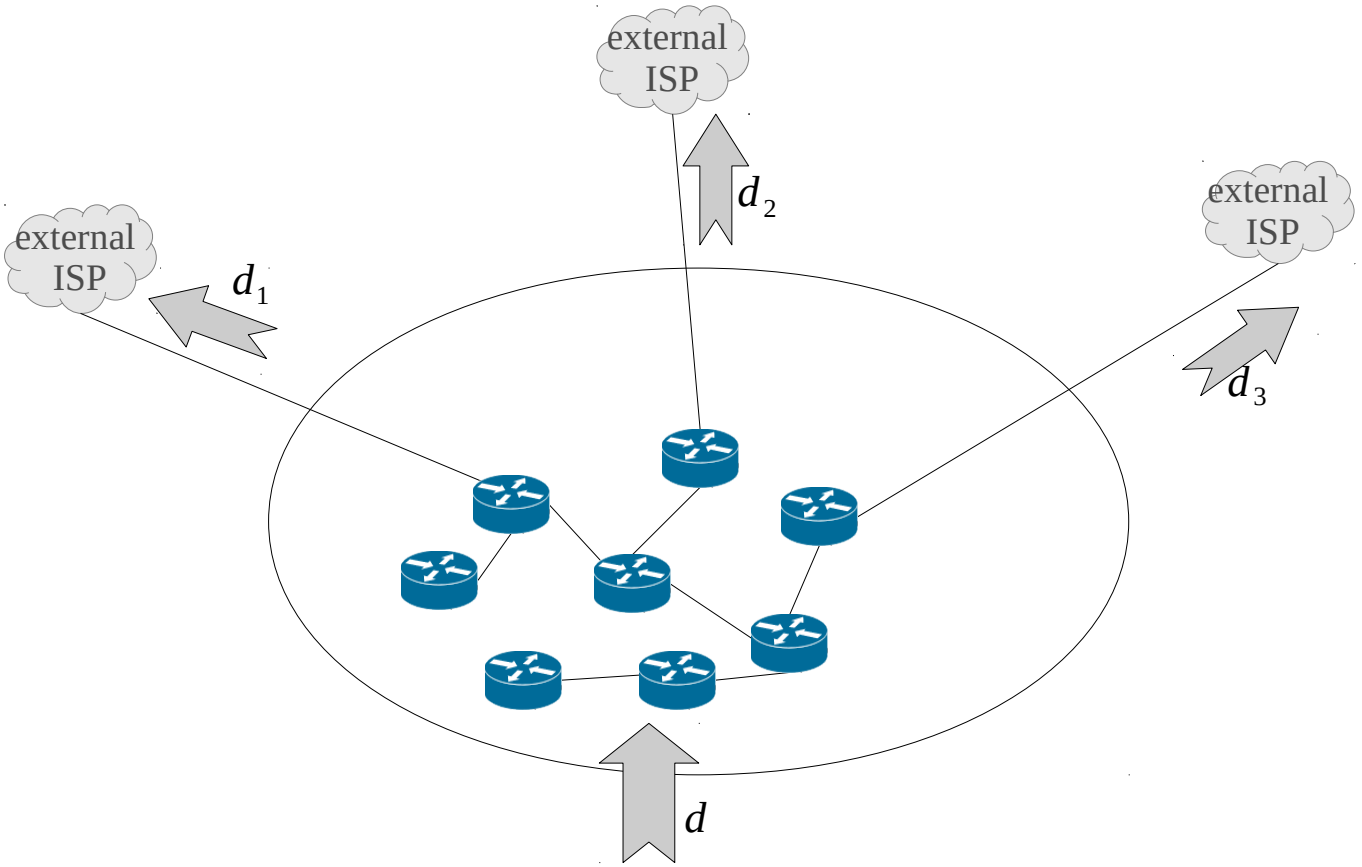
# ISP network



# ISP network

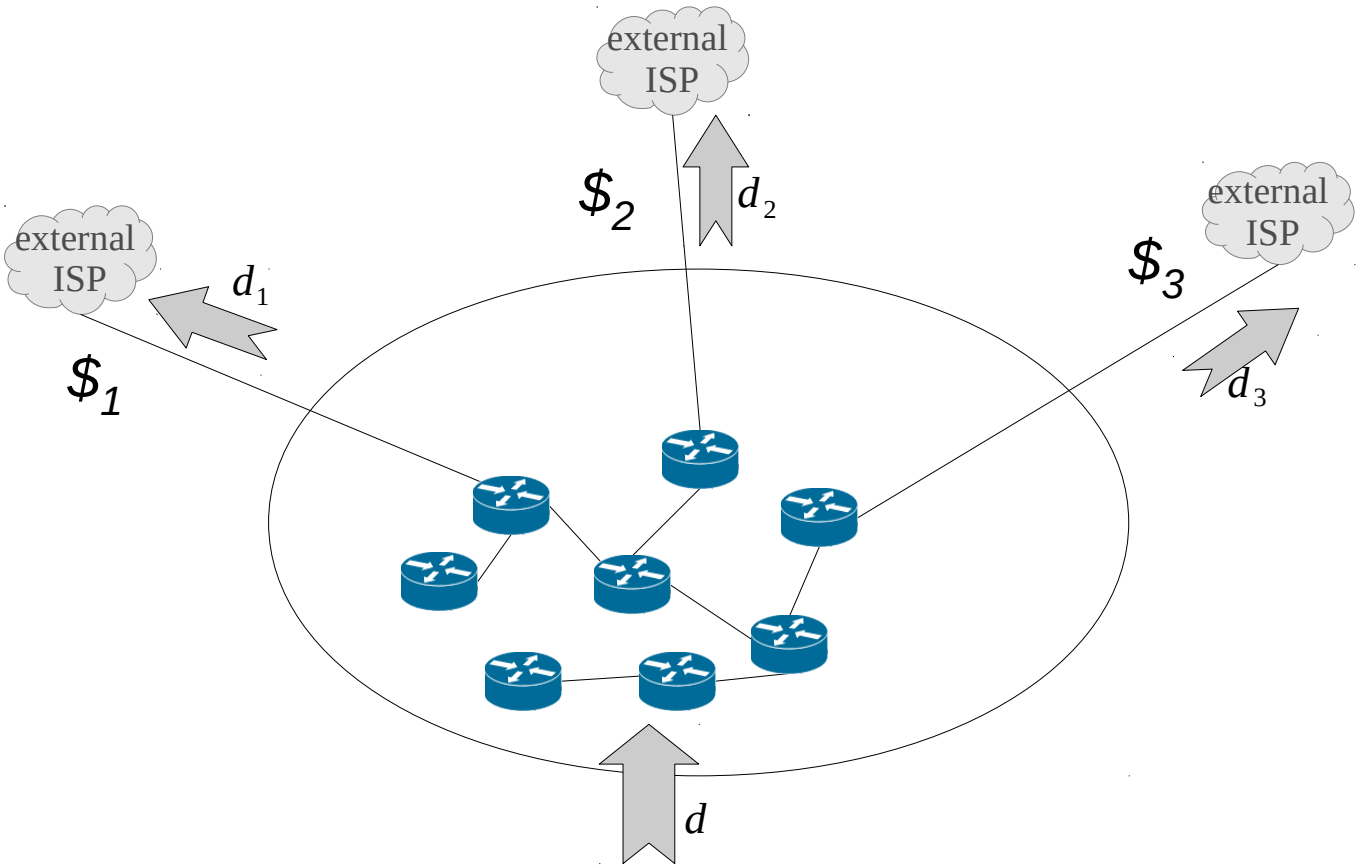


# ISP network



# ISP network

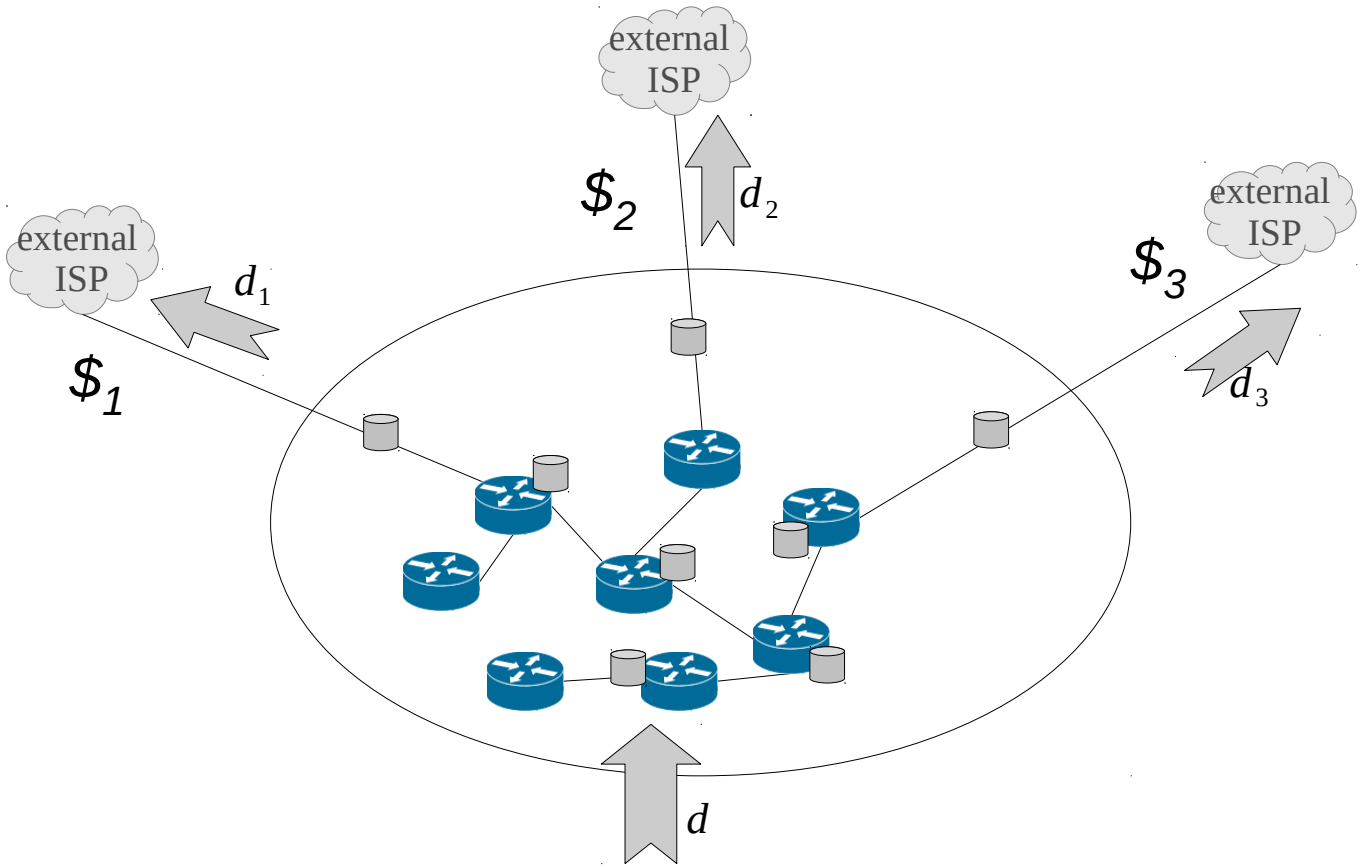
Cost: $C = \sum_i d_i \$_i$




$\$_1 < \$_2 < \$_3$

