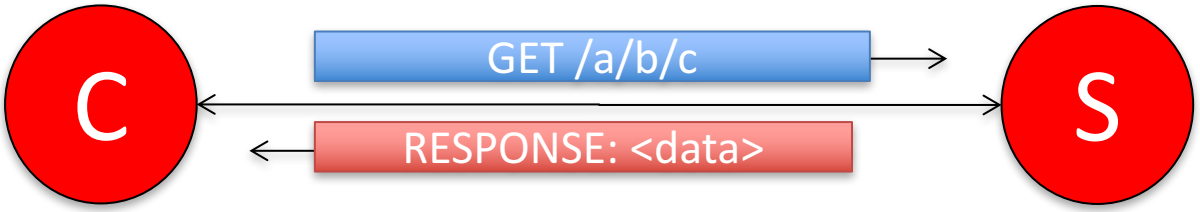
UCI University of California, Irvine

When Encryption is Not Enough

Privacy Attacks in Content-Centric Networking

Cesar Ghali, Gene Tsudik and Christopher A. Wood
University of California, Irvine
{cghali, gene.tsudik, woodc1}@uci.edu

Privacy with IP

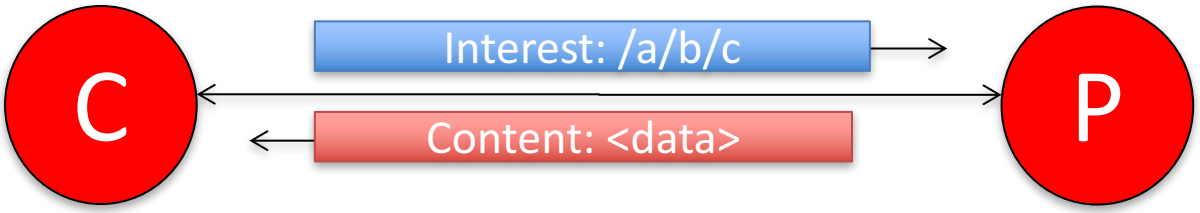


Privacy with IP

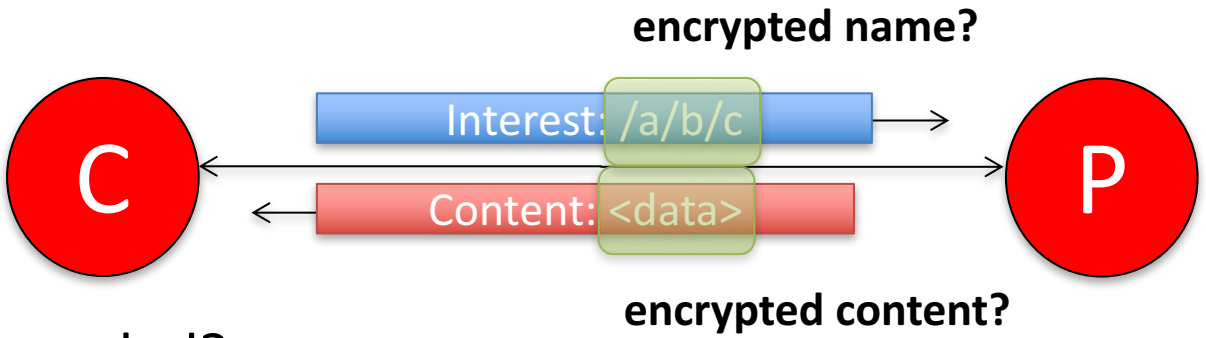
What's revealed?

- Source and destination addresses and port #
- Timing
- Packet sizes

Privacy with CCN



Privacy with CCN



What's revealed?

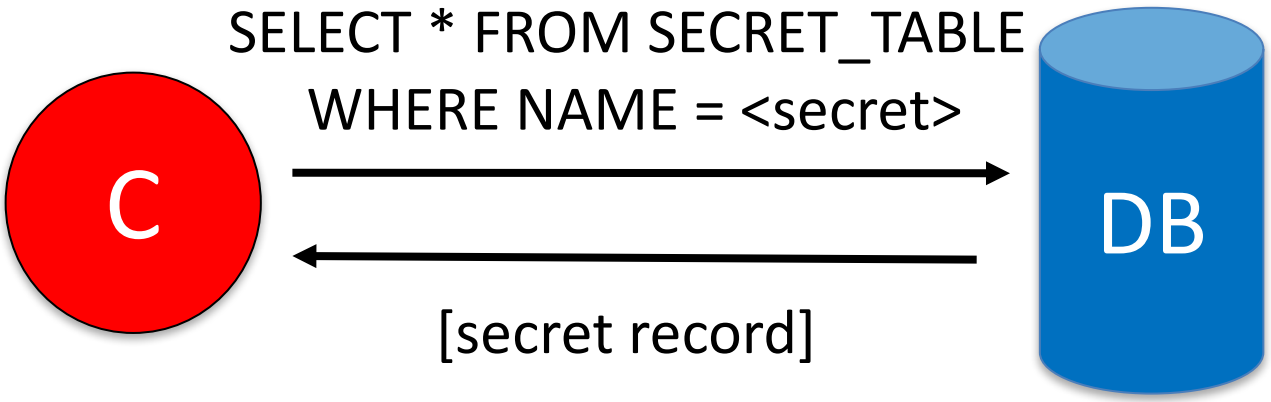
- Consumer and producer locations
- Timing
- Packet sizes
- Producer identity
- **Interest name (and equality)**
- …

Motivating Question

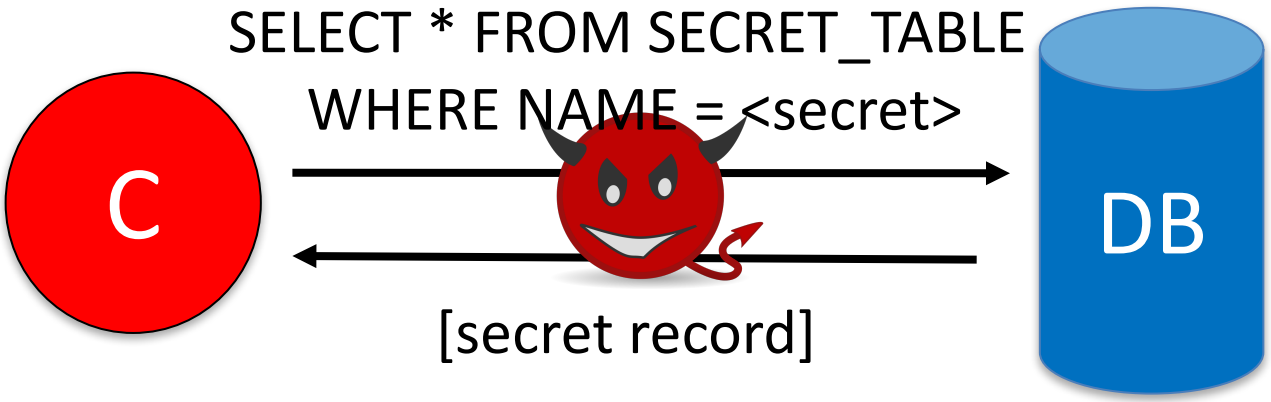
What can an adversary do with interest equality alone?

