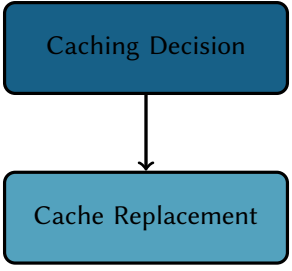
Performance Comparison of Caching Strategies for Information-Centric IoT

Jakob Pfender, Alvin Valera, Winston Seah

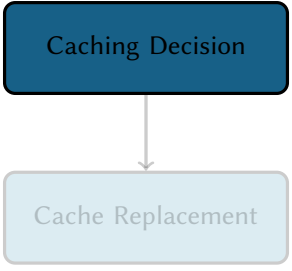
School of Engineering and Computer Science
Victoria University of Wellington, New Zealand

September 22, 2018

Traditional Caching Strategies



Traditional Caching Strategies



Caching Decision — Traditional Approaches

Cache Everything Everywhere (CEE)

- Feasible in traditional ICN thanks to large caches
- Fastest possible propagation of content through network → rapid replication

Caching Decision — Traditional Approaches

Cache Everything Everywhere (CEE)

- Feasible in traditional ICN thanks to large caches
- Fastest possible propagation of content through network $\rightarrow$ rapid replication
- High redundancy
- Non-optimal resource utilisation

Caching Decision — Traditional Approaches

Cache Everything Everywhere (CEE)

- Feasible in traditional ICN thanks to large caches
- Fastest possible propagation of content through network $\rightarrow$ rapid replication
- High redundancy
- Non-optimal resource utilisation

Probabilistic Caching $(Prob(P))$

- Increases cache diversity across the network
- More popular content likelier to be stored

Caching Decision — Traditional Approaches

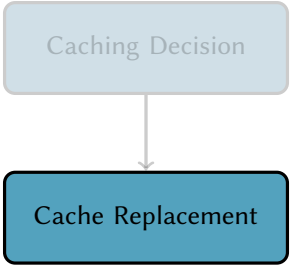
Cache Everything Everywhere (CEE)

- Feasible in traditional ICN thanks to large caches
- Fastest possible propagation of content through network $\rightarrow$ rapid replication
- High redundancy
- Non-optimal resource utilisation

Probabilistic Caching $(Prob(P))$

- Increases cache diversity across the network
- More popular content likelier to be stored
- Desirability of diversity depends on application scenario
- If request pattern has strong skew, diversity hurts performance by wasting resources on unpopular content
- The more uniform the distribution, the more beneficial diversity

Traditional Caching Strategies



Cache Replacement Decision — Traditional Approaches

- **Least Recently Used (LRU):**
  - Unpopular / outdated content more likely to be removed
  - Generally effective, but danger of thrashing
  - Alternative: Least Frequently Used (LFU) → avoids thrashing, performs poorly with variable access patterns & request spikes

Cache Replacement Decision — Traditional Approaches

- **Least Recently Used (LRU):**
  - Unpopular / outdated content more likely to be removed
  - Generally effective, but danger of thrashing
  - Alternative: Least Frequently Used (LFU) $\rightarrow$ avoids thrashing, performs poorly with variable access patterns & request spikes
- **Random Replacement (RR):**
  - Evict a randomly chosen content chunk
  - Simple, fast, no overhead
  - Some argue that cache replacement should be performed as fast as possible
    - Simple and fast more desirable than effective but complex

Traditional ICN caching vs. IoT

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power
  - Available cache space extremely valuable

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power
  - Available cache space extremely valuable
- Unreliable links

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power
  - Available cache space extremely valuable
- Unreliable links
- Small, transient data

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power
  - Available cache space extremely valuable
- Unreliable links
- Small, transient data
- Request distributions tend to be uniform

Traditional ICN caching vs. IoT

Lessons from traditional ICN caching

- CEE is inefficient (high redundancy, low diversity, poor utilisation of resources)
- Caching less $\rightarrow$ better performance
- Cache diversity generally desirable
- Cache replacement policies should be as fast & simple as possible

Key differences in IoT

- Limited memory and processing power
  - Available cache space extremely valuable
- Unreliable links
- Small, transient data
- Request distributions tend to be uniform

Can we apply the lessons from traditional ICN caching to the IoT?