# ISP network

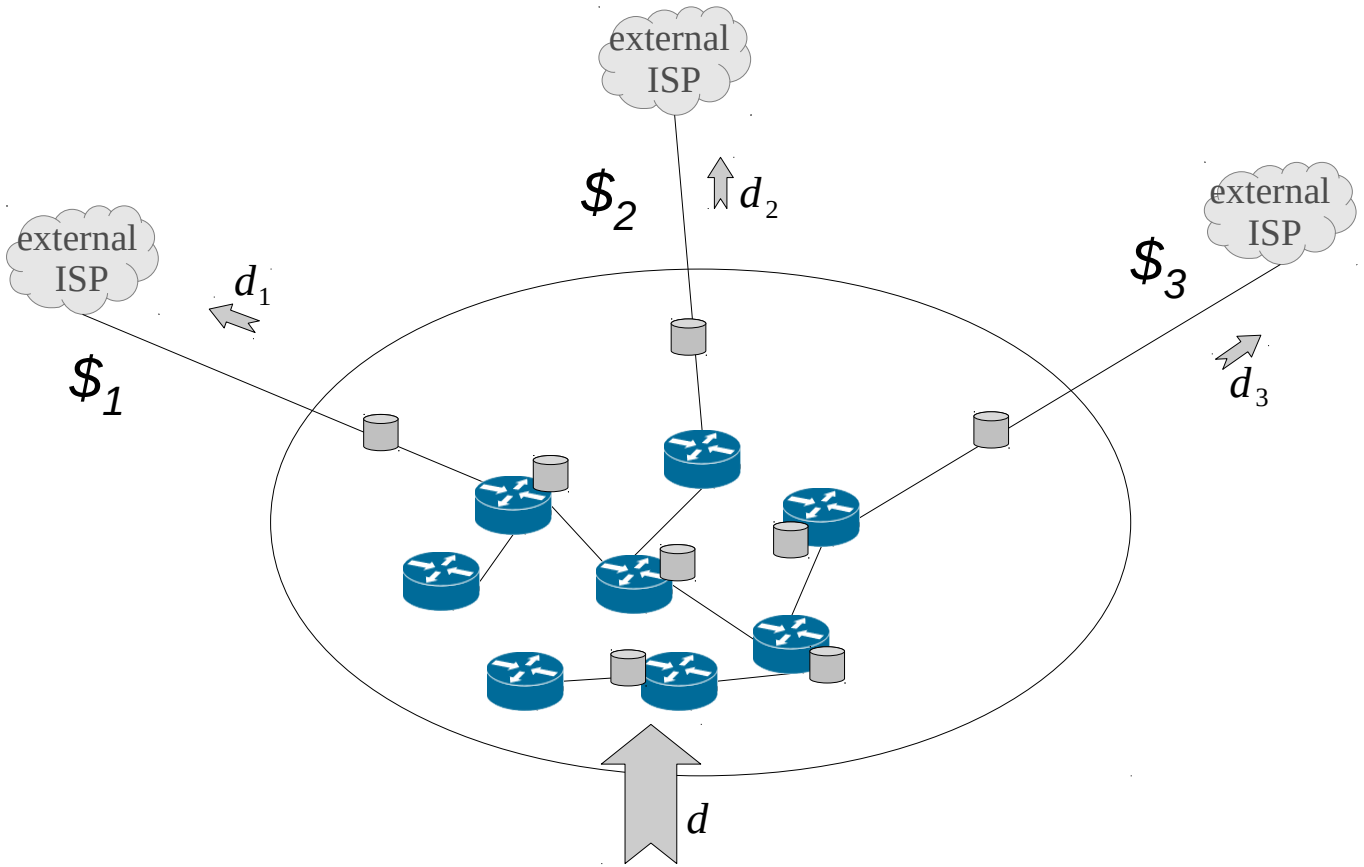
Cost: $C = \sum_i d_i \$_i$




$\$_1 < \$_2 < \$_3$

# ISP network

Cost: $C = \sum_i d_i \$_i$



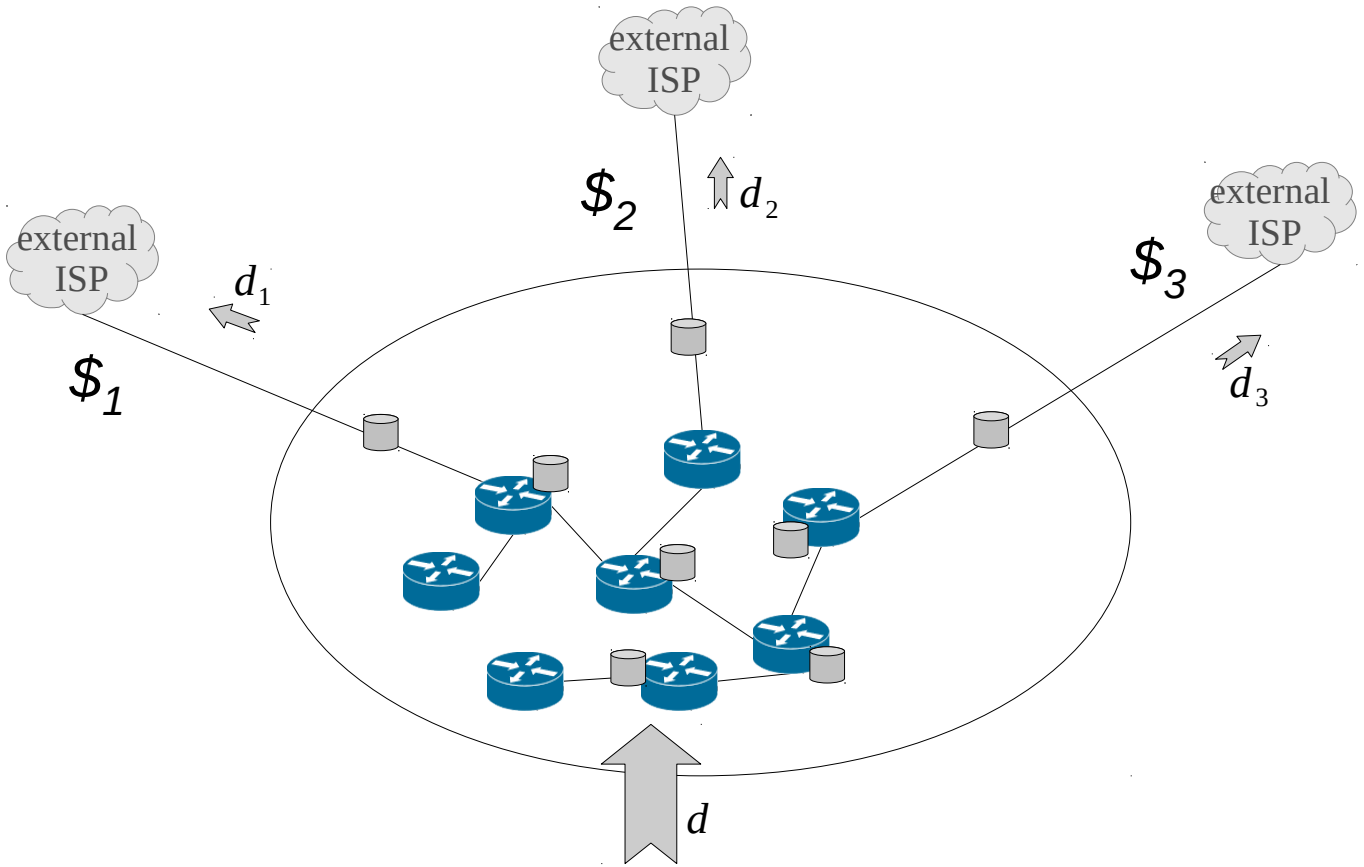

$\$_1 < \$_2 < \$_3$