Over to encrypted databases…

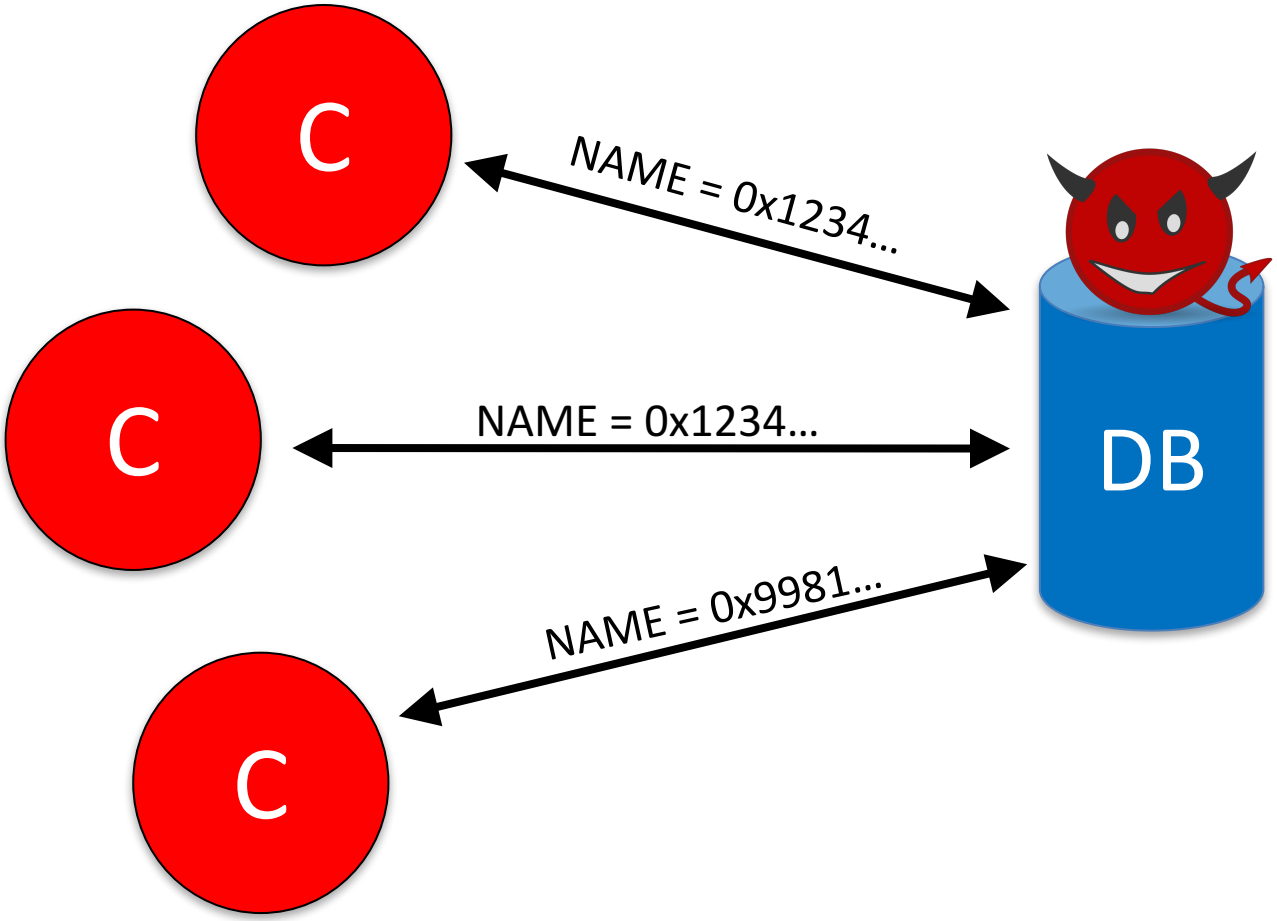
Encrypted Database



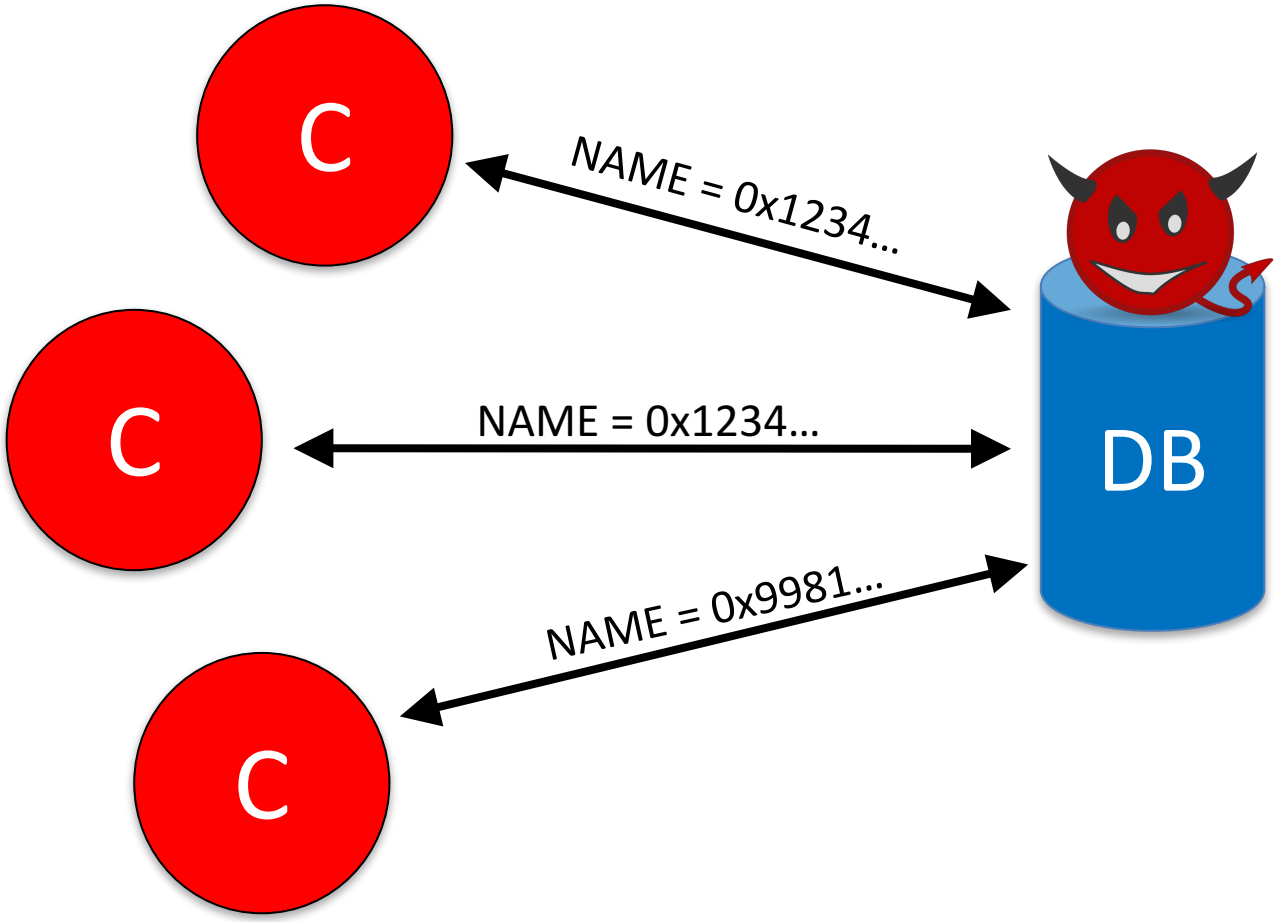
Encrypted Database



Eavesdropping

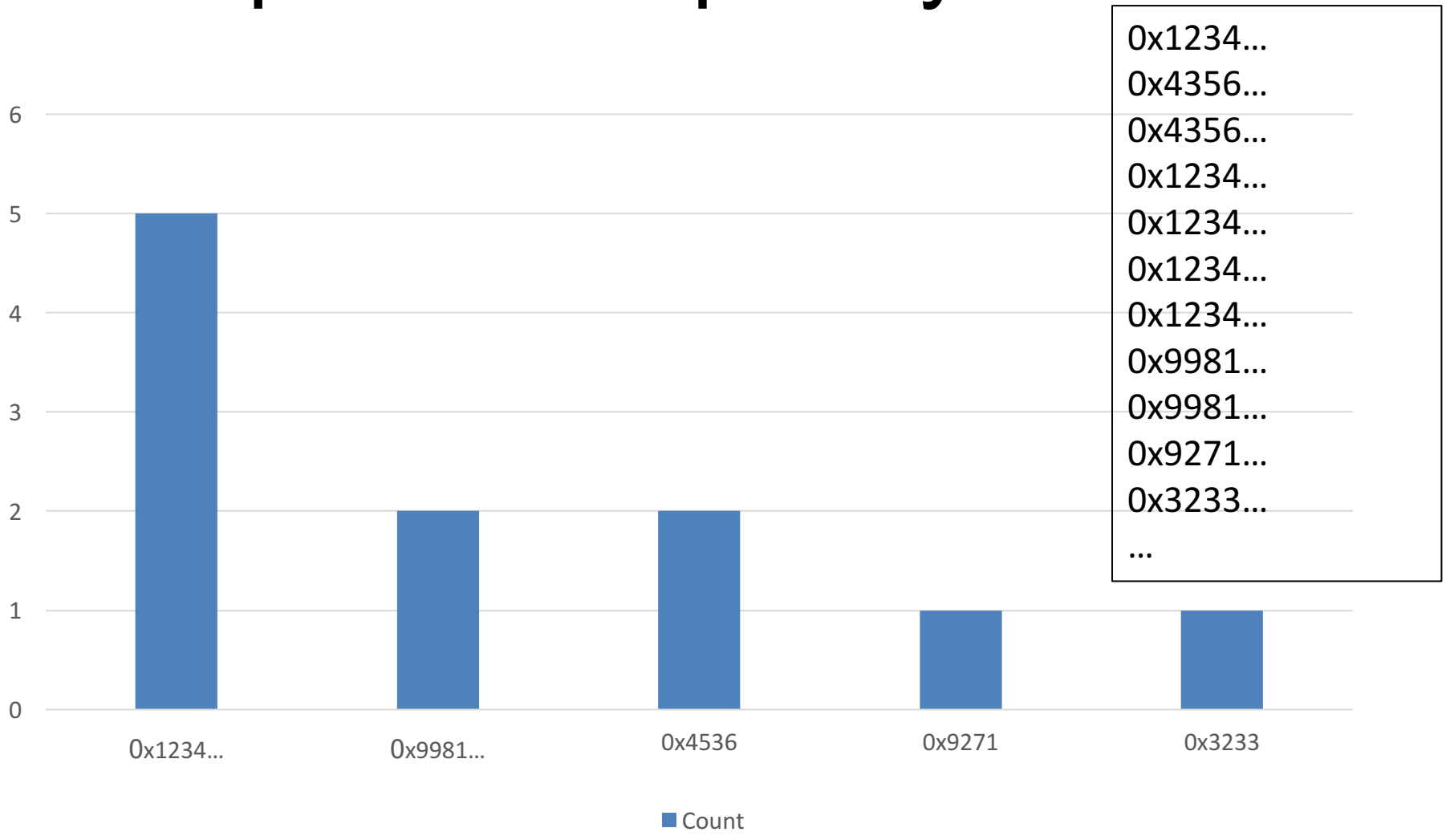


Eavesdropping

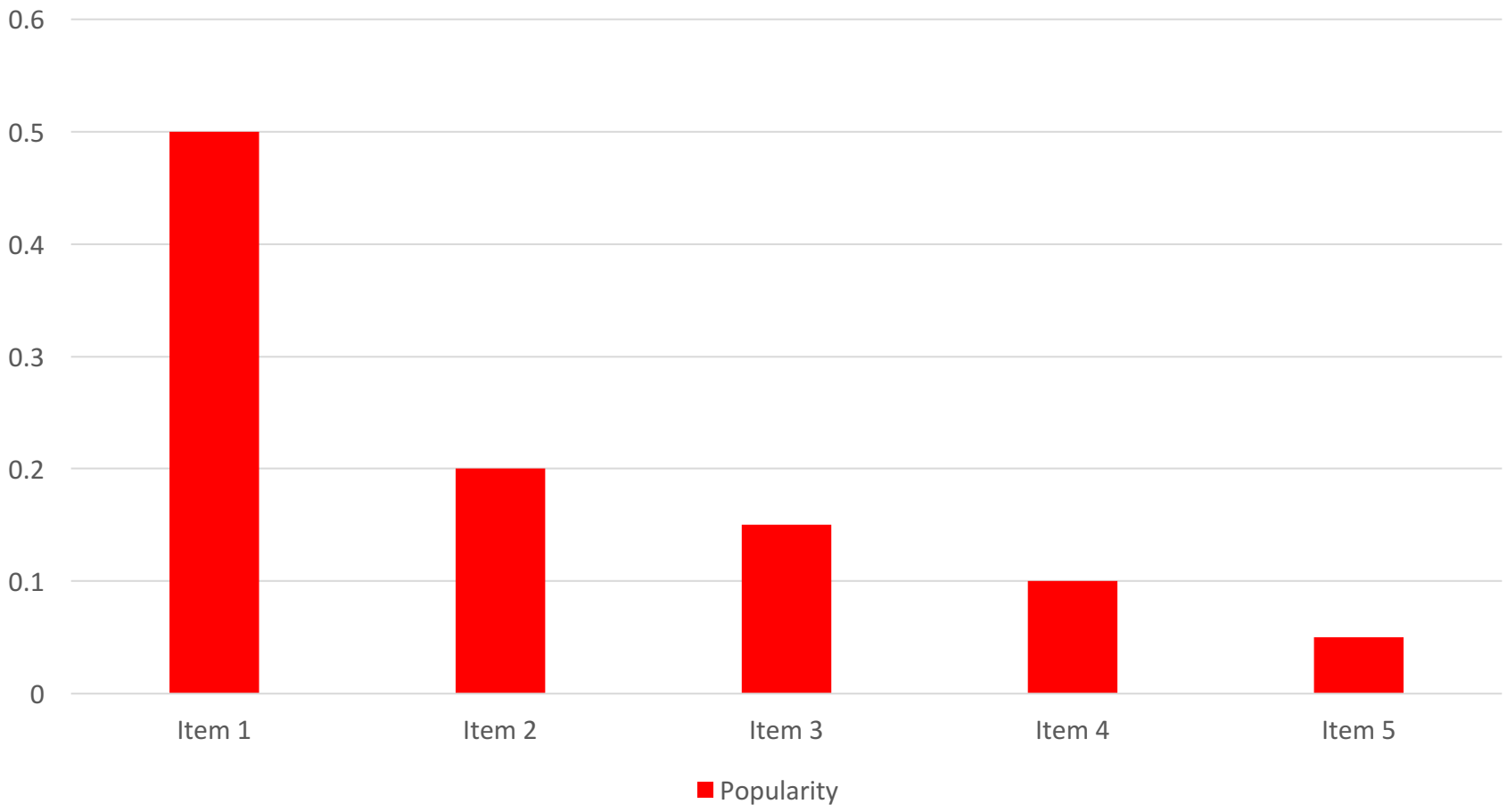


0x1234…
0x4356…
0x4356…
0x1234…
0x1234…
0x1234…
0x1234…
0x9981…
0x9981…
0x9271…
0x3233…
…

