Performance of RAN Caching in LTE Mobile Networks: a Real Traffic Analysis

Presenter: Tao Lin

Tao Lin[†], Hongjia Li[†], Haiyong Xie[‡], Jiasi Chen[#], Huajun Cui[†], Guoqiang Zhang[*], Wei An[†] and Yang Li[†]

[†] Institute of Information Engineering, Chinese Academy of Sciences, China

[‡] University of Science and Technology of China, China

[#] Department of Computer Science and Engineering, University of California, Riverside, USA

[*] School of Computer Science and Technology, Nanjing Normal University, China

IEEE SECON 2016, London

1. Introduction
2. Traffic Characteristics
3. Performance Analysis
4. Implications on Real Systems
5. Conclusion

Introduction

- What is RAN Caching?
  - Content caching on the devices of RAN (Radio Access Networks), e.g., LTE e-NodeB

- What is done in this paper?
  - the first work that thoroughly analyzes the performance of RAN caching in LTE mobile networks using real traffic

Dataset Description

- Over a 7-day measurement period
- Collected from 3 e-NodeBs in a commercial LTE network of China Mobile
- More than 3200 access LTE users per day
- More than 60 Million Http sessions

Data Duration	2015.4.10 – 2015.4.16
Number of e-NodeBs	3
Average Users Number per Day	3249
HTTP Requests	62,104,921
Cacheable Http Requests	12,973,637
Total Download Traffic	1776 GB
Total Cacheable Download Traffic	1044 GB

Dataset Description

Field Meaning	Field Meaning
Timestamp	The time when a HTTP request is initiated
User ID	Anonymous user identifier
Content ID	The unique resource locator (URL) of one requested content
HTTP Method	A token obtained from HTTP request message, e.g., GET, HEAD, POST
Content Type	A value obtained from the content type field in HTTP reply message, including text, image, video, audio, application and others
Content Length	The length of requested content in bytes, obtained from HTTP reply message
Downlink Length	Actual length of downlink transmission data
HTTP Status Code	Indication of the implementation state of a specific HTTP session, e.g., 200, 203, 300

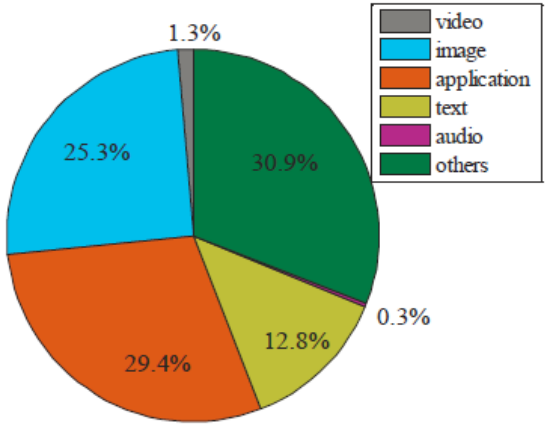
1. Introduction
2. Traffic Characteristics
3. Performance Analysis
4. Implications on Real Systems
5. Conclusion

Main Observations

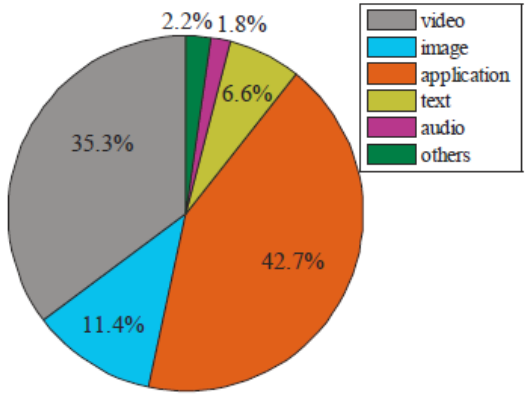
- **Observation 1**: 21.5% HTTP sessions are cacheable. However, the traffic volume of these cacheable HTTP sessions account for 56.2% of the total HTTP traffic