# ISP network

Hit ratio: $H = \frac{d - \sum_i d_i}{d}$

Cost: $C = \sum_i d_i \$_i$

Classic caching: max $H$




$\$_1 < \$_2 < \$_3$

# ISP network

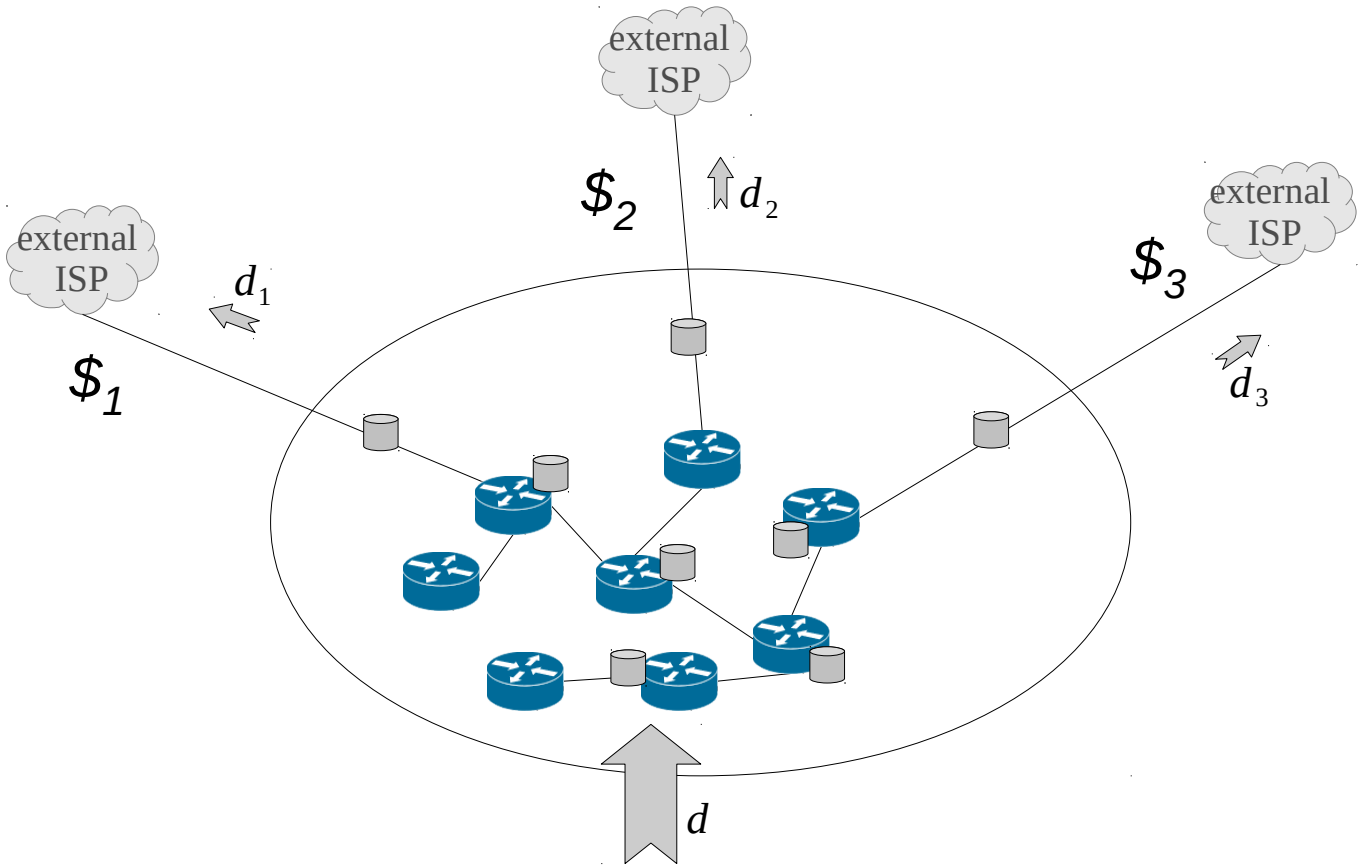
Hit ratio: $H = \frac{d - \sum_i d_i}{d}$

