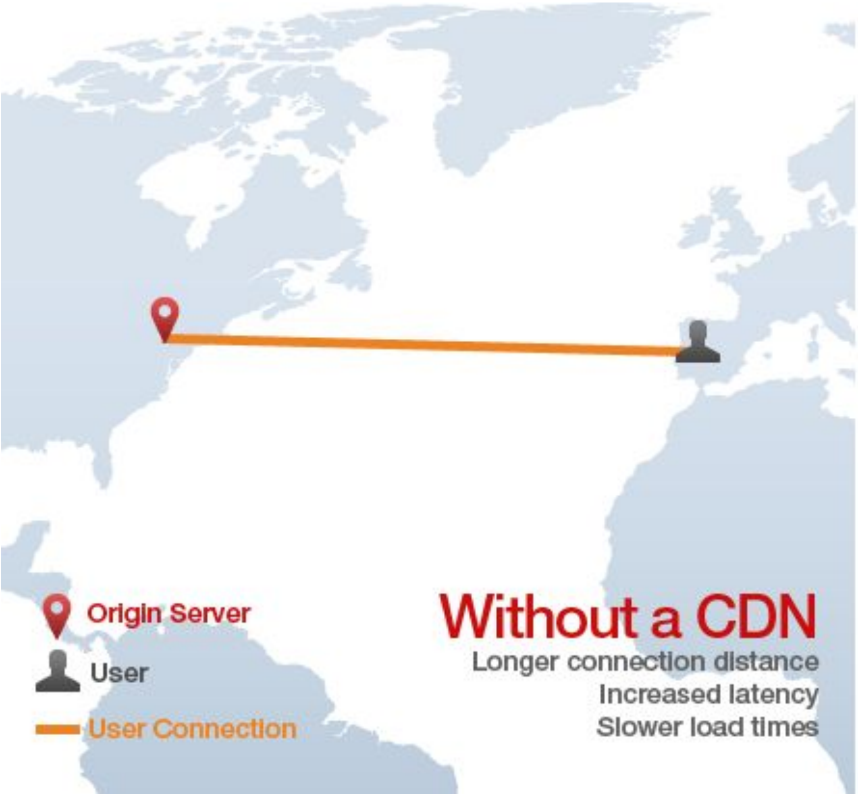
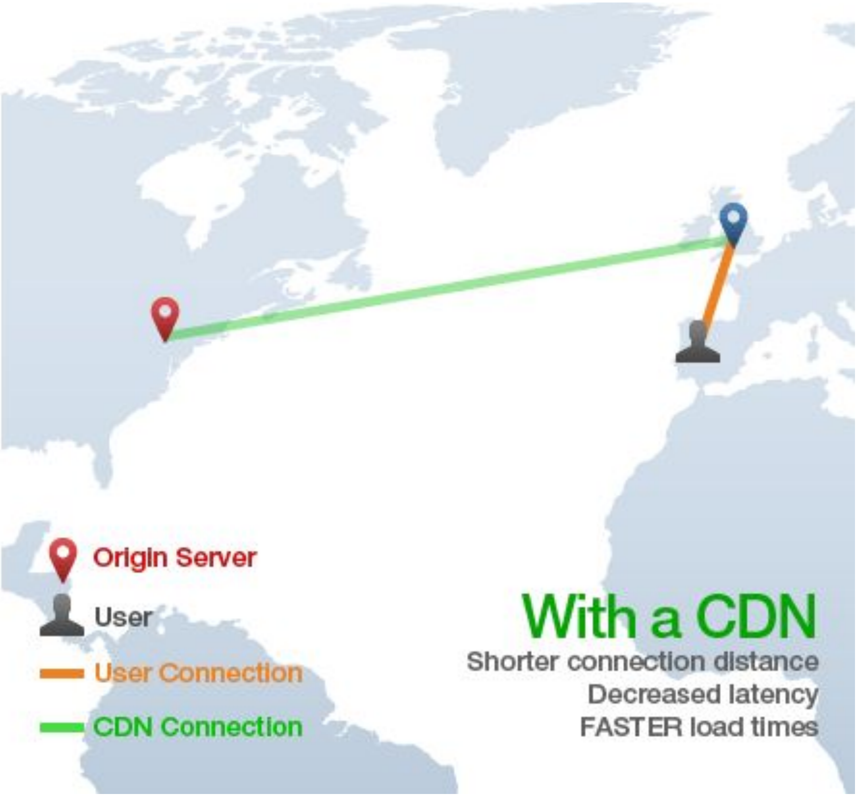
Probabilistic Approaches to Energy Conservation in CDNs