Main Observations

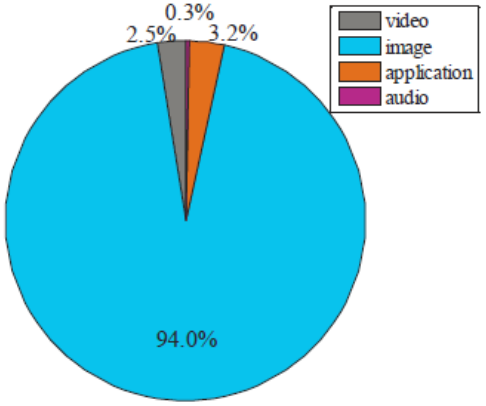
- **Observation 2**: Proportion of Requests and Traffic Volume by Content Type



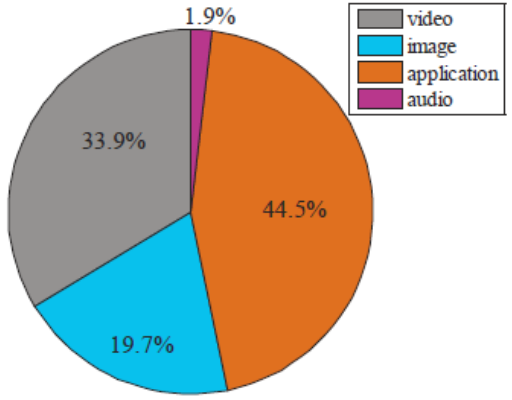
Proportion of All Http Requests



Proportion of All Traffic Volume



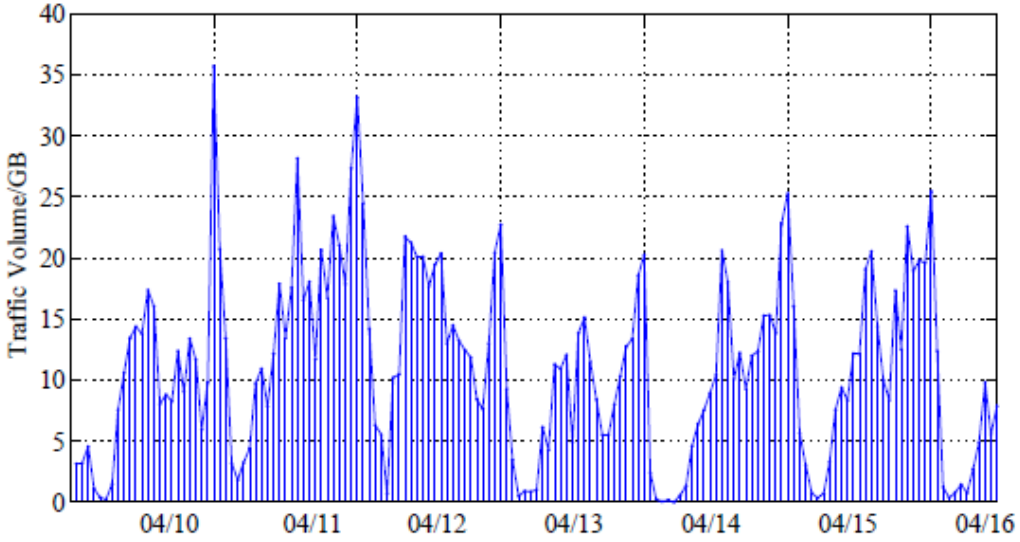
Proportion of Cacheable Http Requests



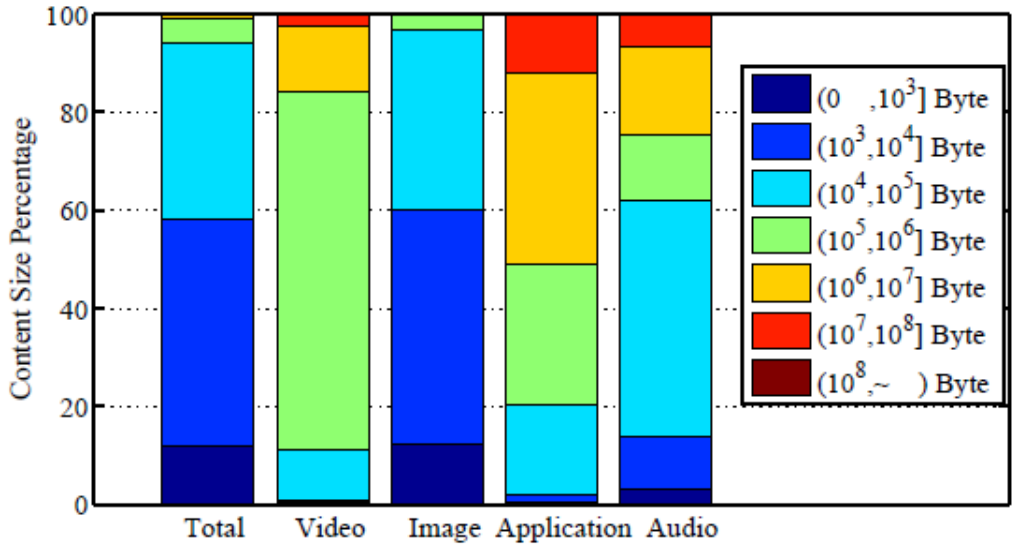