Cost: $C = \sum_i d_i \$_i$

Classic caching: max $H$

max $H$ ⇨ min $C$

(cost)




$\$_1 < \$_2 < \$_3$

# ISP network

Hit ratio: $H = \frac{d - \sum_i d_i}{d}$

Cost: $C = \sum_i d_i \$_i$

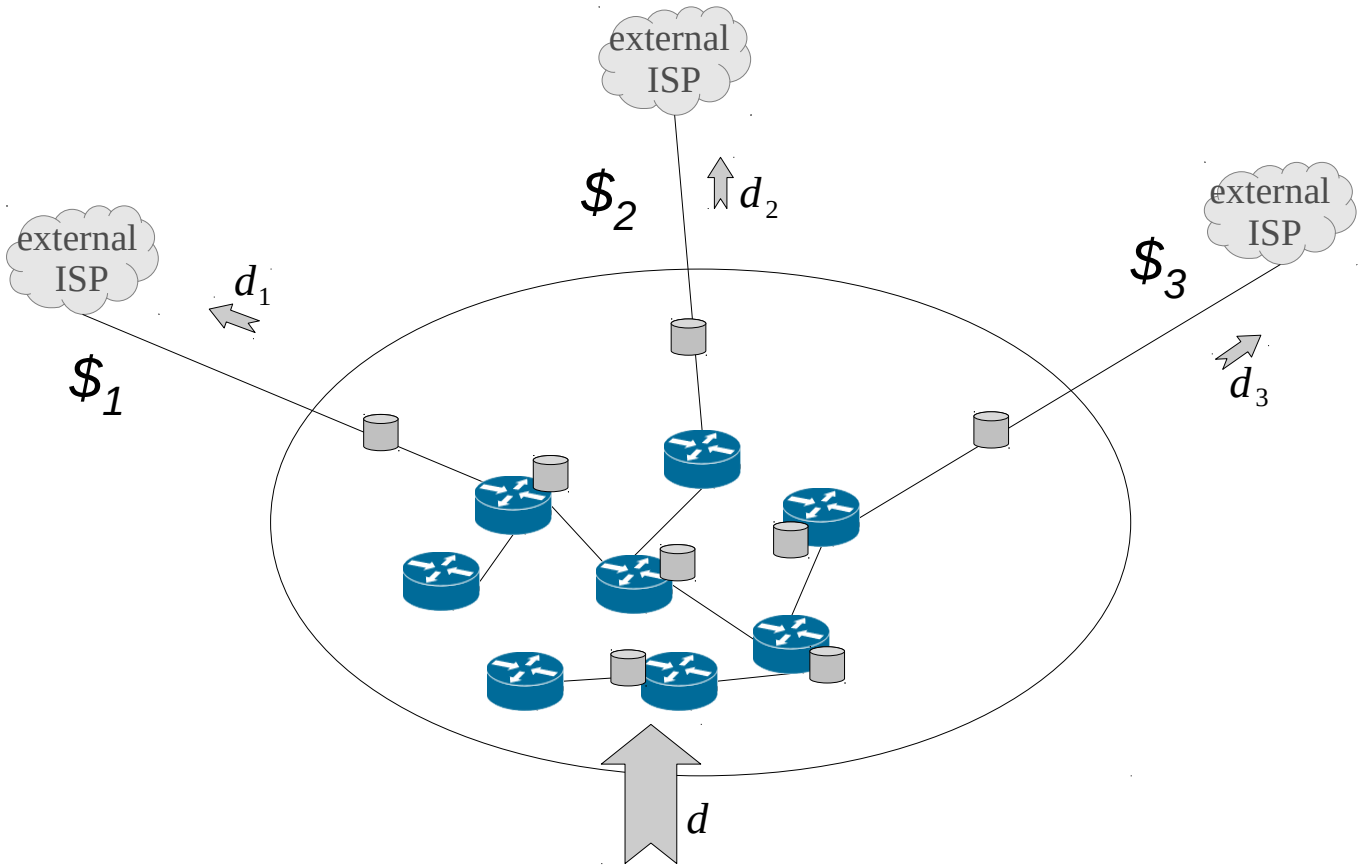
Classic caching: max $H$

~~max $H$ ⇨ min $C$~~

(cost)

H, C are conflicting goals

Price heterogeneity




$\$_1 < \$_2 < \$_3$

# Optimization model

Hit ratio: $H = \frac{d - \sum_i d_i}{d}$

Cost: $C = \sum_i d_i \$_i$

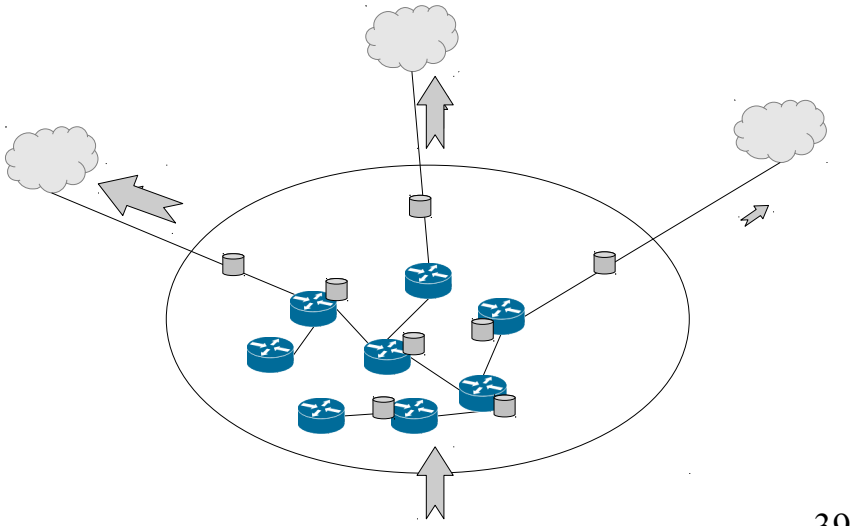
Classic caching: MAX-HIT

vs.

Cost-aware caching: MIN-COST

We find the optimal

- object placement
- cache sizing
- forwarding



# Greedy algorithms

MAX-HIT (classic)