Advanced Caching Strategies (for the IoT)

Advanced Caching Strategies

Dynamic Caching Probability

- Dynamically compute caching probability for each node and/or each content chunk, based on available information
- Caching behaviour adapts to the state of the network

Advanced Caching Strategies

Dynamic Caching Probability

- Dynamically compute caching probability for each node and/or each content chunk, based on available information
- Caching behaviour adapts to the state of the network
- Example: **pCASTING** (Hail *et al.* 2015)
  - Consider content age, node battery, cache occupancy
  - Values normalised & weighted by relative importance
  - Fully distributed, no communication overhead
  - Uses purely local information

Advanced Caching Strategies

Dynamic Caching Probability

- Dynamically compute caching probability for each node and/or each content chunk, based on available information
- Caching behaviour adapts to the state of the network
- Example: **pCASTING** (Hail *et al.* 2015)
  - Consider content age, node battery, cache occupancy
  - Values normalised & weighted by relative importance
  - Fully distributed, no communication overhead
  - Uses purely local information

Max Diversity Most Recent (MDMR)

- Hahm *et al.* 2017
- Cache replacement strategy developed specifically for information-centric IoT
- Aim: Maximise diversity while maintaining freshness
- Always attempt to replace older content from same producer as new content, otherwise oldest chunk from producer with more than one chunk in cache, else oldest chunk

Performance Metrics

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network
- **Data retrieval delay**
  - Also affected by network congestion, density, etc.

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network
- **Data retrieval delay**
  - Also affected by network congestion, density, etc.
- **Interest retransmission ratio**
  - On-path caching may reduce hop count for retransmissions

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network
- **Data retrieval delay**
  - Also affected by network congestion, density, etc.
- **Interest retransmission ratio**
  - On-path caching may reduce hop count for retransmissions
- **Total cache evictions**
  - How well does the strategy adapt to content popularity & propagation?

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network
- **Data retrieval delay**
  - Also affected by network congestion, density, etc.
- **Interest retransmission ratio**
  - On-path caching may reduce hop count for retransmissions
- **Total cache evictions**
  - How well does the strategy adapt to content popularity & propagation?
- **Diversity Metric (DM)**
  - $D = \frac{|C_{disj}|}{|S|}$
    - $|S|$: Number of content producers
    - $|C_{disj}|$: Number of disjoint name prefixes in all caches

Performance Metrics

- **Server load / cache hit ratio**
  - Indicates how well popular content is distributed across the network
- **Data retrieval delay**
  - Also affected by network congestion, density, etc.
- **Interest retransmission ratio**
  - On-path caching may reduce hop count for retransmissions
- **Total cache evictions**
  - How well does the strategy adapt to content popularity & propagation?
- **Diversity Metric (DM)**
  - $D = \frac{|C_{disj}|}{|S|}$
    - $|S|$: Number of content producers
    - $|C_{disj}|$: Number of disjoint name prefixes in all caches
- **Cache Retention Ratio (CRR)**
  - Measures ratio of distinct objects in caches to all generated objects:
    - $C = \frac{D_q}{D_p}$

Evaluation

Evaluation — Experiment Setup

- All experiments conducted on **FIT IoT-LAB** open testbed using **M3 nodes**
  - STM32 (ARM Cortex M3), 512 kB ROM, 64 kB RAM, Atmel AT86RF231 2.4 GHz transceiver on IEEE 802.15.4
- Simple **RIOT** application using **CCN-lite** as ICN implementation, modified to support the different caching strategies

Evaluation — Experiment Setup

- All experiments conducted on **FIT IoT-LAB** open testbed using **M3 nodes**
  - STM32 (ARM Cortex M3), 512 kB ROM, 64 kB RAM, Atmel AT86RF231 2.4 GHz transceiver on IEEE 802.15.4
- Simple **RIOT** application using **CCN-lite** as ICN implementation, modified to support the different caching strategies
- 60 M3 nodes distributed evenly in a single building
- Multihop setup, average path length 2–3 hops

Evaluation — Experiment Setup

- All experiments conducted on **FIT IoT-LAB** open testbed using **M3 nodes**
  - STM32 (ARM Cortex M3), 512 kB ROM, 64 kB RAM, Atmel AT86RF231 2.4 GHz transceiver on IEEE 802.15.4
- Simple **RIOT** application using **CCN-lite** as ICN implementation, modified to support the different caching strategies
- 60 M3 nodes distributed evenly in a single building
- Multihop setup, average path length 2–3 hops
- Prefix announcements recorded with hop count and rebroadcast with increased hop count
- Interests forwarded according to lowest hop count, broadcast fallback