Artemas Radik, Andrew Palacci, Adarsh Hiremath

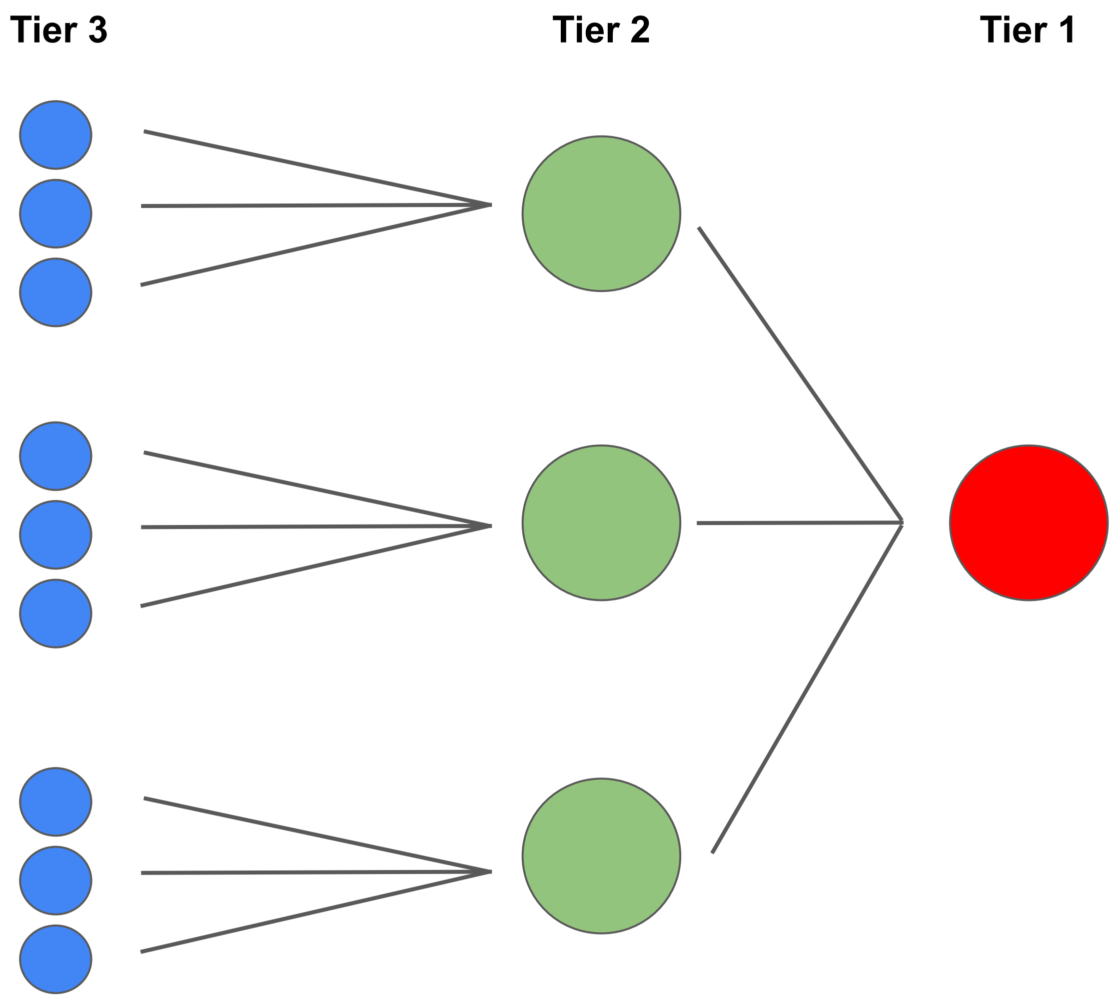
Background

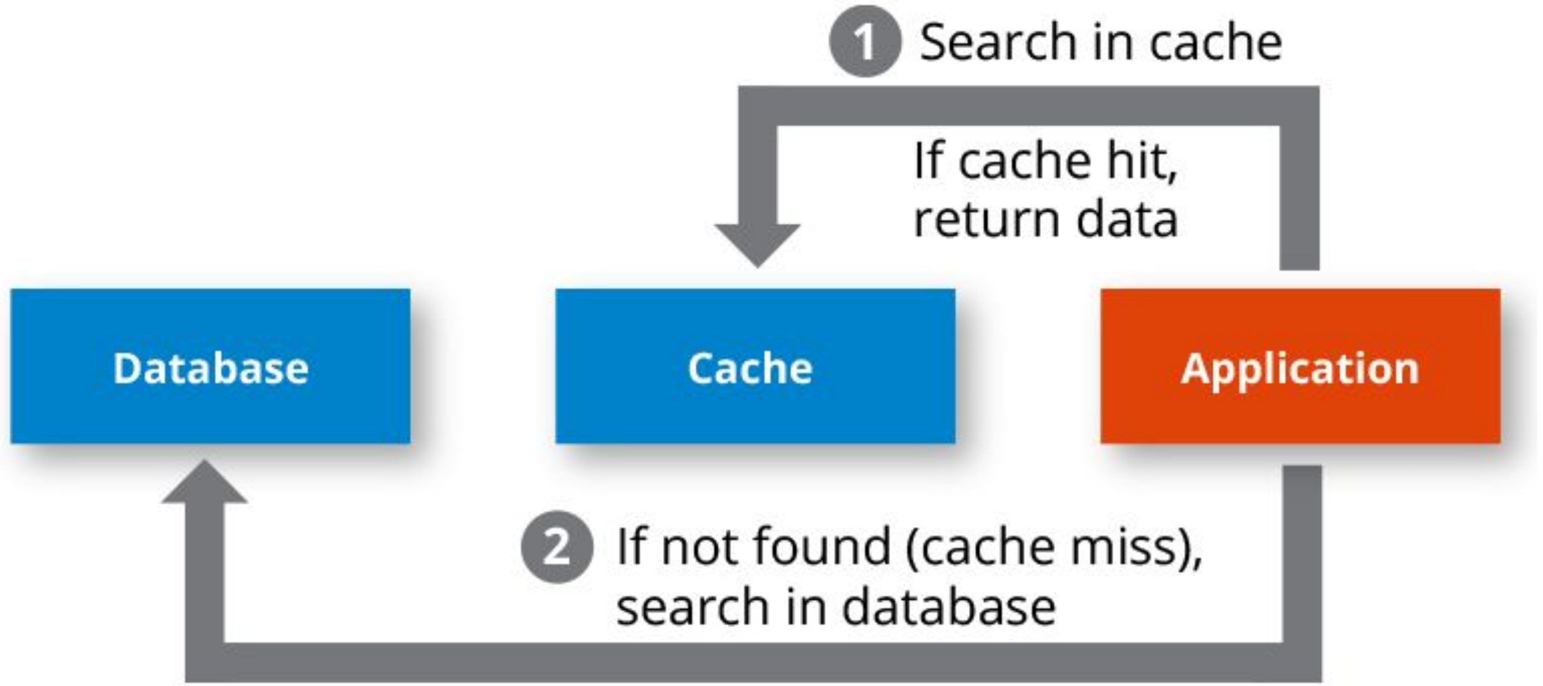


Origin Server
User
User Connection
Without a CDN
Longer connection distance
Increased latency
Slower load times



Origin Server
User
User Connection
CDN Connection
With a CDN
Shorter connection distance
Decreased latency
FASTER load times


