Edge Caching

Living on the edge: The Role of Proactive Caching in 5G Wireless Networks

Cache in the Air: Exploiting Content Caching and Delivery Techniques for 5G systems

Approximation Algorithms for Mobile Data Caching in Small Cell networks

We have seen a rise in wireless services and traffic over the past decade,with a projection of 500 fold increase over the next decade. Clearly, more advanced techniques are needed to increase the network capacity.

Edge Caching : A potential solution.

Small Cell Networks (SCN)

A way to meet unprecedented traffic demands is to deploy SCNs, which provide short-range, low-power, low-cost small base stations.

But they are not able to solve peak traffic demands using the existing *reactive* networking paradigm:

users trafic requests must be served urgently as they come or dropped causing outages.

Also, a large scale deployment of SCNs is not viable due to site acquisition, installation and *backhaul* costs.

The Role of Proactive Caching in 5G Wireless Networks

The proposed networking paradigm is proactive in essence.

It involves *network nodes* exploiting users' context information, anticipate users' demands and use predictive abilities to save resources and cache potential future requests.
Intelligent pre-allocation of resources and effective scheduling of user requests.

Propose to utilize network nodes' ability to anticipate and predict demands, Device-to-Device communication, aspects of context-awareness and Social Networks.

Leveraging device capabilities and Social Networks

Leveraging the powerful processing capabilities and large memory storage of devices enables network operators to proactively serve predictable peak-hour requests during off-peak times.

Need an optimal tradeoff between predictions that result in content being retrieved but not needed, and content needed but not anticipated timely.

Key challenge: Data sparsity

Possible solution: Social relationships and networks.