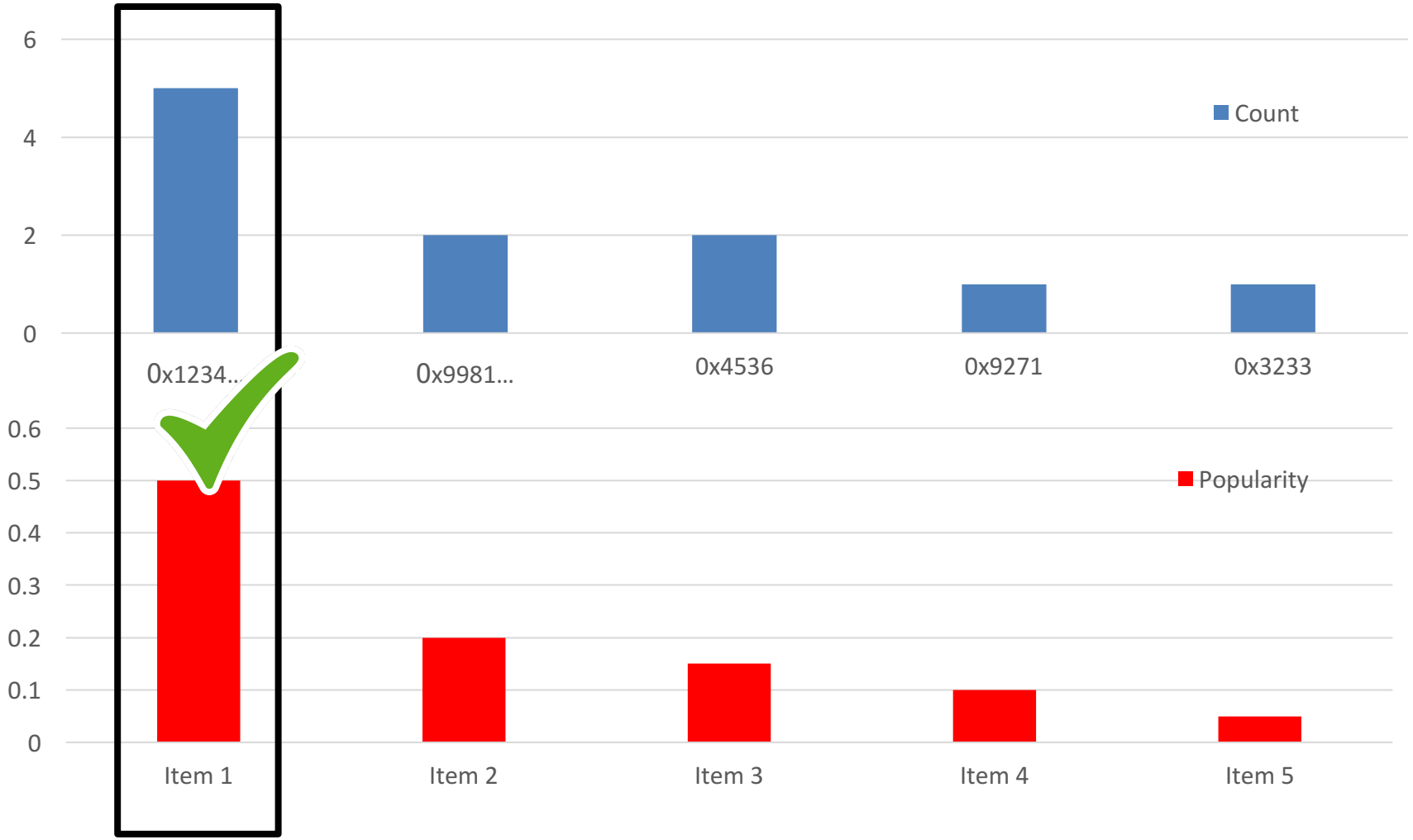
Empirical Frequency Counts



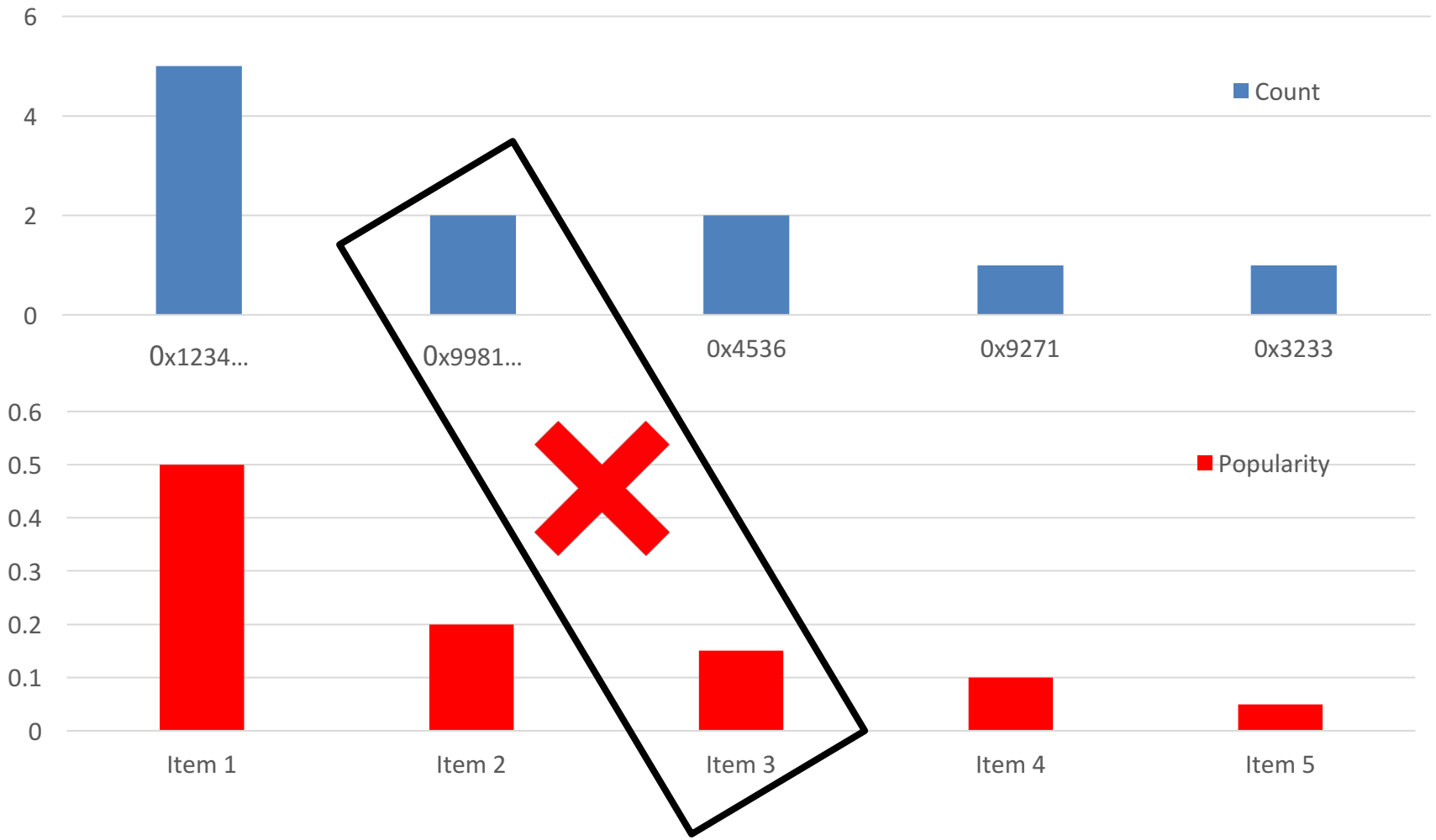
Auxiliary Popularity Info



Frequency Analysis Attack

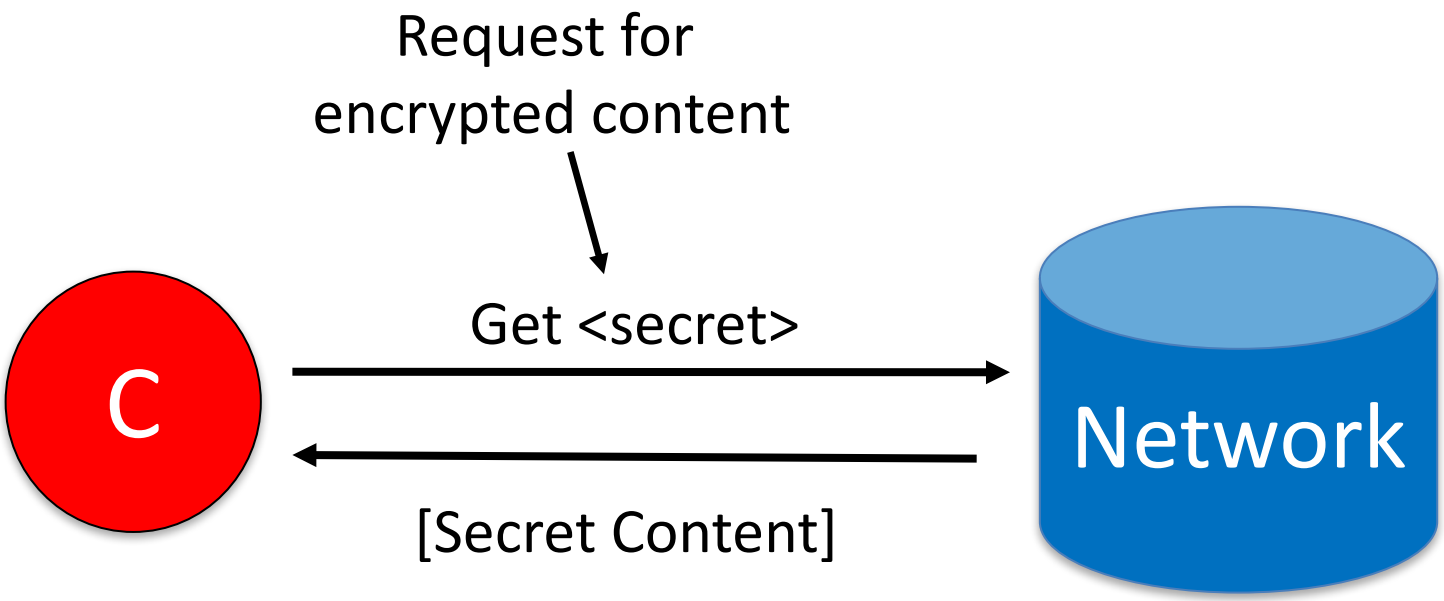


Frequency Analysis Attack



Back to CCN…

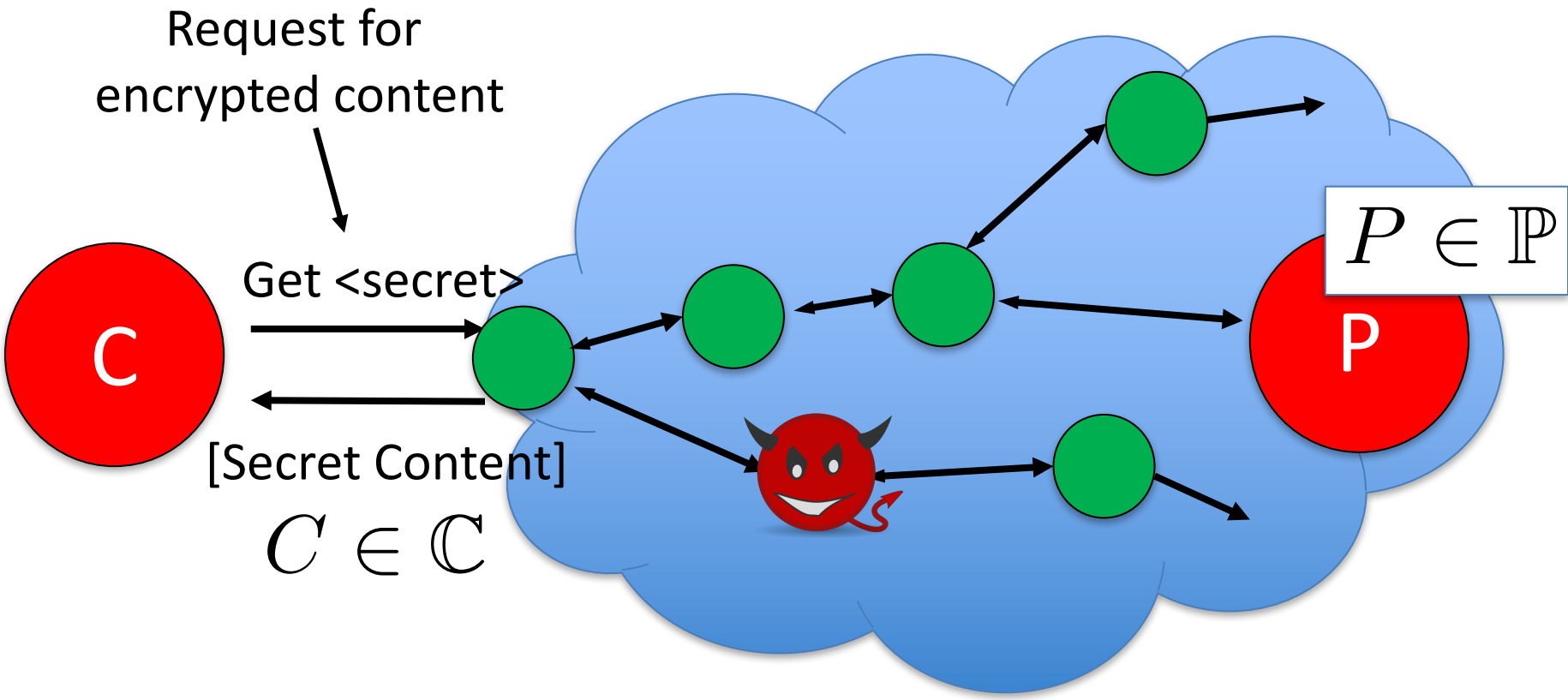
CCN as a Content Database