Proportion of Cacheable Traffic Volume

Main Observations

- **Observation 3**: Traffic Volume and Content Size



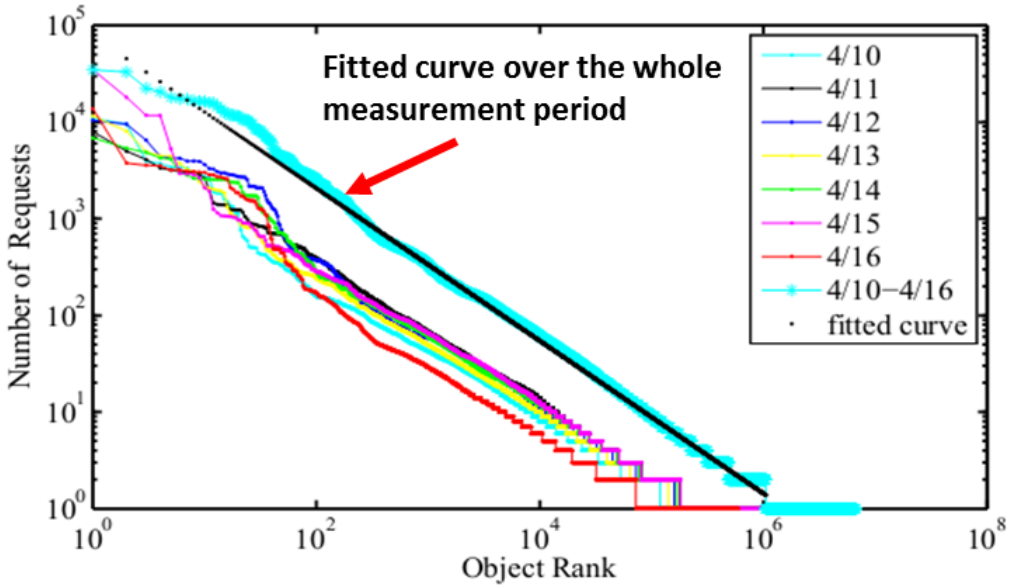
Traffic volume for one week



Cacheable content size distribution

Main Observations

- **Observation 3**: the popularity of cacheable contents follows a Zipf distribution with parameter α equal to 0.8, with 3200 average access users per day.