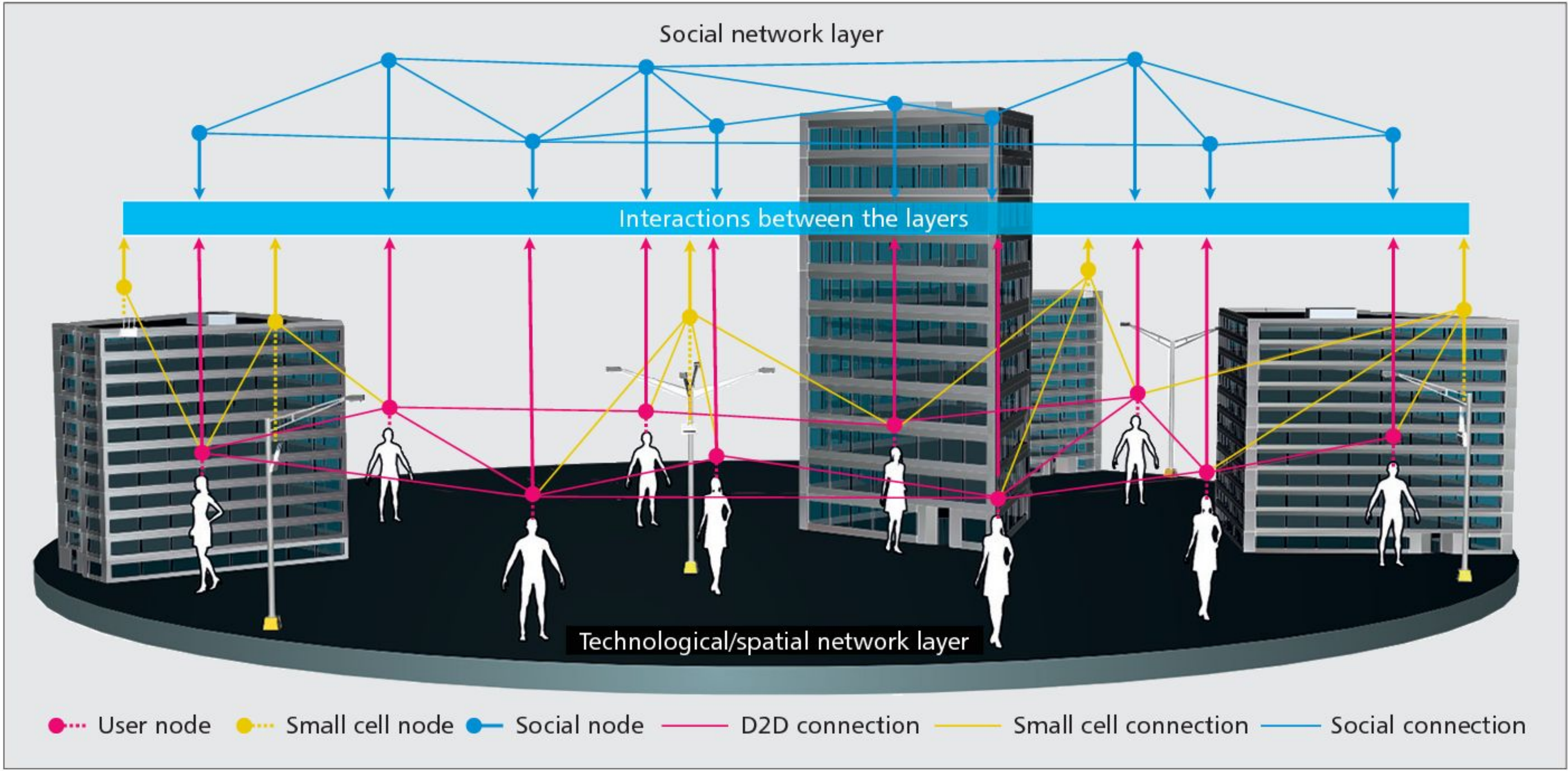


Figure 1. An illustration of an overlay of socially interconnected and technological/spatial network.

Using Proactive SCNs

Store files based on their highest popularity, until the storage capacity is achieved.

Therein, SBSs have information of the popularity matrix : *large, sparse, partially unknown.*

In order to train a predictive model, a Least Square Minimization problem was solved.

Parameters used :
Satisfied Requests, Backhaul load, Vs Cache Size and Number of requests.

Social Networks Aware Caching via D2D

By exploiting the interplay between social and technological networking, each SBS tracks and learns the set of influential users using the social graph.

Social community formation: Eigenvector centrality.
Eigenvectors and eigenvalues of the adjacency matrix of a “*social graph*”: weighted, tells us the communication probability between 2 users..

The largest Eigenvalue is taken as a parameter ‘K’ as the *K most influential users* for the set. The communities are formed via K-means clustering.

After knowing the influential users and their communities, the next step is to determine the content dissemination process inside each community.

This is modeled as a *Chinese restaurant process (CRP)*.

Using this, a *Popularity matrix* is constructed, and the popular files in each community can be stored *greedily* in the cache (D2D) of influential users.

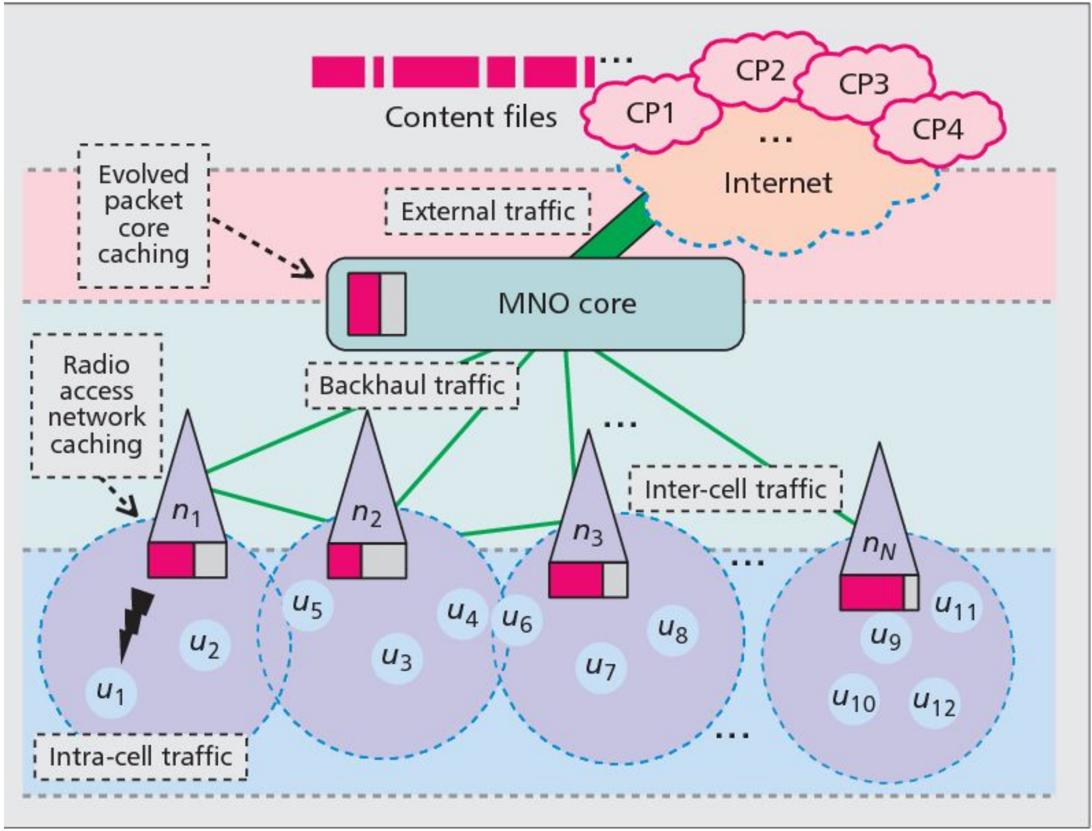
Cache in the Air: Exploiting Content Caching and Delivery Techniques for 5G systems