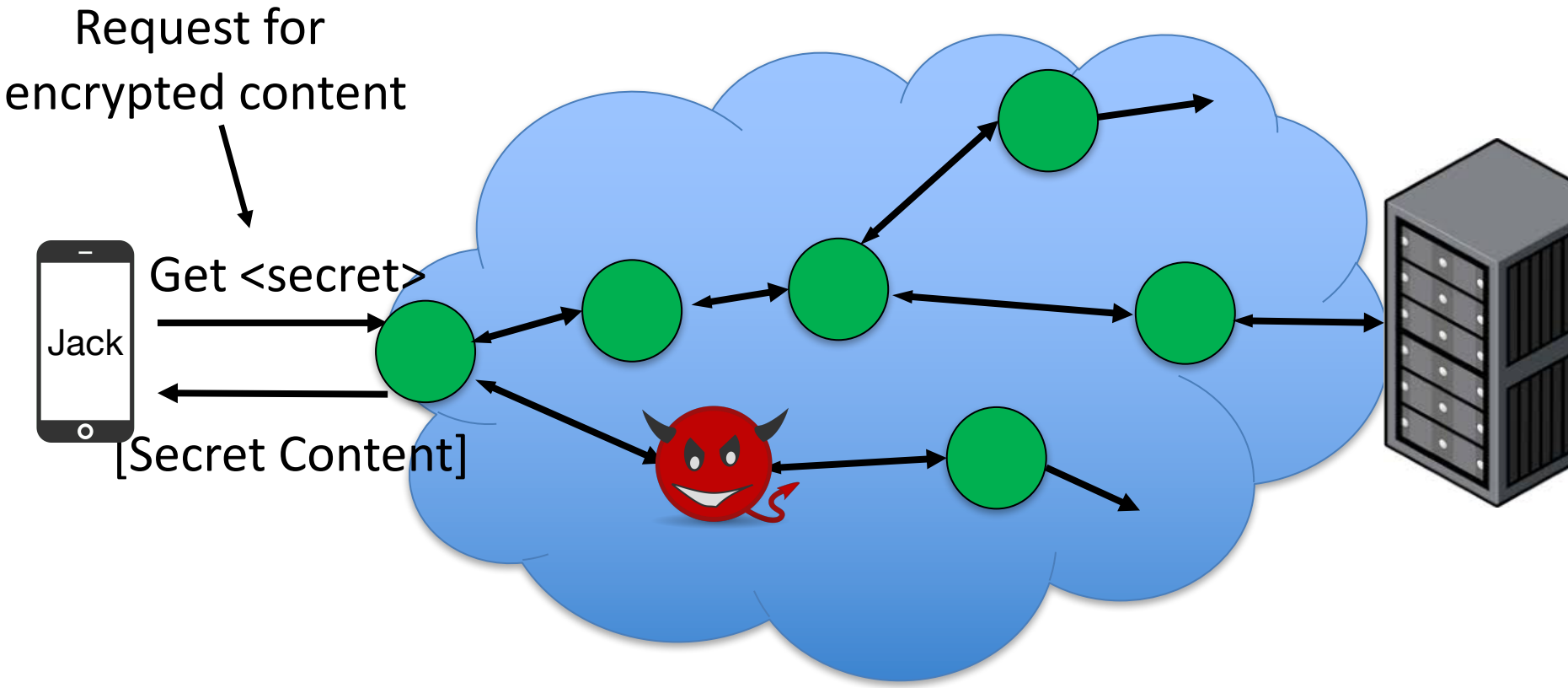
$\mathbb{P}$ Application data items

$\mathbb{C}$ Encrypted data items

CCN as a Content Database



CCN as a Content Database



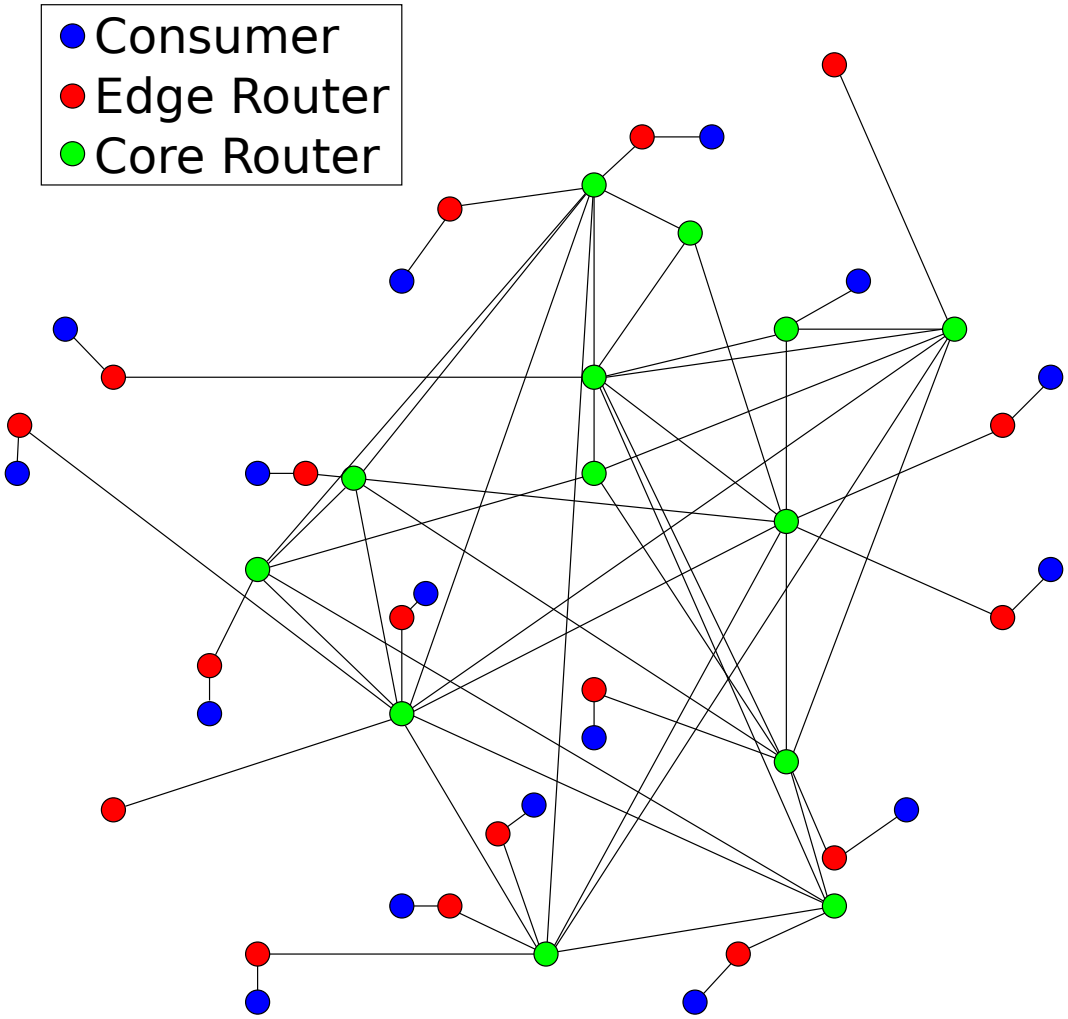
Relevant Distributions

- Real popularity distribution $\mathcal{D}_R(\mathbb{P})$
- Auxiliary information distribution $\mathcal{D}_A^{\mathcal{A}}(\mathbb{P})$