Evaluation — Experiment Setup

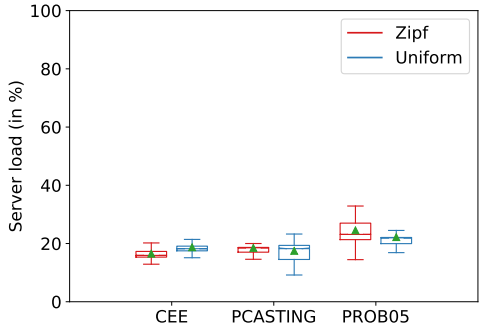
- All experiments conducted on **FIT IoT-LAB** open testbed using **M3 nodes**
  - STM32 (ARM Cortex M3), 512 kB ROM, 64 kB RAM, Atmel AT86RF231 2.4 GHz transceiver on IEEE 802.15.4
- Simple **RIOT** application using **CCN-lite** as ICN implementation, modified to support the different caching strategies
- 60 M3 nodes distributed evenly in a single building
- Multihop setup, average path length 2–3 hops
- Prefix announcements recorded with hop count and rebroadcast with increased hop count
- Interests forwarded according to lowest hop count, broadcast fallback
- Nodes produce random content chunks prefixed with their ID every 1–5 seconds
- Nodes request random existing content every 0.5–1.5 seconds, using uniform or Zipfian pattern

Evaluation — Caching decision policies

Evaluation — Caching decision policies

Server load

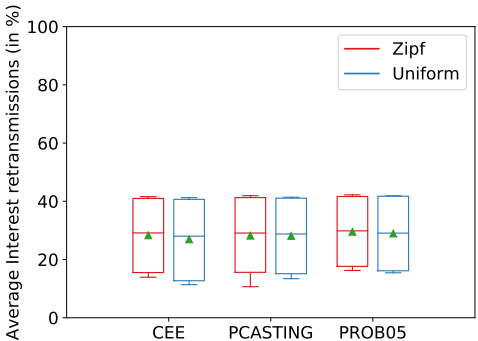
- Probabilistic caching results in lower probability that given content can be found in given cache
- pCASTING seems to cache with higher average probability than $p = 0.5$
- CEE is better for skewed request pattern because it increases replication of popular content



Evaluation — Caching decision policies

Interest retransmission ratio

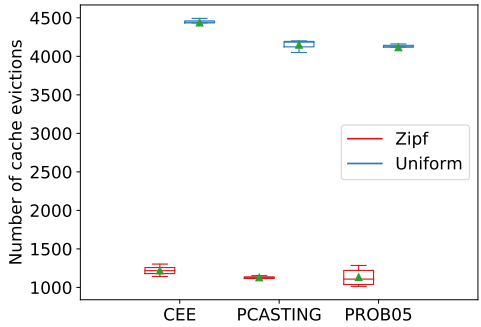
- No statistically significant effect
- Influence of network factors (topology, congestion) stronger than caching policy
- Previous work (Hail *et al.*) found more significant differences using simulation



Evaluation — Caching decision policies

Cache evictions

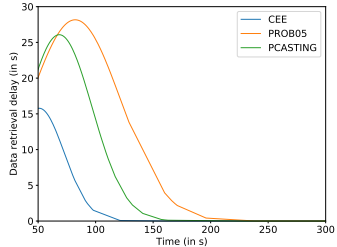
- Correlated with rate at which caches are filled
- Skewed distribution → reduced thrashing
- Total number of transmissions during experiment run: 40,000-75,000
  - Uniform requests: 5%-12% evictions
  - Zipfian requests: 1%-4%
- Strong dependence on cache size

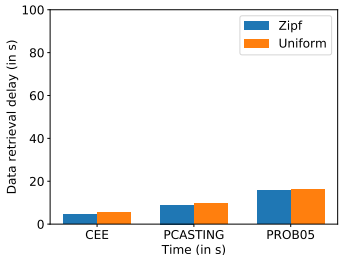


Evaluation — Caching decision policies

Data retrieval delay

- Large initial latency because content takes time to propagate
- Cache hit probability lower for probabilistic policies → higher peak and slower decline as caches take longer to fill
- All strategies reach minimal delay very quickly

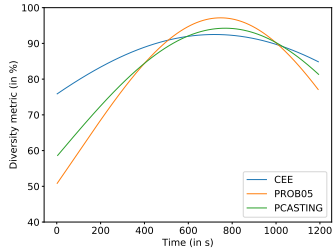


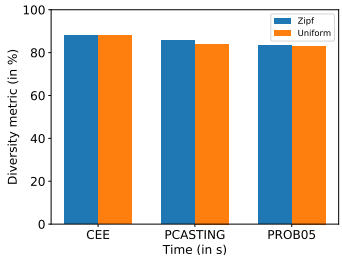


Evaluation — Caching decision policies

Diversity metric

- Majority of content producers represented in network at any given time
- CEE caches all contents from the start → caches fill more quickly
  - But since everything is cached, cache contents are highly redundant
- Probabilistic methods have higher probability of representing all content producers