Proposes edge caching scheme based on the concept of content-centric networking or information-centric networking.

Gives possible answers to questions:
Where and How to Cache?

RAN: Radio Access Network

EPC: Evolved packet core.



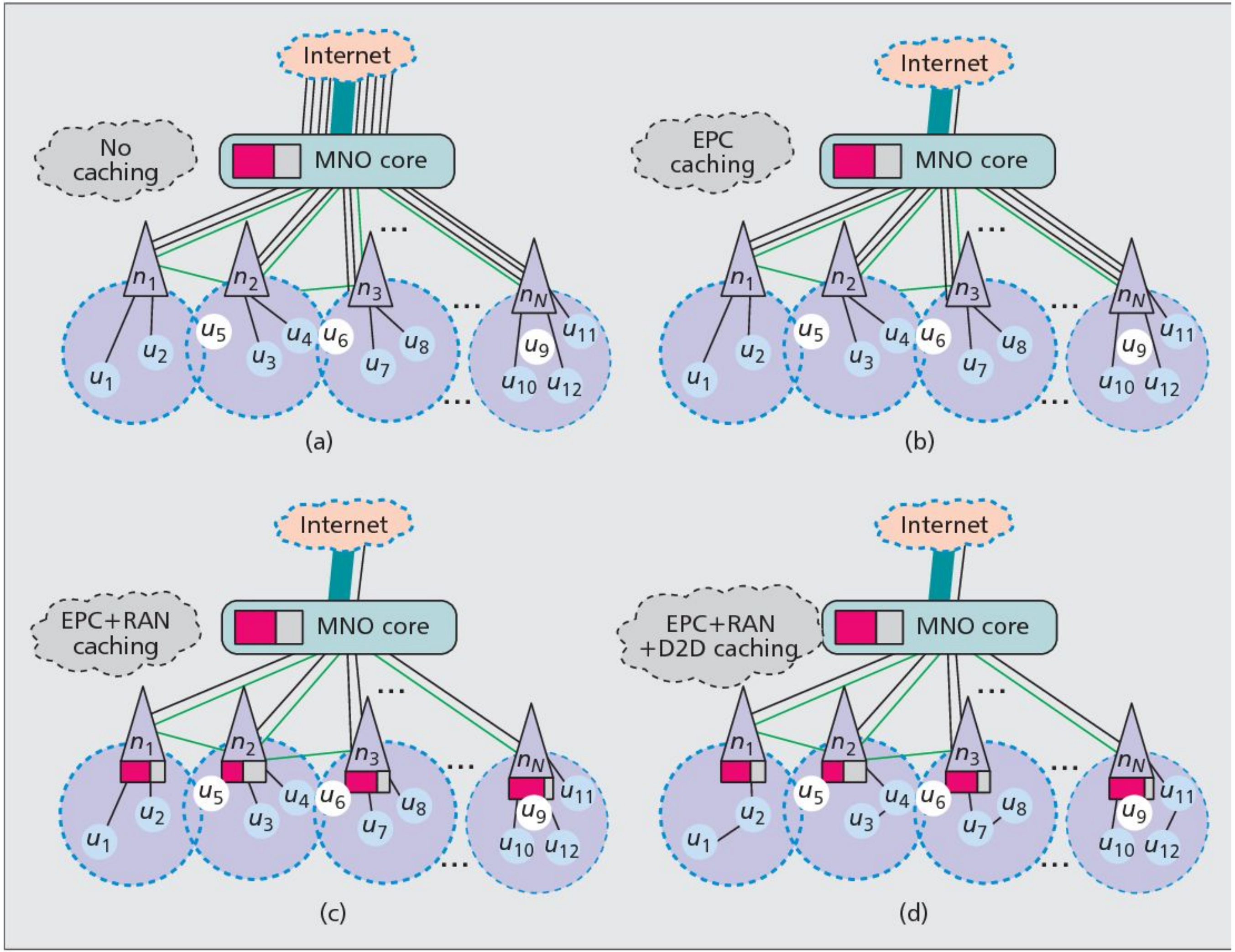


Figure 2. *Comparison of a) no caching; b) EPC caching; c) EPC+RAN caching; d) EPC+RAN+D2D*

Content Centric Networking(CCN) Based Caching

The primary goal of CCN is to facilitate in-network data storage for universal caching in every network node.