- Cache the |C| objects with the highest number of requests $d^o$

cache size

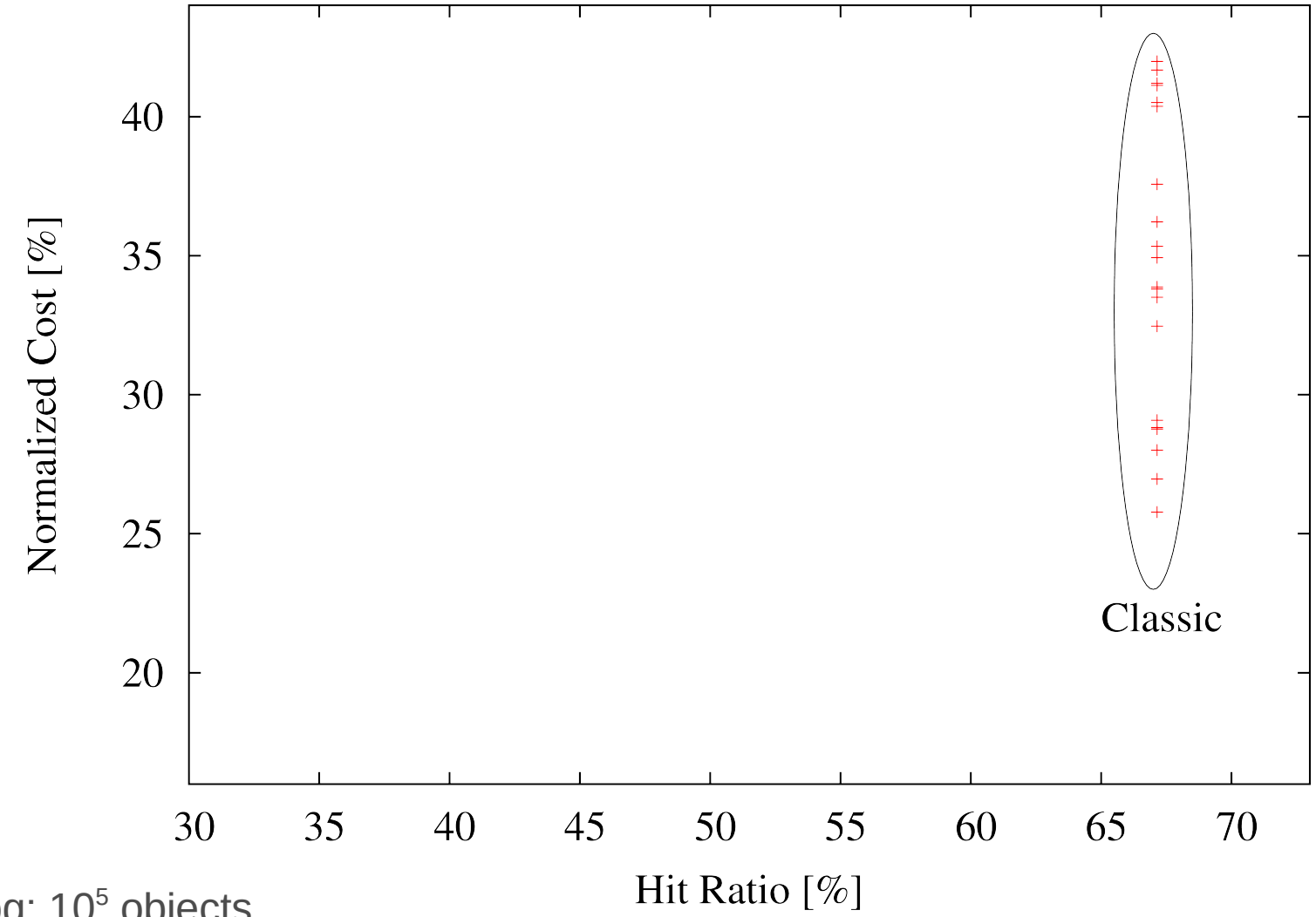
Polynomial time

MIN-COST (cost-aware)

- Cache the |C| objects with the highest cost $d^o \cdot \$^o$

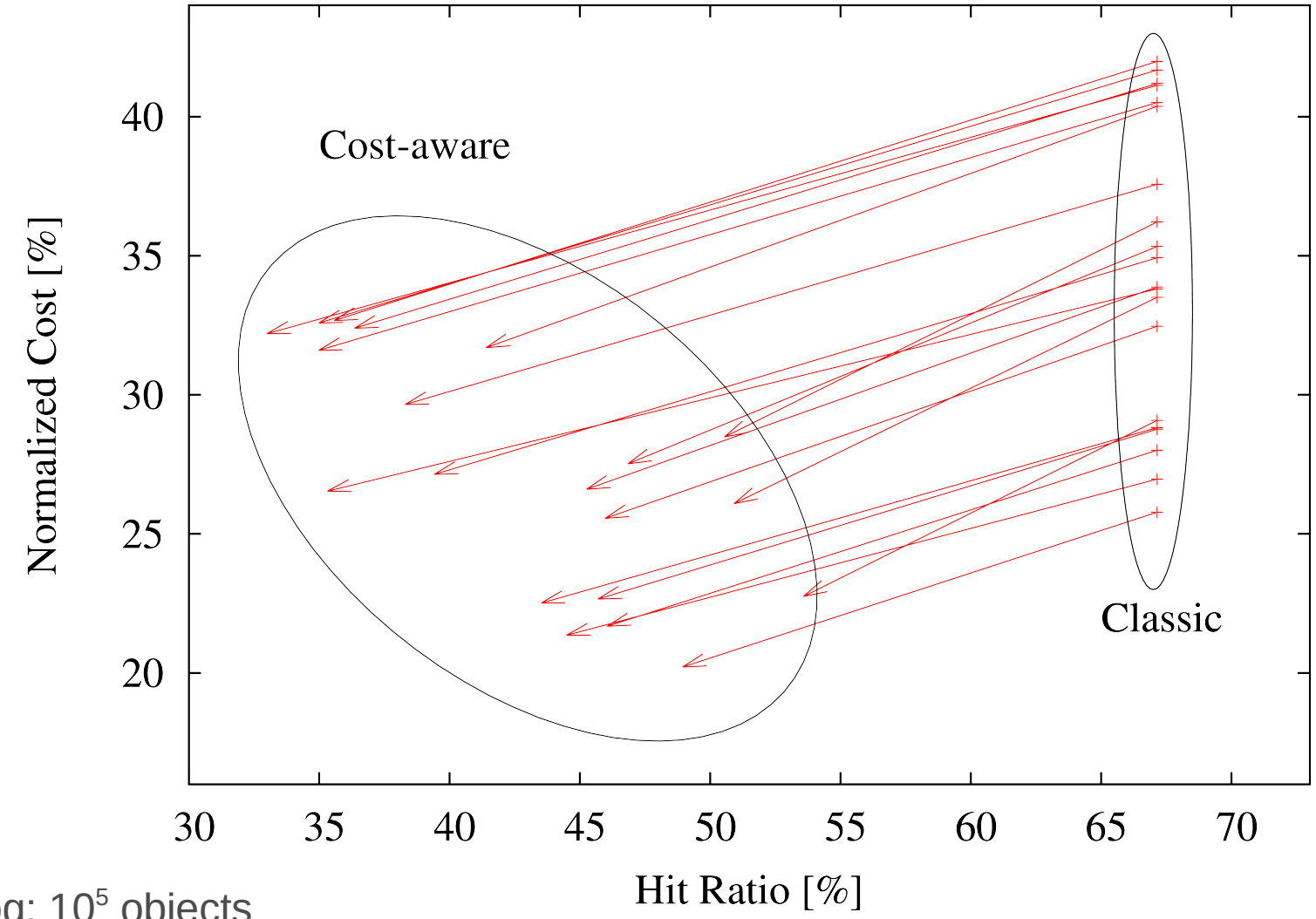
Price of the link giving access to object $o$

# Hit-ratio vs. cost tradeoff




Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

# Hit-ratio vs. cost tradeoff



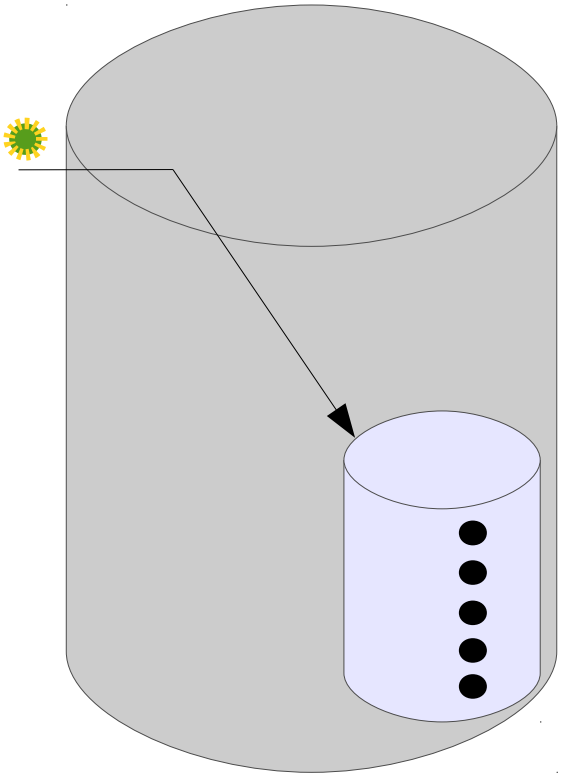

Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

Parc, Alcatel, Cisco are working on Information Centric Routers operating at 10Gbps.