Tier 3
Tier 2
Tier 1


1 Search in cache
If cache hit,
return data
Database
Cache
Application
2 If not found (cache miss),
search in database

Distributions

Uniform
Zipf
Poisson
Pareto
Bimodal

How to measure energy consumption?

$$E_{tot} = E_{storage} + E_{server} + E_{synch} + E_{tx}$$

[Bianco et al. 2016]

$$E_{storage} = \sum_{m} B n_m P_{st} t$$

B	size of each content
n_m	number of replicas for content *m*
P_{st}	storage power consumption per bit
t	time period of the analysis

$$E_{server} = \sum_{m} B r_m E_{sr}$$

B	size of each content
r_m	requests for content *m*
E_{sr}	server energy consumption per bit

$$E_{synch} = \sum_{m} B\, m_m n_m [E_r (H_{ps} + 1) + E_l H_{ps}]$$

B	size of each content
m_m	modifications to content *m*
n_m	number of replica for content *m*
E_r	router energy consumption per bit
H_{ps}	hops from primary to surrogate servers
E_l	link energy consumption per bit

$$P_A = \frac{S}{T_2} \cdot \frac{1}{g_3} \cdot P_{hit} \qquad P_B = \frac{S}{T_2} \cdot \left(1 - \frac{1}{g_3}\right) \cdot P_{hit} \qquad P_C = 1 - (P_A + P_B)$$

S	# of surrogate servers
T_2	# of Tier 2 ISPs
g_3	# of Tier 3 ISPs connected to Tier 2 ISP
P_{hit}	cache hit probability

$$
\begin{aligned}
E_{tx} = P_A \sum_m Br_m[E_r(H_{sd}^A + 1) + E_l H_{sd}^A] + \\
P_B \sum_m Br_m[E_r(H_{sd}^B + 1) + E_l H_{sd}^B] + \\
P_C \sum_m Br_m[E_r(H_{sd}^C + 1) + E_l H_{sd}^C]
\end{aligned}
$$

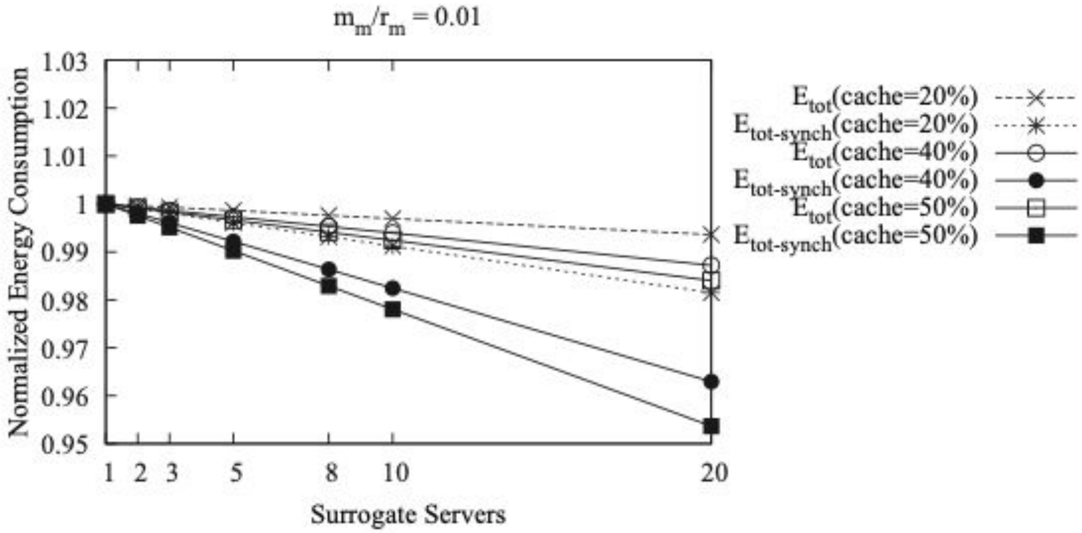
B	size of each content
r_m	requests for content *m*
E_r	router energy consumption per bit
E_l	link energy consumption per bit
H^A_{sd}	hops to fetch content from same T3 ISP
H^B_{sd}	hops to fetch content from same T2 ISP
H^C_{sd}	hops to fetch content from core network

$$E_{tot} = E_{storage} + E_{server} + E_{synch} + E_{tx}$$

How do we minimize E_{tot}?