Cacheable content size distribution

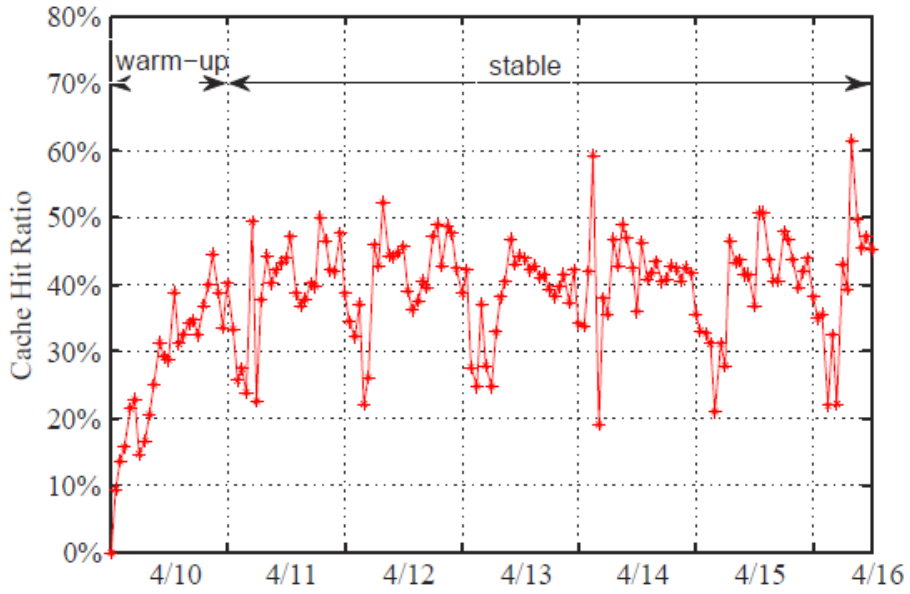
Data	α	C	r
2015/04/10	0.7584	3.8551	0.9747
2015/04/11	0.7425	4.0530	0.9974
2015/04/12	0.8403	4.2719	0.9728
2015/04/13	0.7556	3.9325	0.9892
2015/04/14	0.7885	4.1235	0.9819
2015/04/15	0.7356	3.9844	0.9847
2015/04/16	0.9246	4.1816	0.9568
2015/04/10-16	0.8077	4.9569	0.9925

the Obtained Parameters of Zipf Distribution

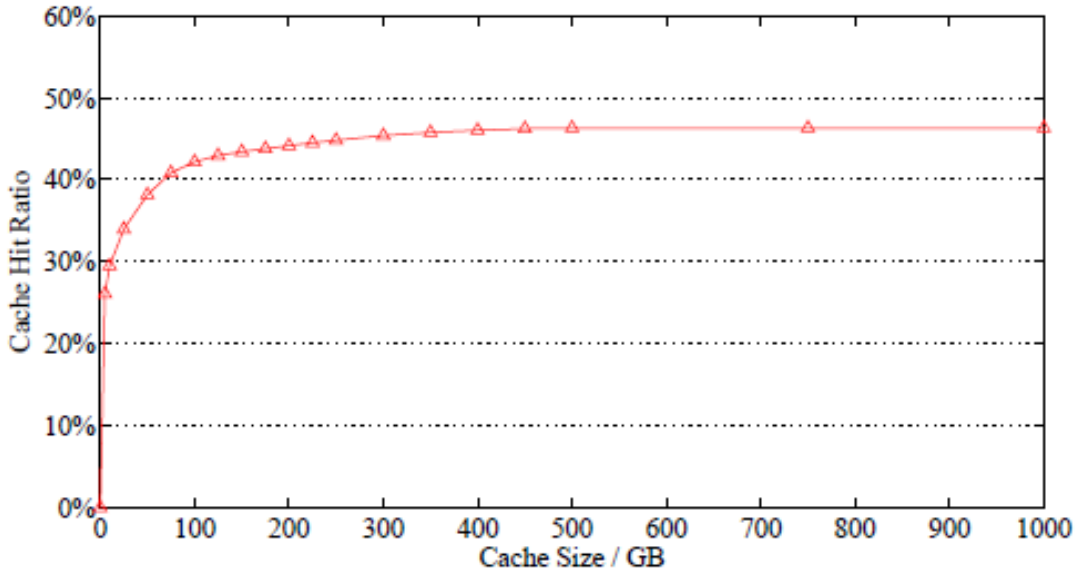
1. Introduction
2. Traffic Characteristics
3. Performance Analysis
4. Implications on Real Systems
5. Conclusion

Cache hit ratio based on the total three eNodeBs

- *Effect of Cache size on Hit Ratio*
  - With cache size increasing from 5 GB to 1000 GB, the cache hit ratio rises correspondingly and converges to a maximum hit ratio of 46.3%.
  - The growth of hit ratio flattens out when the cache size reaches 100 GB, and the corresponding hit ratio is 42.2%, which is close to the maximum value.