- Empirical frequency distribution $\mathcal{D}_E(\mathbb{C})$

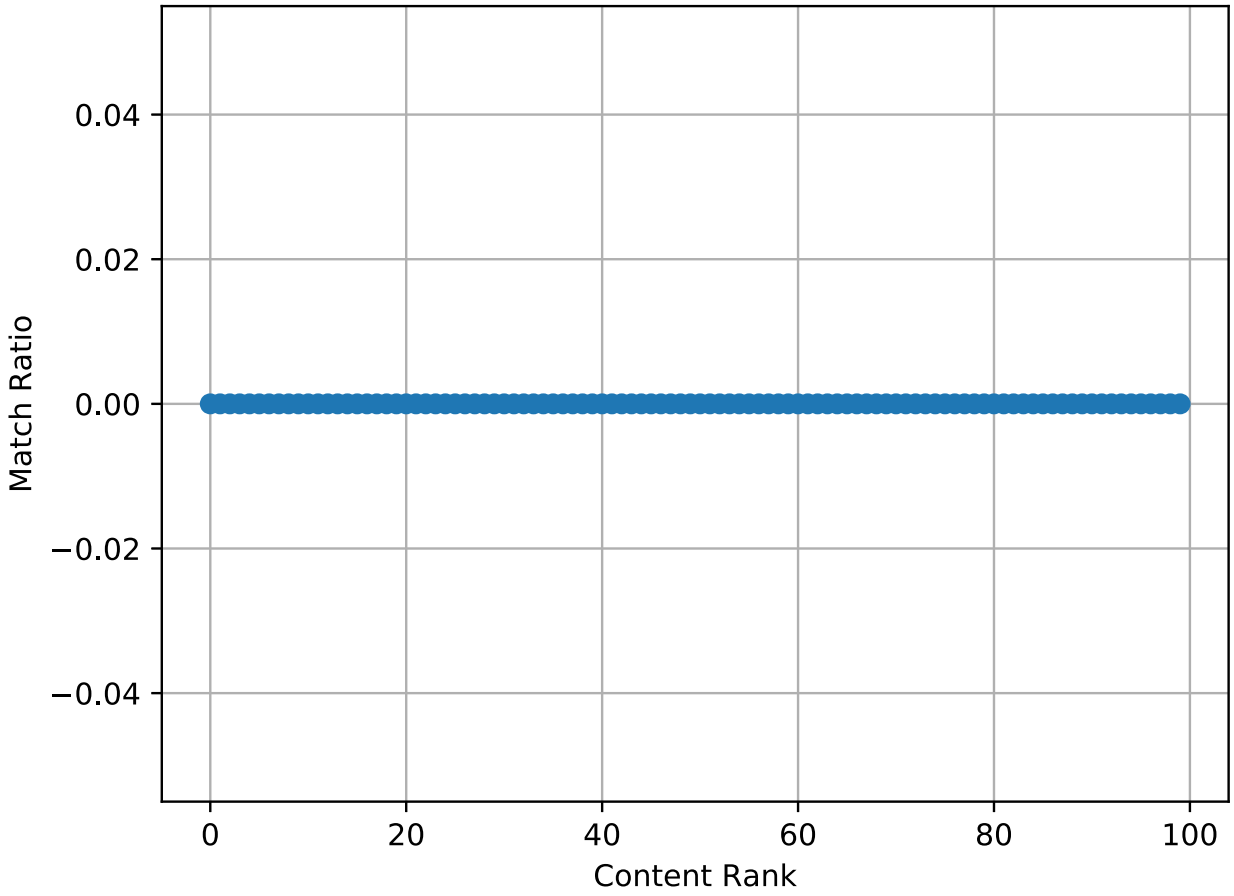
Global Eavesdropping Adversary

- Nefarious ISPs, nation states, etc.
- Questions:
  - To what extent does auxiliary information accuracy matter?
  - To what extent does universe size matter?