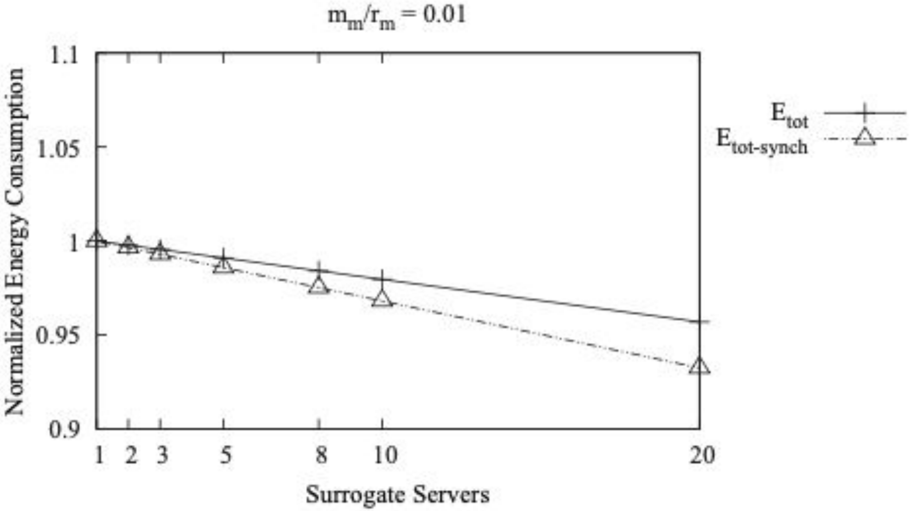
Cache Policies

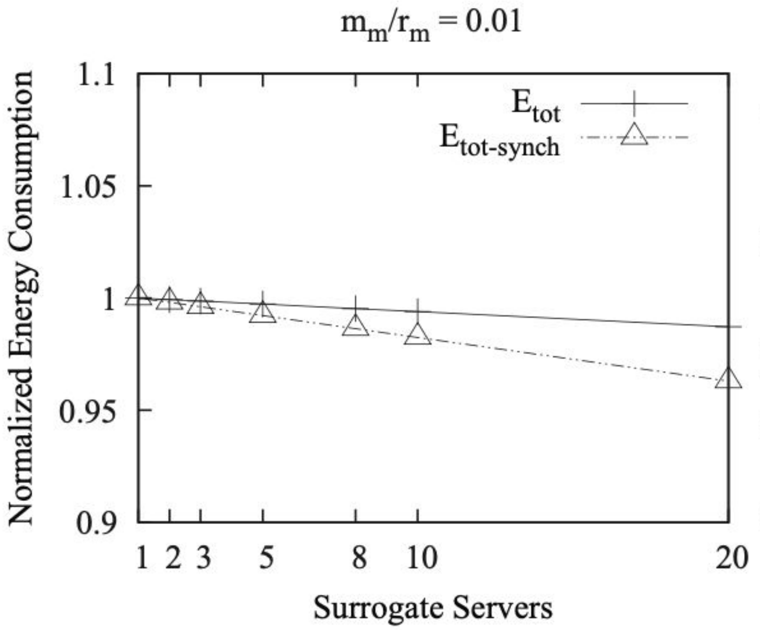
Uniform



$m_m/r_m = 0.01$
1.03
1.02
1.01
1
0.99
0.98
0.97
0.96
0.95
Normalized Energy Consumption
1 2 3 5 8 10 20
Surrogate Servers
E_{tot}(cache=20%)
$E_{tot\text{-}synch}$(cache=20%)
E_{tot}(cache=40%)
$E_{tot\text{-}synch}$(cache=40%)
E_{tot}(cache=50%)
$E_{tot\text{-}synch}$(cache=50%)