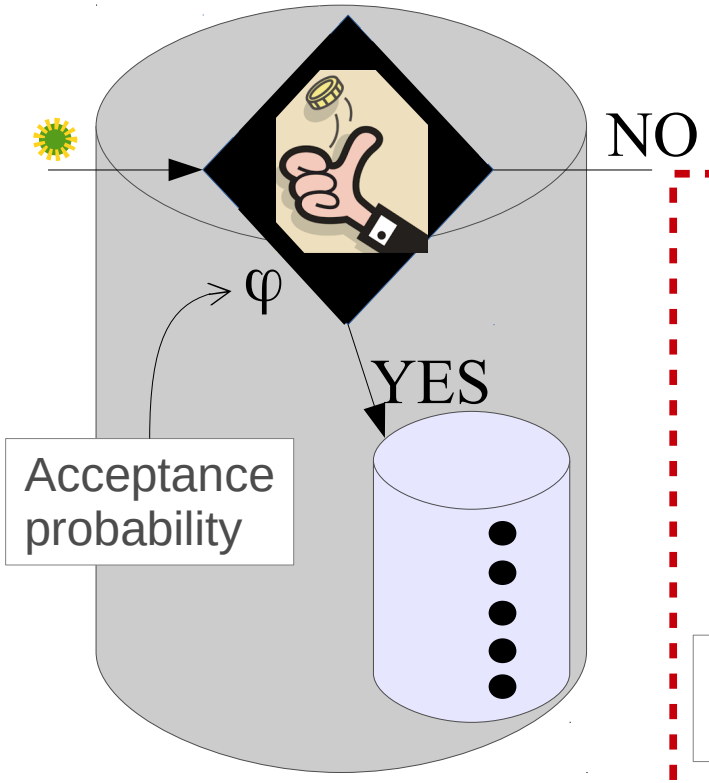
Can we implement a cost-aware cache strategy?

A. Araldo, D. Rossi, F. Martignon, “Design and Evaluation of Cost-aware Information Centric Routers”, in **ACM SIGCOMM Conference on Information-Centric Networking** (ICN), Paris, **2014**