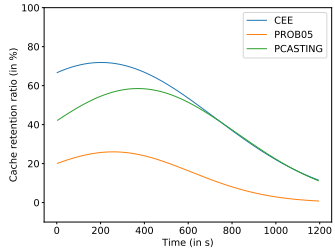


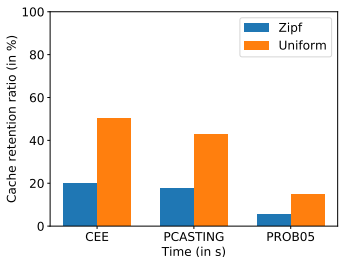


Evaluation — Caching decision policies

Cache Retention Ratio

- Content inevitably fades from network after a while
- Content creation rate is constant, but probabilistic approaches take longer to fill caches
- More diverse caches → delayed decline
- Zipfian skew means less popular content fades much quicker → caches become more homogeneous over time
- Uniform distribution ensures similar lifespans for all content



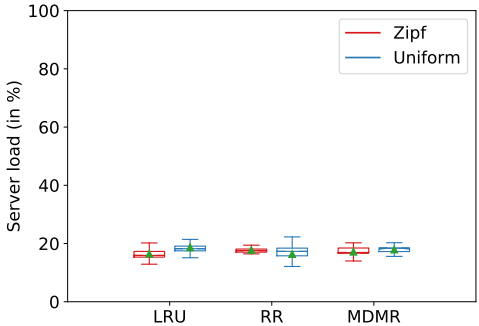


Evaluation — Cache replacement policies

Evaluation — Cache replacement policies

Server load

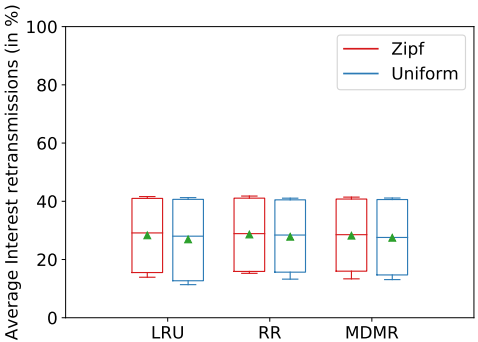
- Cache replacement strategy has negligible effect on server load
- RR performs on par with much more complex strategies



Evaluation — Cache replacement policies

Interest retransmission ratio

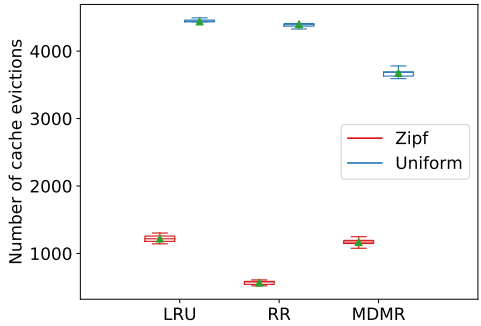
- No statistically significant effect
- Thus: Any observed variations in retransmission rates have causes independent from caching strategy



Evaluation — Cache replacement policies

Cache evictions

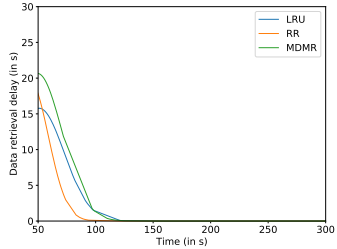
- Significant effect of distribution
- MDMR explicitly designed to maximise diversity for uniform patterns, but can't outperform LRU given skewed distribution
  - Zipfian distribution already favours popular content, reducing impact of cache-shaping strategies
- Surprising: RR outperforms other approaches given Zipfian distribution
  - Thrashing at popularity tail end counteracted by RR?

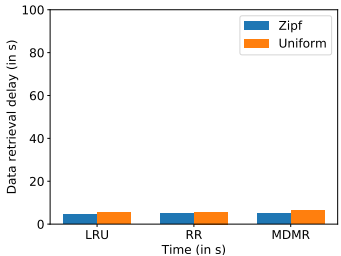


Evaluation — Cache replacement policies

Data retrieval delay

- MDMR prioritises content freshness
  - Freshness only becomes significant factor after some time has elapsed ⇒ MDMR requires minimum time to become effective
- Significance of early spikes mostly in terms of strain on individual nodes, especially if devices battery-powered

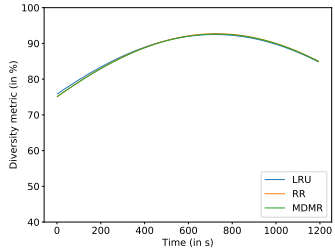


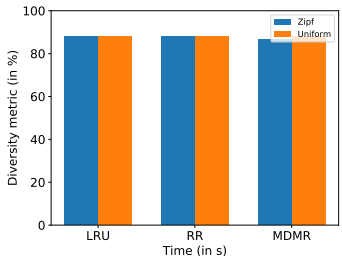


Evaluation — Cache replacement policies

Diversity Metric

- Why no impact?