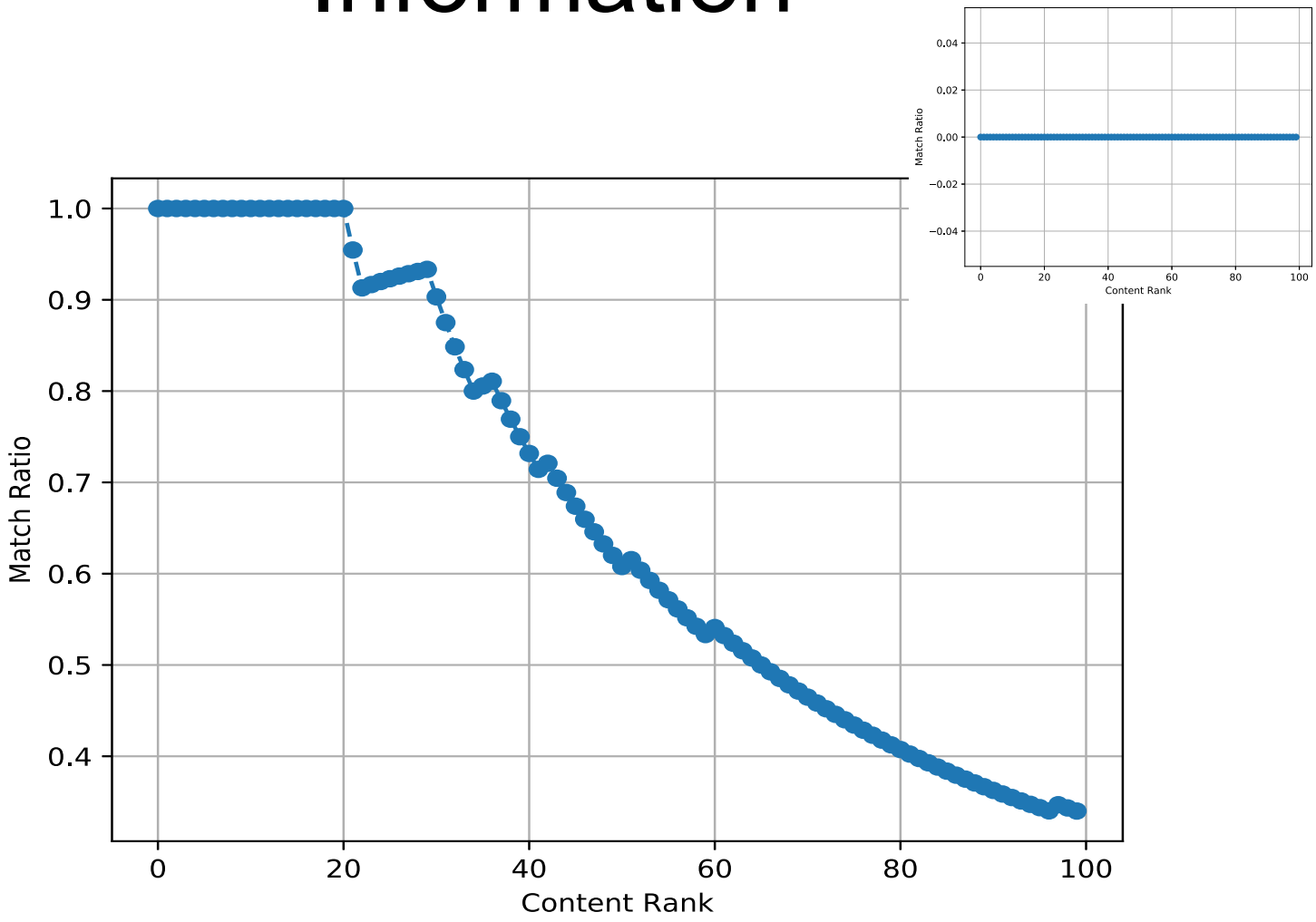
Topology



Different Auxiliary and Popularity Information



Matching Auxiliary and Popularity Information

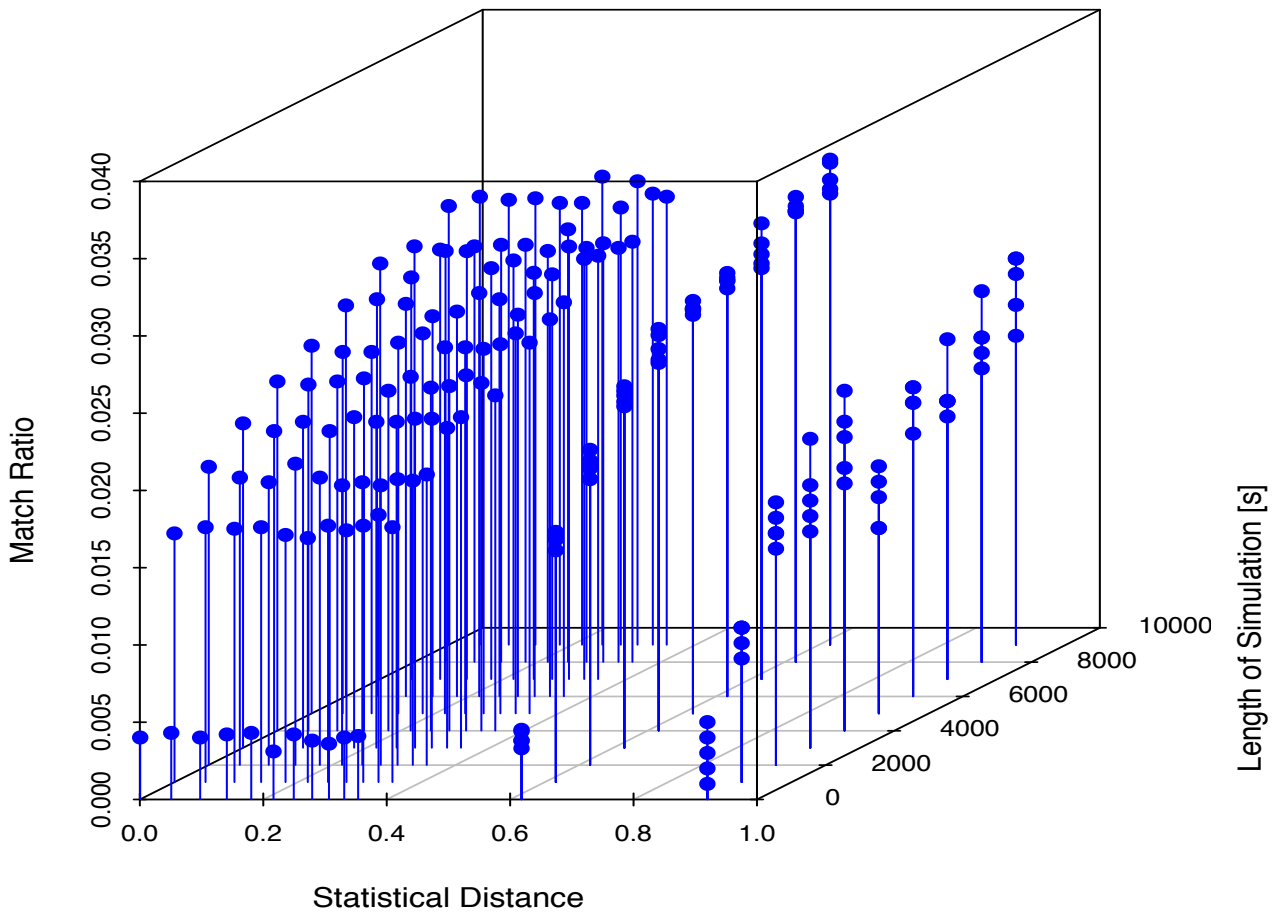


Takeaway

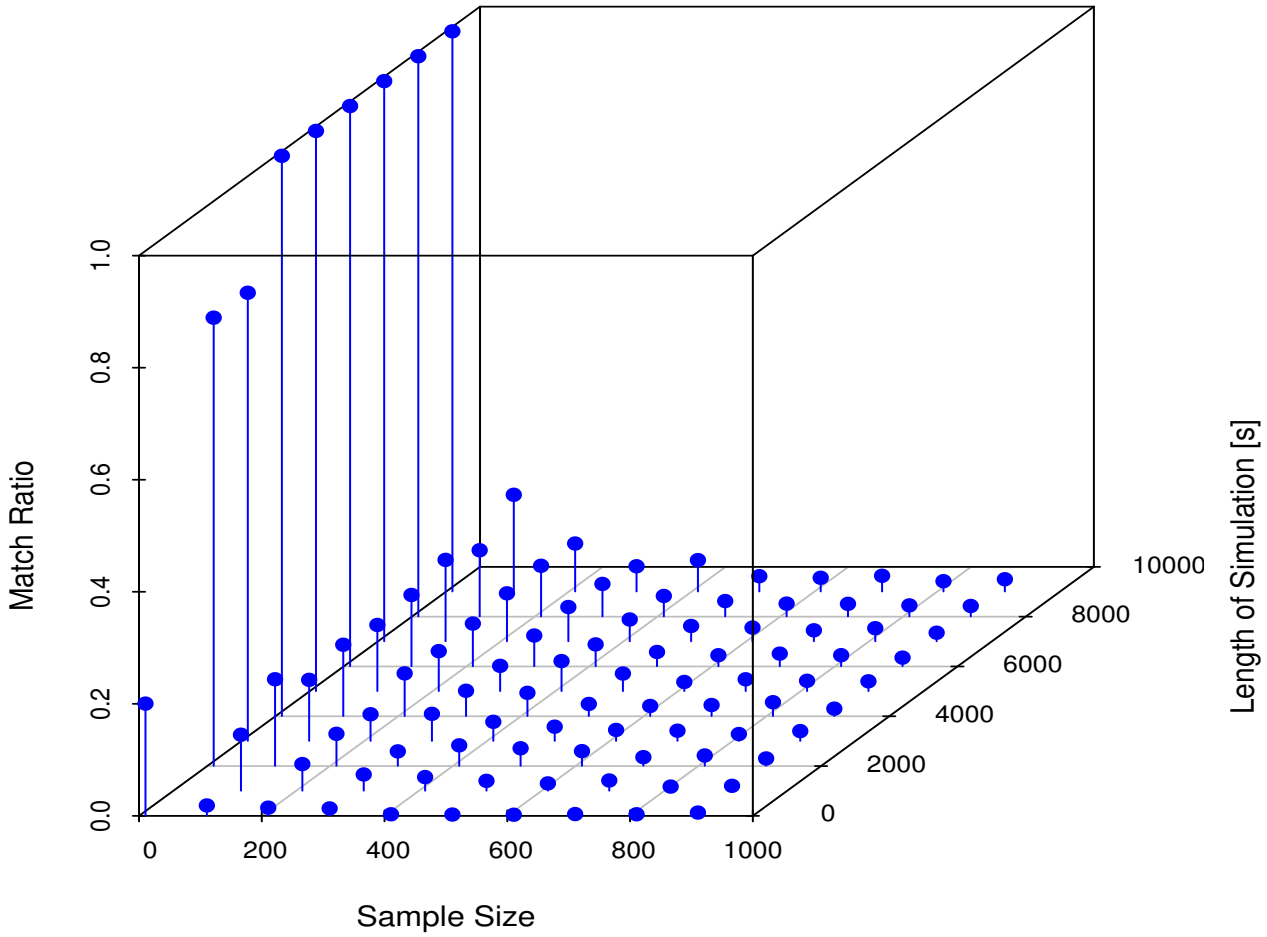
$$\Delta(\mathcal{D}_A^{\mathcal{A}}(\mathbb{P}), \mathcal{D}_R(\mathbb{P})) \approx 0.0$$

$$\Delta(\mathcal{D}_E(\mathbb{C}), \mathcal{D}_A^{\mathcal{A}}(\mathbb{P})) \approx 0.0$$