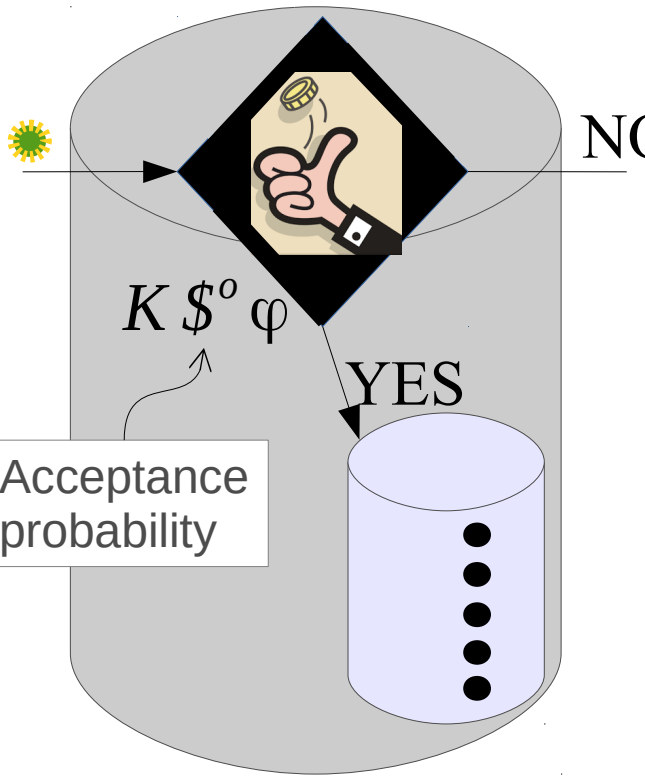
# Decision policies



Always
cache

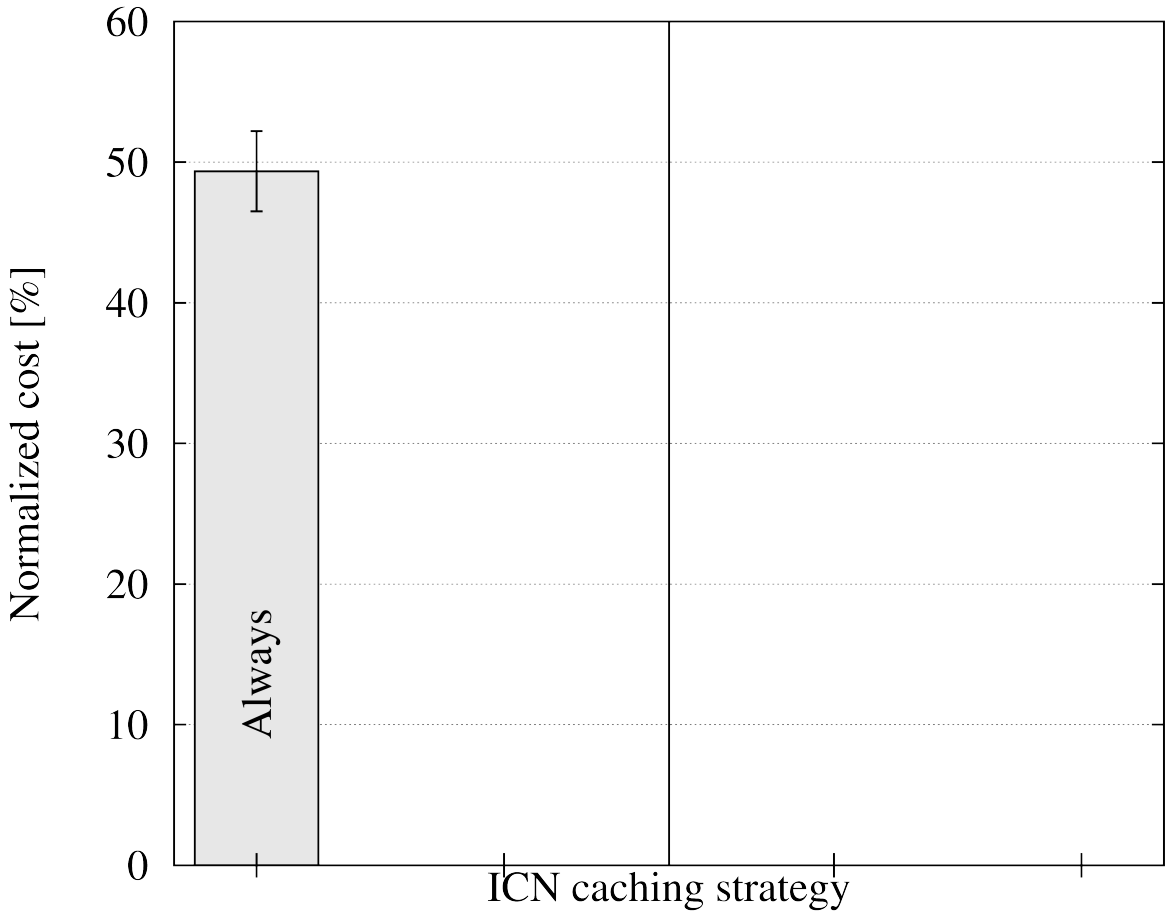



Uniform
probabilistic



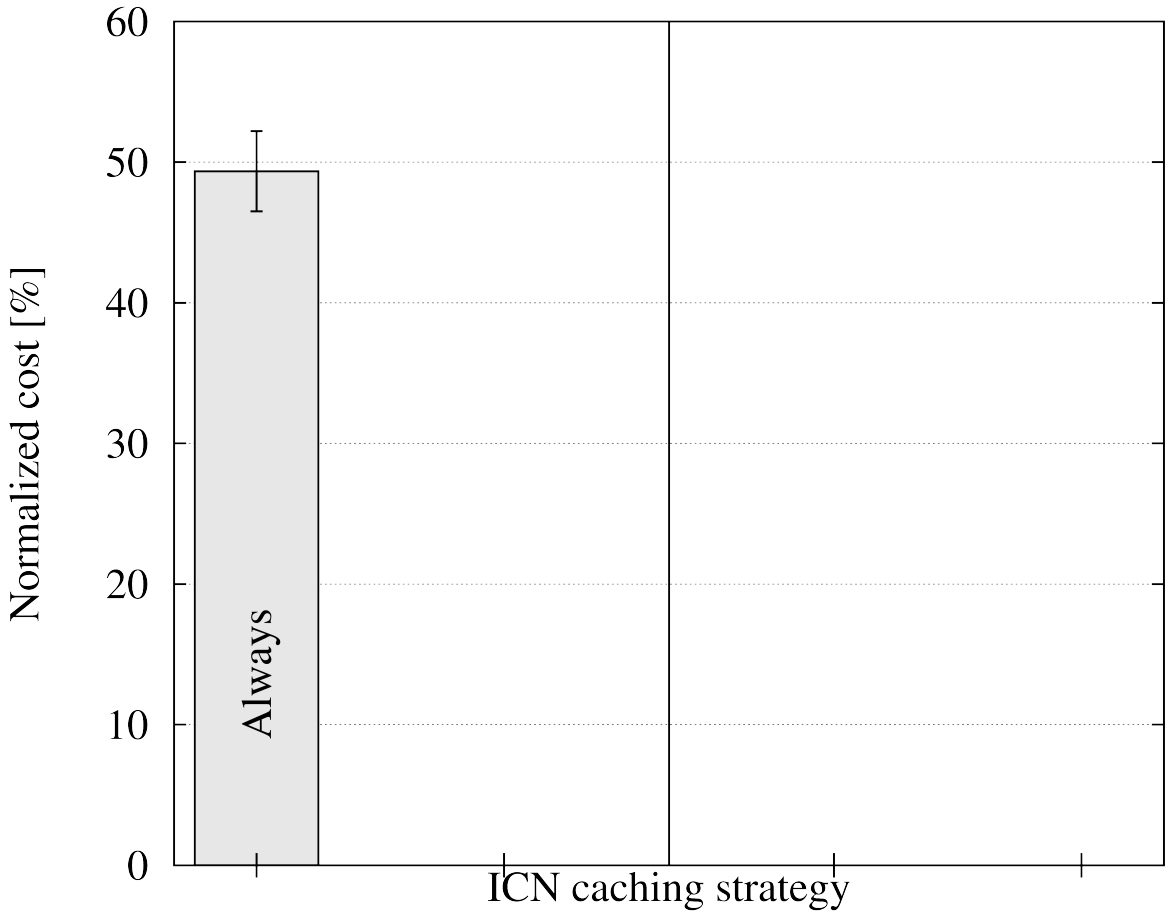

Cost-Aware
(CoA)