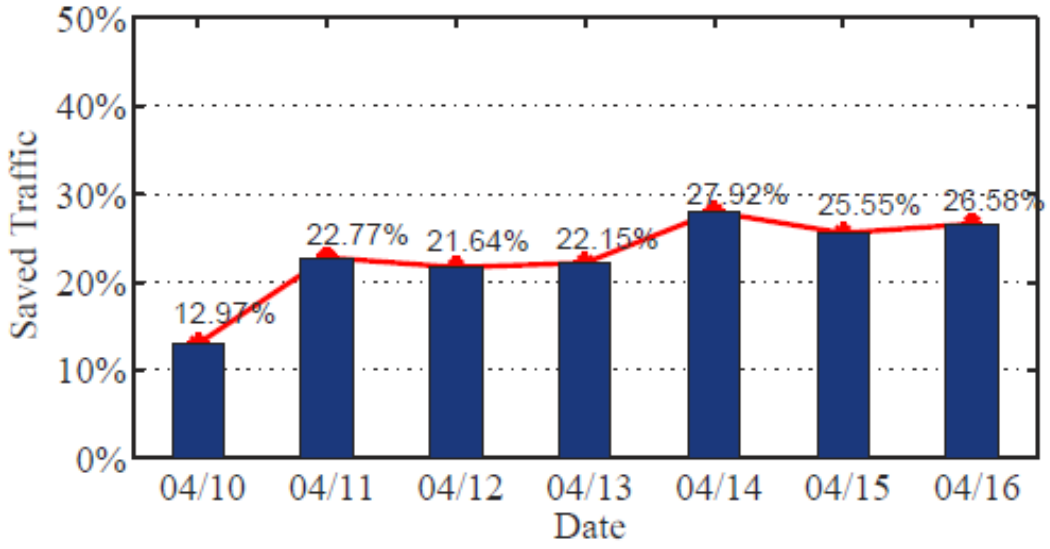
Cache Hit Ratio of the total three eNodeBs



Hit ratio over different cache size

Saved Traffic based on the total three eNodeBs

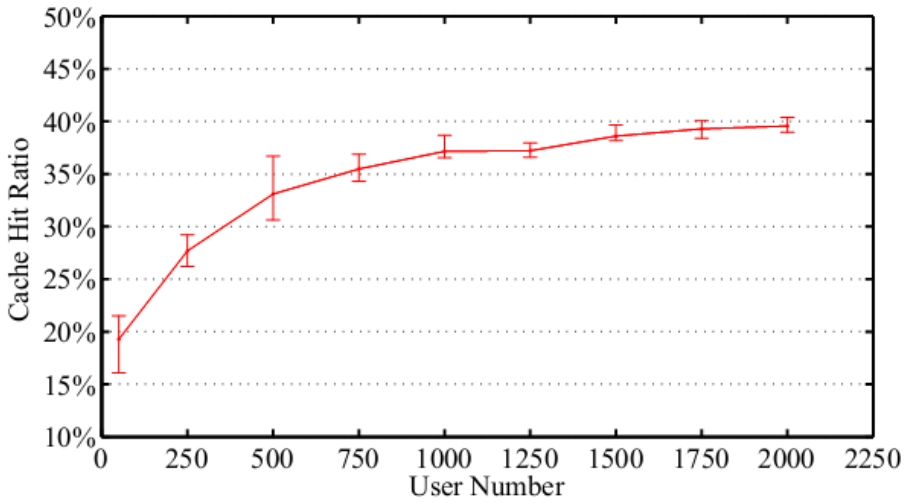
- *Saved Traffic based on the total three eNodeBs*
  - With 100 GB cache size, RAN caching can save as much as 24.4% of the total HTTP downstream traffic (including both cacheable and un-cacheable traffic) on average, with maximum savings of 27.9%