Auxiliary Information Gap



Content Universe Size



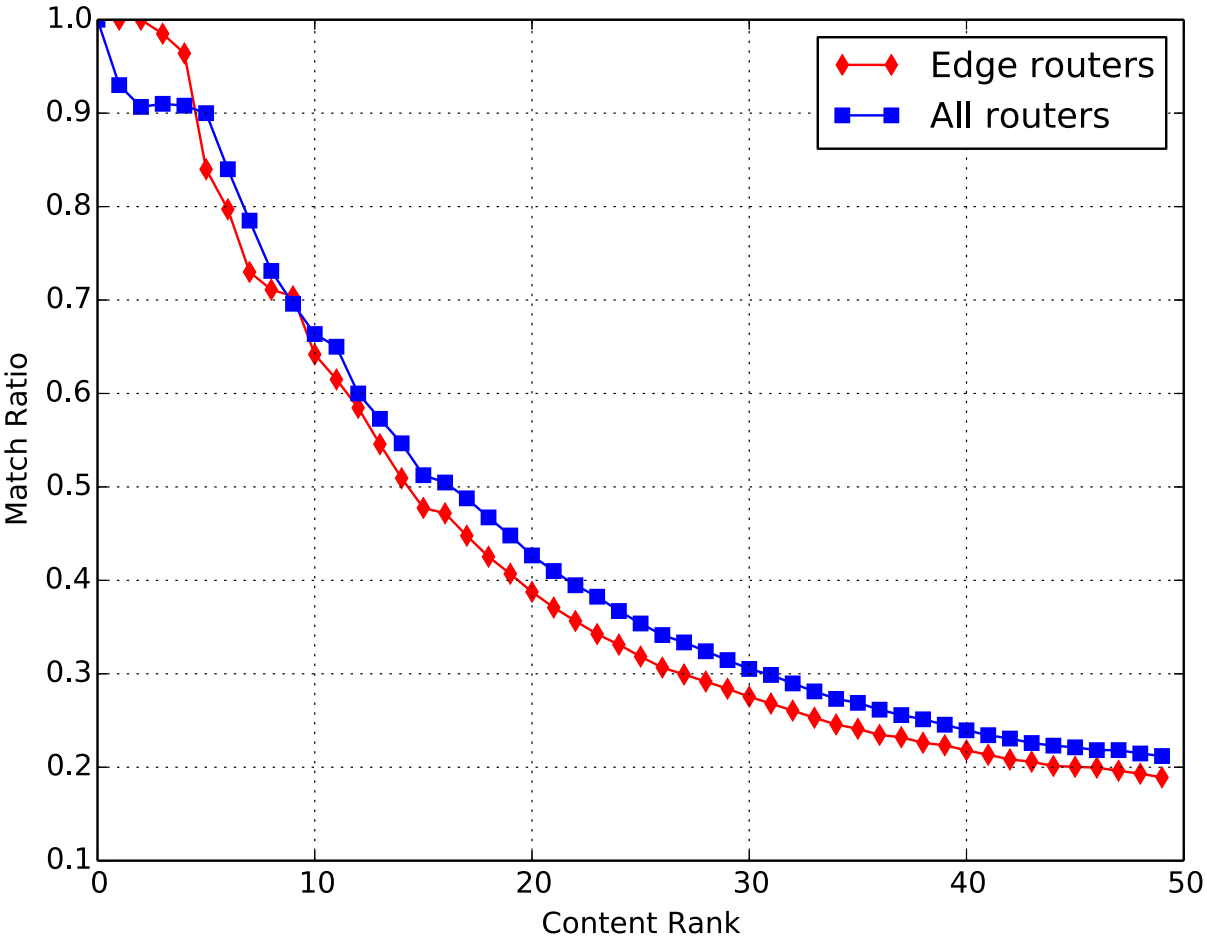
Takeaway

Auxiliary information accuracy is not as important as sample size

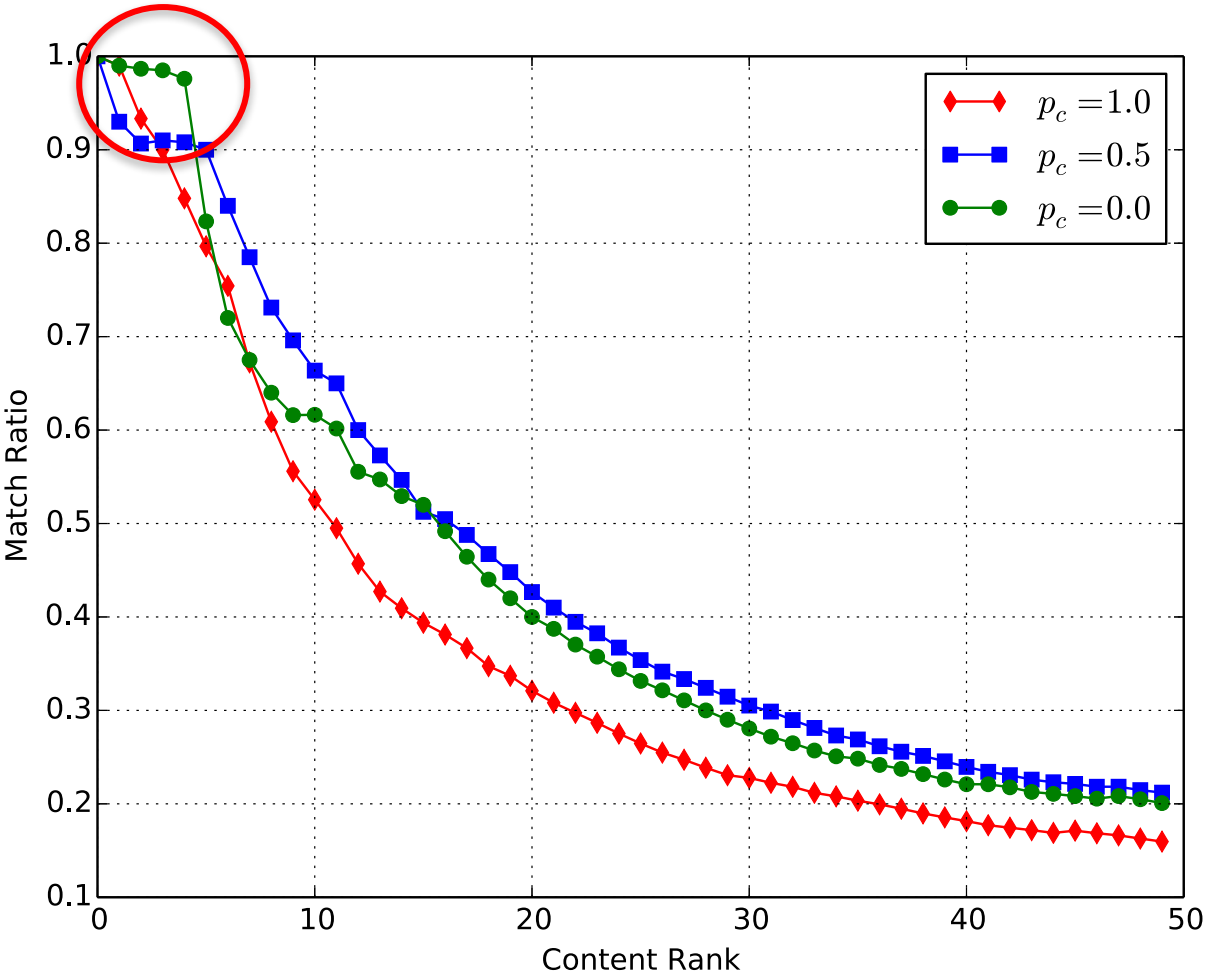
Distributed Adversary

- Access point, enterprise network middlebox, compromised transit router, etc.
- Questions:
  - Where does the adversary have the best chance at succeeding?
  - To what extent does caching dampen attack efficacy?
  - Can content replication (across different producers) help?

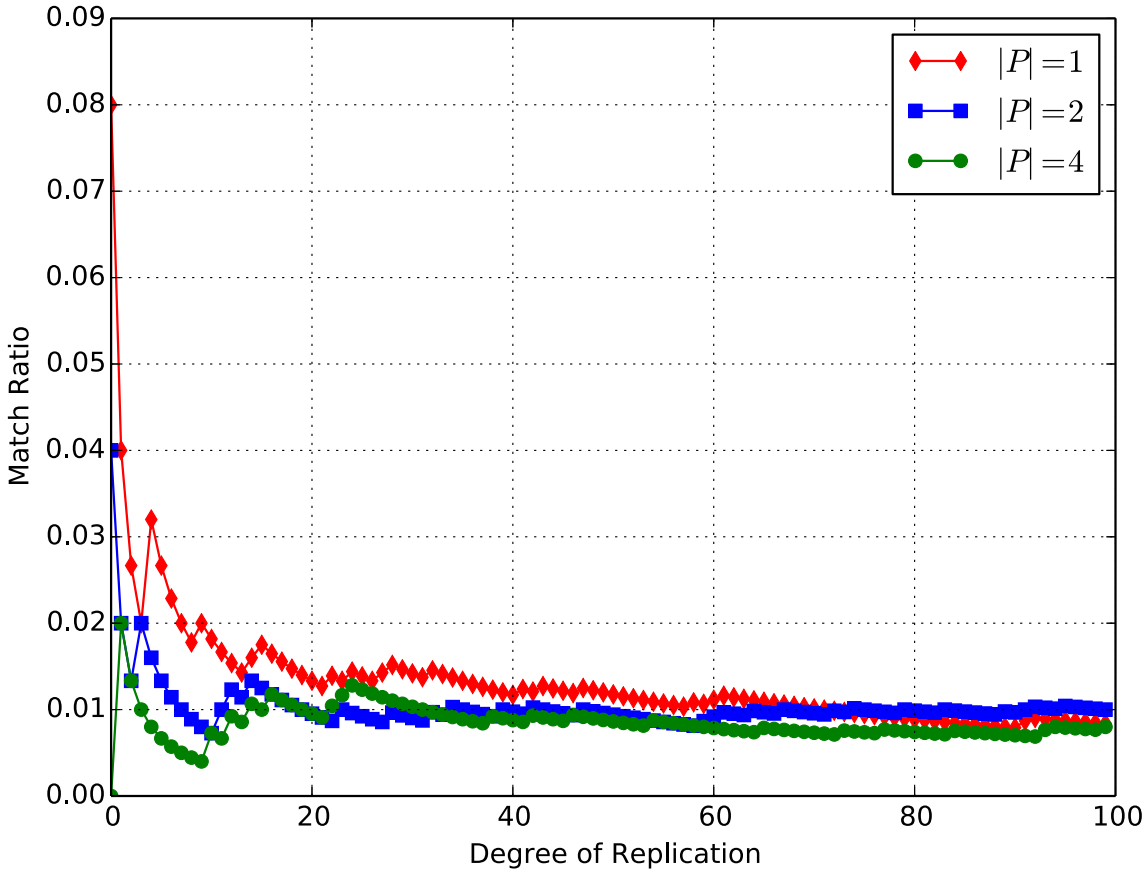
Edge vs Inner Router



Cache Presence



Replication



Probing for Popularity

- What does  do if it has no popularity information?
- Exploit caches to learn popularity
  - Assumes plaintext and ciphertext equivalents are fetched with equal distributions

Summary

- Caching both helps and hurts privacy
- Eavesdropping at the edge is enough
- Content replication helps bypass adversaries
- Preventing namespace enumeration is key to mitigating the attack

Future Work

- Expand simulator and widen experiments
- Analytically quantify the attack match percentage given distributions, network topologies, and cache hit probabilities