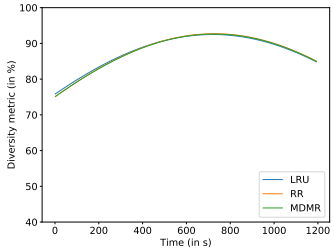


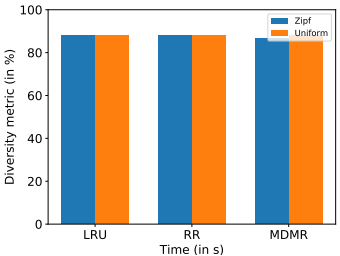


Evaluation — Cache replacement policies

Diversity Metric

- Why no impact?
- Note: MDMR intended for scenario with periodic sleeping, which is not the case here

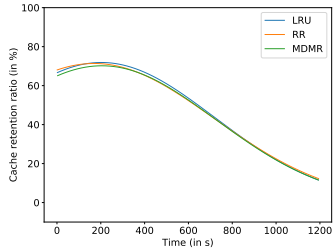


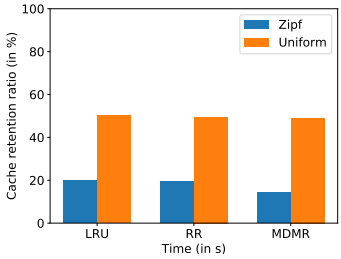


Evaluation — Cache replacement policies

Cache Retention Ratio

- Content lifetime solely influenced by caching decision, not by replacement
- Faster fading given skewed distribution unsurprising





Revisiting lessons from traditional ICN

Revisiting lessons from traditional ICN

- Caching less $\rightarrow$ better performance

Revisiting lessons from traditional ICN

- Caching less $\rightarrow$ better performance — **Yes**, at least in terms of diversity

Revisiting lessons from traditional ICN

- Caching less $\rightarrow$ better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable

Revisiting lessons from traditional ICN

- Caching less $\rightarrow$ better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable — **Yes**, if request patterns are uniform

Revisiting lessons from traditional ICN

- Caching less $\rightarrow$ better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable — **Yes**, if request patterns are uniform
- CEE is inefficient

Revisiting lessons from traditional ICN

- Caching less → better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable — **Yes**, if request patterns are uniform
- CEE is inefficient — **Depends on scenario**

Revisiting lessons from traditional ICN

- Caching less → better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable — **Yes**, if request patterns are uniform
- CEE is inefficient — **Depends on scenario**
- Stateless cache replacement policies are sufficient

Revisiting lessons from traditional ICN

- Caching less → better performance — **Yes**, at least in terms of diversity
- Cache diversity generally desirable — **Yes**, if request patterns are uniform
- CEE is inefficient — **Depends on scenario**
- Stateless cache replacement policies are sufficient — ✓

Further conclusions

Further conclusions

- The performance of simple stateless strategies is **encouraging**, because it means effective caching can be achieved even on resource-constrained IoT devices

Further conclusions

- The performance of simple stateless strategies is **encouraging**, because it means effective caching can be achieved even on resource-constrained IoT devices
- Identifying ideal caching decision strategy depends on application & resources

Further conclusions

- The performance of simple stateless strategies is **encouraging**, because it means effective caching can be achieved even on resource-constrained IoT devices
- Identifying ideal caching decision strategy depends on application & resources
- Is data homogeneous & needs to be distributed as rapidly as possible or is cache diversity more important than response time?
- How much strain can we afford to place on individual nodes?

Further conclusions

- The performance of simple stateless strategies is **encouraging**, because it means effective caching can be achieved even on resource-constrained IoT devices
- Identifying ideal caching decision strategy depends on application & resources
- Is data homogeneous & needs to be distributed as rapidly as possible or is cache diversity more important than response time?
- How much strain can we afford to place on individual nodes?
- Results presented here offer no universal solution

Future work

- One experimental setup cannot reflect the diversity of IoT applications
- Many more proposed caching decision strategies
- Take into account topological factors, content popularity, other aspects
- More comprehensive survey/evaluation desirable

Thank you!