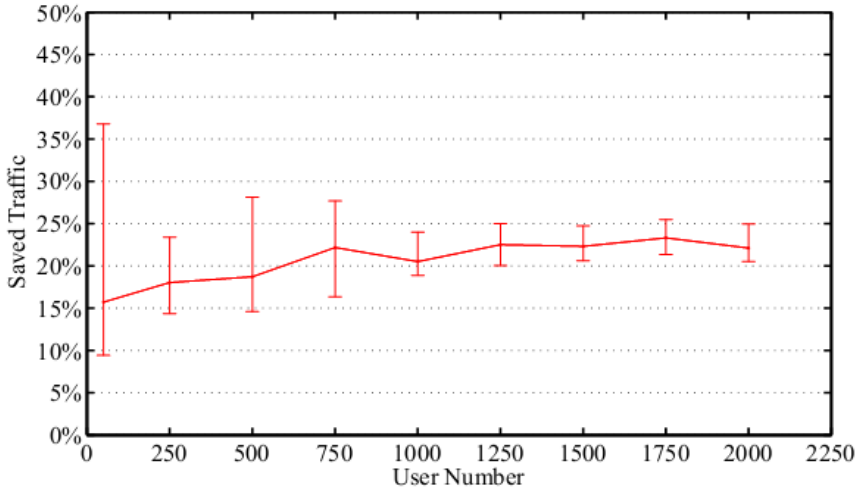
Saved traffic in different days

Effect of User Number

- *Effect of User Number on Hit Ratio*
  - We investigate the effect of user number ranging from 50 to 2000, on the cache hit ratio. Not surprisingly, with increasing user number, the cache hit ratio correspondingly improves from an average value of 19.1% to 39.7%



- *Effect of User Number on Traffic Reduction*
  - The effect of user number on the saved traffic is also investigated. When the user number ranges from 50 to 2000, the volume of total HTTP downstream traffic is reduced by 15.9% to 23.2% on average.



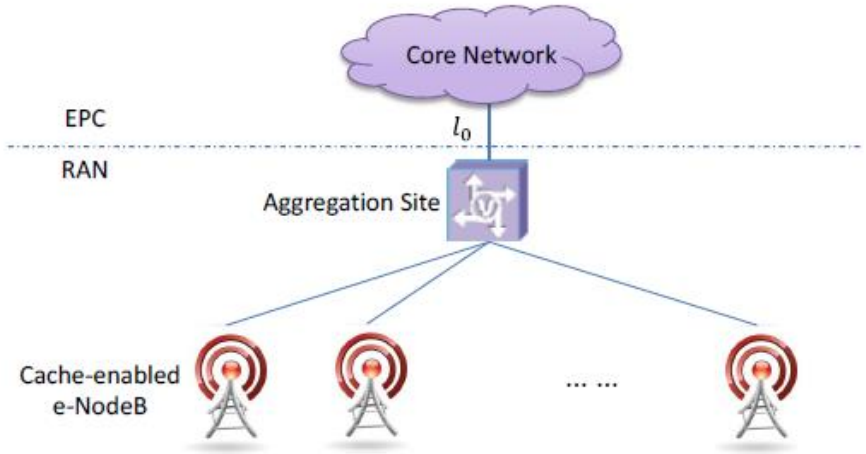
1. Introduction
2. Traffic Characteristics
3. Performance Analysis
4. Implications on Real Systems
5. Conclusion