Zipf Popularity-Based





m_m/r_m = 0.01
m_m/r_m = 0.1
Normalized Energy Consumption
1.1
1.05
1
0.95
0.9
1.3
1.25
1.2
1.15
E_tot
E_tot-synch
1 2 3 5 8 10 20
Surrogate Servers

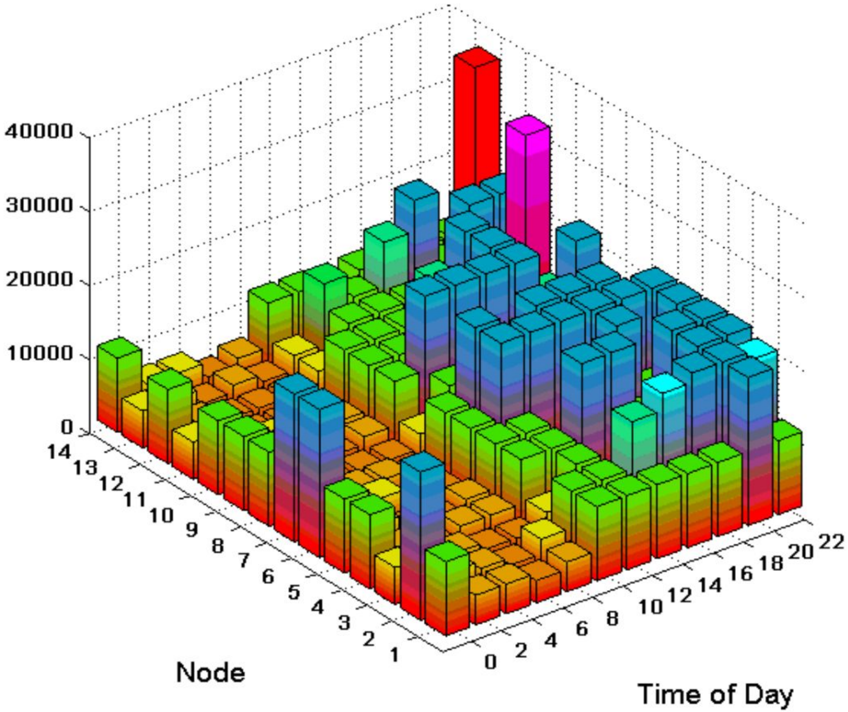
Zipf is Obsolete

TikTok

Pareto / Bimodal Caching to Improve Energy Efficiency

Mixed-Integer Linear Programming (MILP) For Variable Cache Sizes

$$\sum_{t\in T}\left(\sum_{i\in N} Pp\left(AP_{it} + \sum_{j\in Nm_i: i\neq j} C_{ijt}\right) + \sum_{i\in N} Po_{it} + \sum_{i\in N}\sum_{j\in Nm_i} Pt\cdot w_{ijt} + \sum_{i\in N}\sum_{j\in Nm_i} Pa\cdot Amp_{ij}\cdot f_{ij} + \sum_{i\in N}\sum_{j\in Nm_i} Pmd\cdot f_{ij} + \sum_{i\in N} \emptyset M \right)$$



[Osman 2014]

Project direction

- Mathematically modeling the efficiency of Pareto / Bimodal cache policies
- Running model simulations to verify predictions for various use cases
- Considering other scenarios where using CDNs increases E_{tot}
- Allowing for variable cache storage (MILP)
- Extensions to other CDN architectures besides 3-tier

References

Bianco, A., Mashayekhi, R., & Meo, M. (2016). Energy consumption for data distribution in content delivery networks. The Institute of Electrical and Electronics Engineers, Inc. (IEEE) Conference Proceedings, 1.

Breslau, L., Pei Cao, Li Fan, Phillips, G., & Shenker, S. (1999). Web caching and Zipf-like distributions: evidence and implications. IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future Is Now (Cat. No.99CH36320), 1, 126–134 vol.1. https://doi.org/10.1109/INFCOM.1999.749260

Osman, N. I. (2014). Energy efficiency in content delivery networks. ProQuest Dissertations Publishing.