# Savings



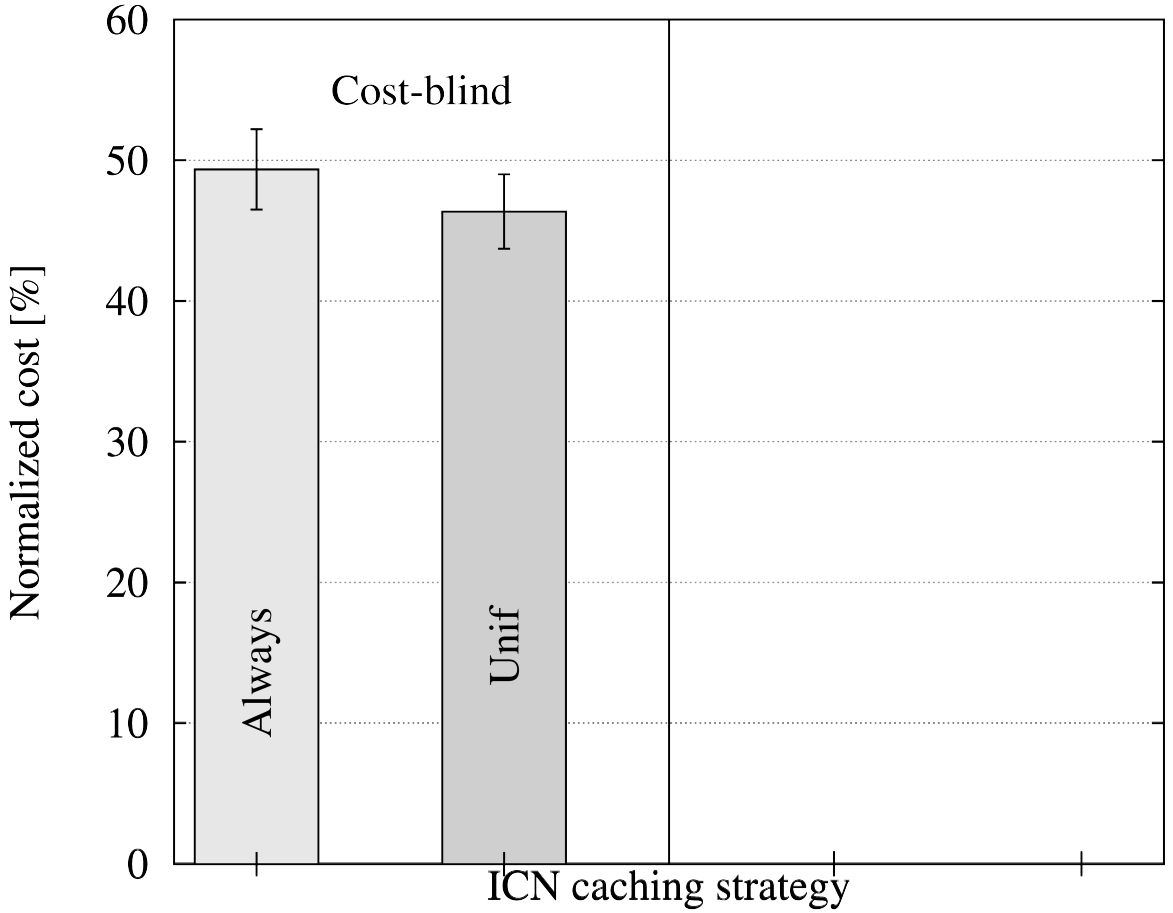

Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

# Savings



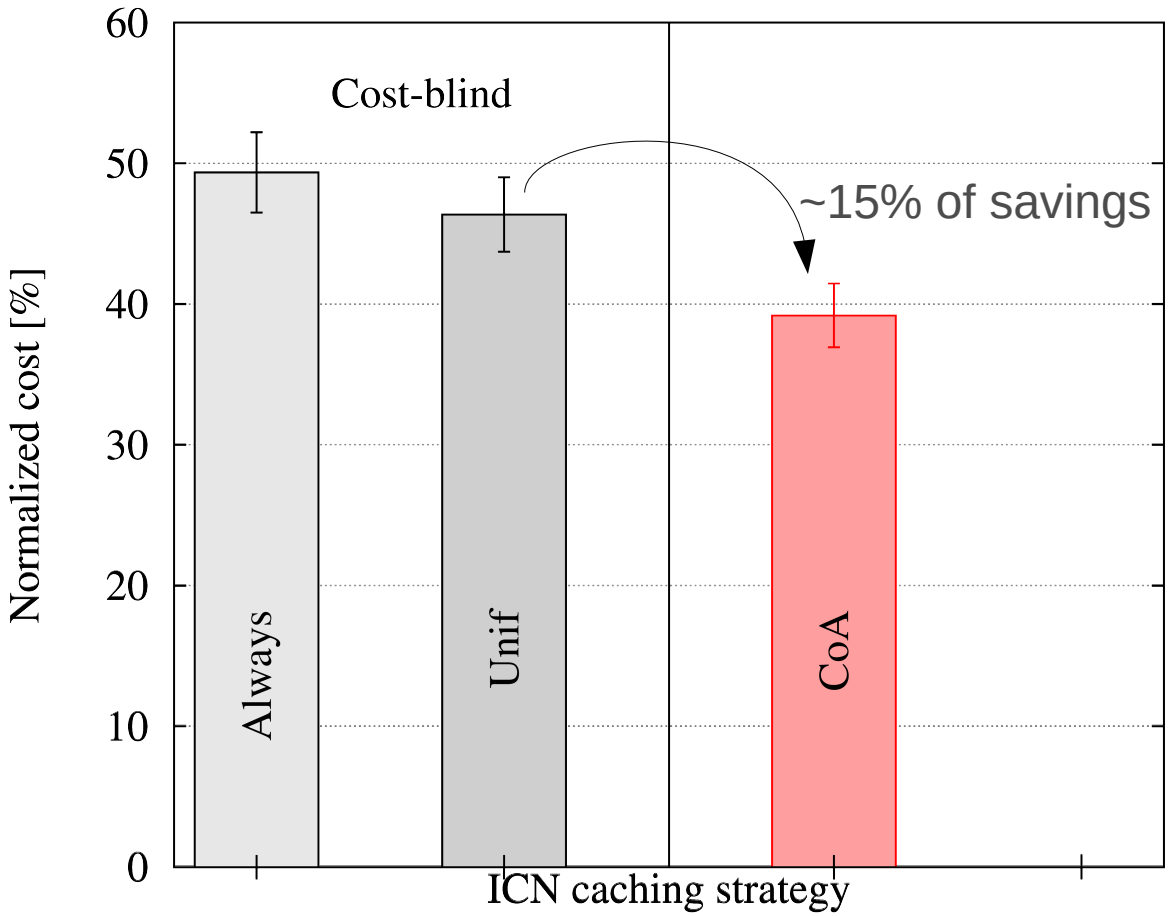

Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

# Savings




Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

# Savings



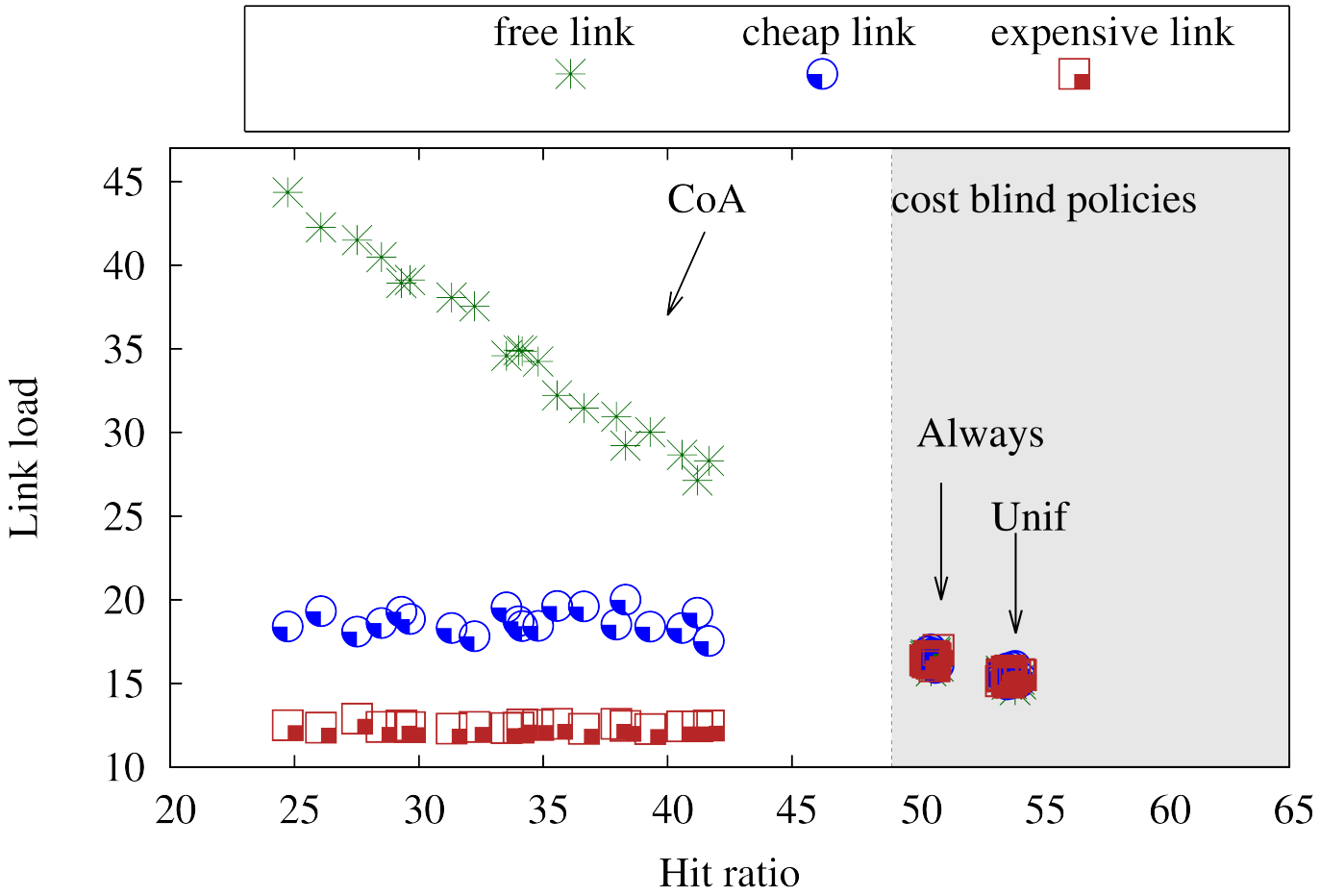

Catalog: $10^5$ objects
Cache: $10^3$ objects
Zipf exponent: 1
Prices: ($\$_1$,$\$_2$,$\$_3$) = (0,1,10)

- Operational cost reduction accumulates over years
- It comes for free (cache deployment cost does not change)

# Conclusion

- Our contribution:
  - Optimization model of ISPs operational costs
  - Polynomial time greedy algorithms
  - Design of a Cost-Aware cache strategy
  - Evaluation through simulations
- Take-away message:
  - Do not look at hit ratio when evaluating a cache system
    - Look at your real objective
  - Classic caching is cost-ineffective
  - Cost-aware caching provides consistent savings in theory and in practice

# Link load




$$load(l) = \frac{\text{downloads from } l}{\text{requests}}$$

- Cost unaware policies distribute load blindly across all links
- On the contrary, CoA effectively **minimizes the load on paid link**
- The paid link load is constant, even when hit ratio decreases

CoA effectively **distributes the additional miss-stream to the free link**