On Fundamental Design Issues

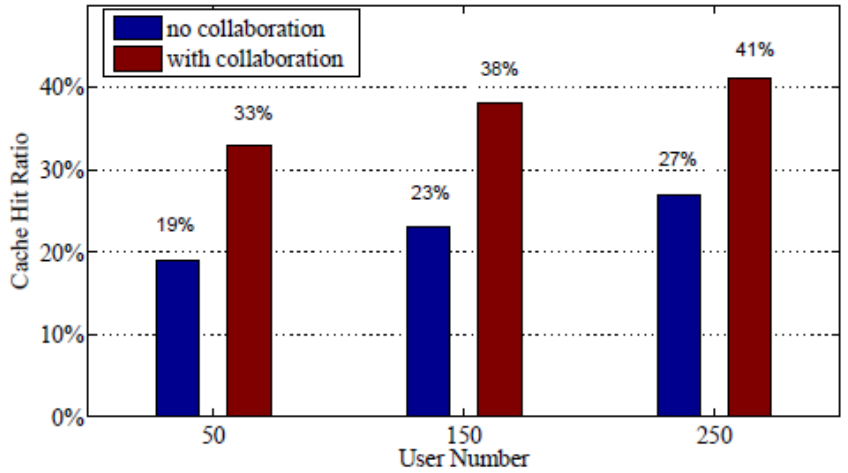
- What type of content should be stored in the RAN cache?
  - In addition to videos and applications, the content type of image should not be overlooked, especially considering the vast majority of request proportion (94%) and a remarkable hit ratio (41.9%).
- How large the storage size of the RAN cache should be?
  - a cache size of several hundreds of GB is enough for a single eNodeB of medium traffic load, and provides a good tradeoff between the performance and the cost of the RAN cache.
- Which e-NodeB is suitable for deploying RAN caches?
  - the access user number of eNodeB is an important factor when deciding which eNodeBs are good candidates for cache

On the Collaborative Caching Issue

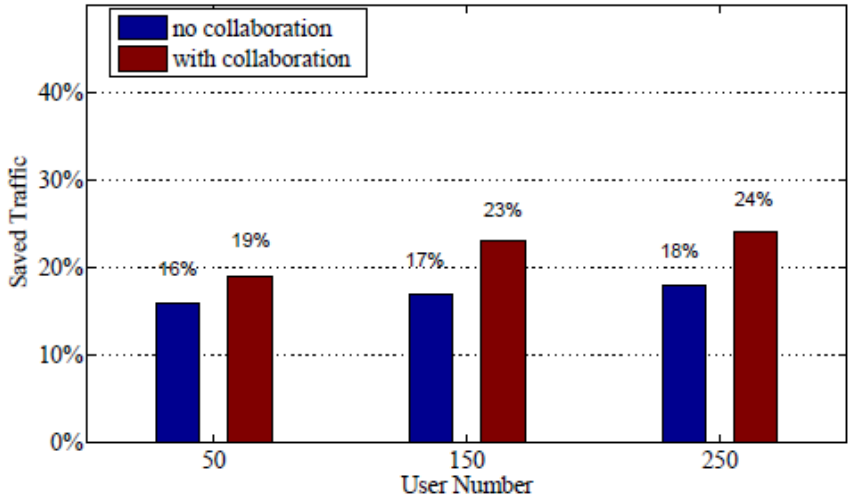
- What is the potential gain of collaborative RAN caching?
  - Experiment setup (10 eNodeBs, the user number of each eNodeB set to be 50, 150, 250)
  - collaborative caching helps improve the performance of a caching system
  - benefits vs. cost



Example of collaborative RAN caching



Comparison of cache hit ratio over different user number



Comparison of saved traffic over different user number

1. Introduction
2. Traffic Characteristics
3. Performance Analysis
4. Implications on Real Systems
5. Conclusion

Conclusion

- the first work that thoroughly analyzes the performance of RAN caching in LTE mobile networks using real traffic
- the popularity of cacheable contents approaches to a Zipf distribution with parameter α equal to 0.8, with 3200 average access users per day
- provide evidence of the potential benefits of deploying RAN caches
- the hit ratio of RAN caching reaches 42.2% on average and that downstream Http traffic can be reduced by 24.4%, with 100 GB cache size and around 3200 mobile users
- present the implications on some fundamental design issues of practical RAN caching systems

Thank You!

Tao Lin
lintao@iie.ac.cn