- Study attack on CDNs today

/this/is/the/end/version=0x00/chunk=0x01/PID=0x02

Questions?

Probing for Popularity

- What does  do if it has no popularity information?

Probing for Popularity

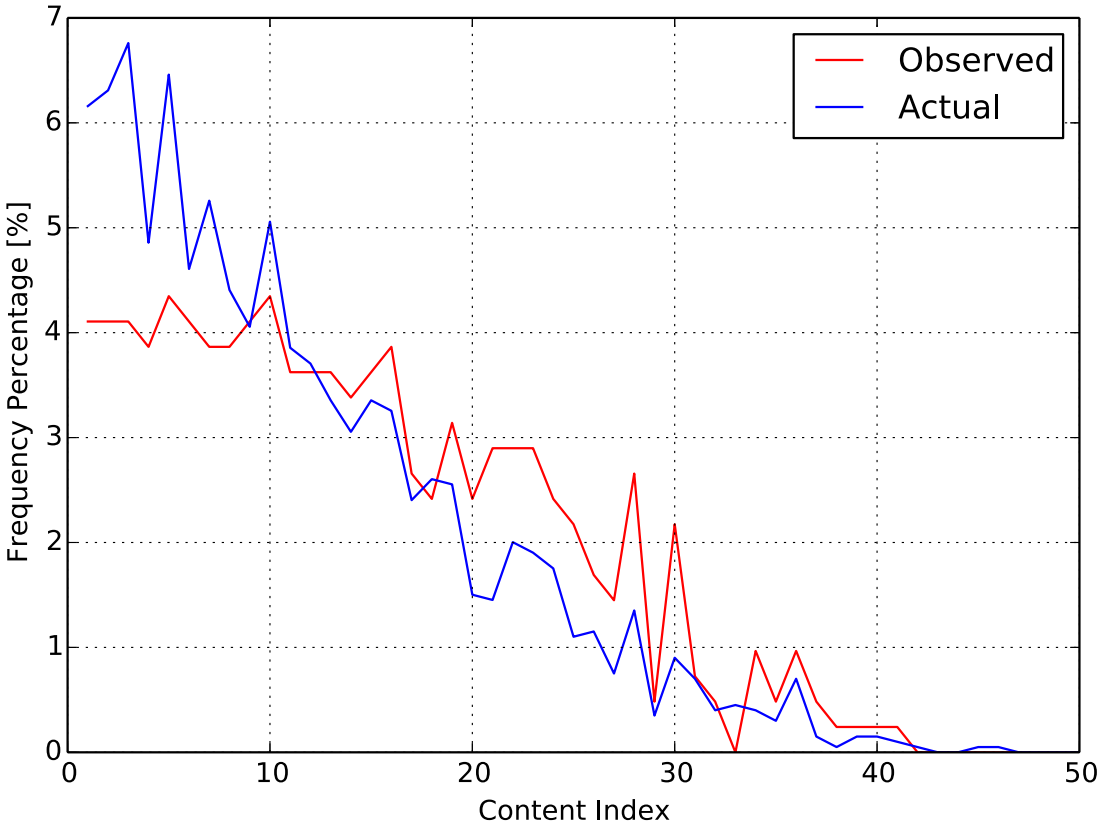
- What does  do if it has no popularity information?
- Exploit caches to learn popularity
  - Assumes plaintext and ciphertext equivalents are fetched with equal distributions

Probing Algorithm

Algorithm 1 InferPopularity

Input: $\mathcal{N}, r, t_c, \epsilon$
Output: $\alpha : \mathcal{N} \rightarrow \mathbb{N}$
for $N \in \mathcal{N}$ **do**
    $\alpha[N] = 0$
end for
for $i = 1, \ldots, r$ **do**
    for $N \in \mathcal{N}$ **do**
        $N_h = N$; $N_m = \text{AppendRandomComponent}(N, 128)$
        $t_N = \text{now}()$
        Send requests for N_h and N_m in parallel and record their time of arrival in t_N^h and t_N^m
        $\Delta_N = ||(t_N^h - t_N)| - |(t_N^m - t_N)||$
        if $\Delta_N > \epsilon$ **then**
            $\rho[N] = \rho[N] + 1$
        end if
        Sleep for t_c
    end for
end for
return α

Probe Results (S = 50)



Probe Results (S = 100)