Major CCN designs have the following common attributes:

- Receiver-oriented and chunk-based transport.
- In-network per-chunk caching
- Name-based forwarding
- Uniquely identifiable content naming

Due to the wide distribution of caching resources, cooperative caching policies should carefully consider content popularity, freshness, diversity, and replica locations over the network topology in order to achieve the following objectives:

- Minimization of *Inter*-ISP traffic
- Minimization of *Intra*-ISP traffic
- Minimization of content access delay for users

Open Issues and Challenges as addressed by this paper

Distributed cache resource management and cooperative caching policy.

Caching with Multicast and Device-to-Device (D2D).

Integration with Virtualization of Cellular Networking Devices.

Approximation Algorithms for Mobile Data Caching in Small Cell networks

Local caching of popular content items at the Small Cell Base stations (SBSs) has been proposed in order to decrease the costly transmissions from the macrocell base stations without requiring high capacity backhaul links for connecting the SBSs with the core network.

However, the caching policy design is a challenging problem especially if one considers realistic parameters such as the bandwidth capacity constraints of the SBSs that can be reached in congested urban areas.

Prior works assume that the transmission capacity is rarely the bottleneck for the caching base stations.

Joint Routing and Caching for Unsplittable requests (JRC-UR)

NP - HARD : further compounded by additional bandwidth constraints of SBSs.

User requests are considered unsplittable.

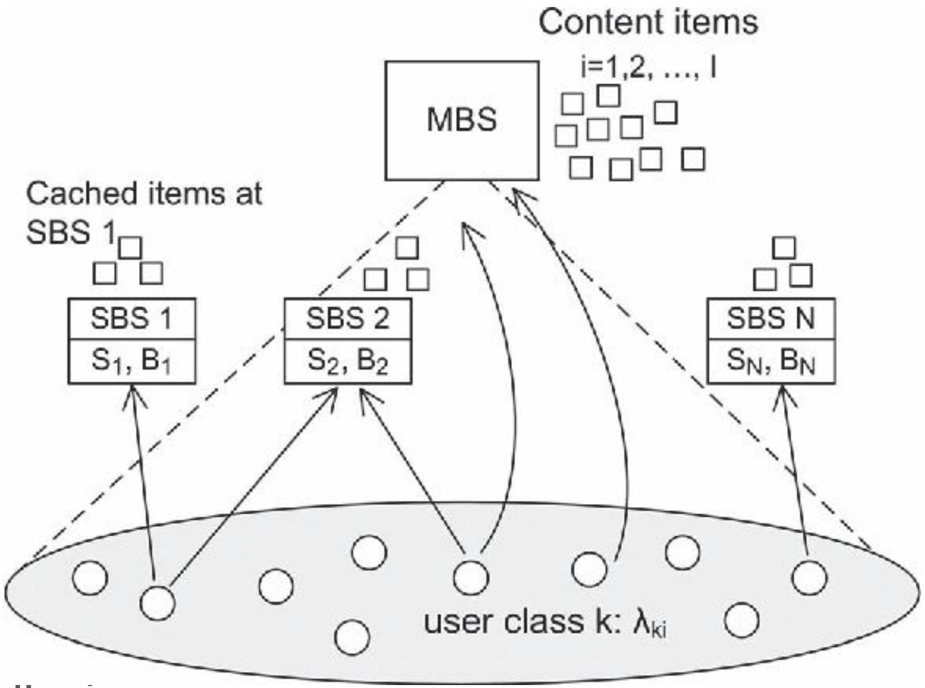
- Proposed a novel mapping of this problem to a a variant of the *facility location problem* known as the Unsplittable Hard-Capacitated Metric Facility Location Problem (UHCMFL).
- to the best of the authors knowledge, this is the first work that provides a solution with provable approximation guarantees for the joint content placement and request routing problem in small cell networks.

System Model

User Class: κ: {1,2 … ,K}.

SBSs which operate in conjunction with a macrocell: N
Transmission capacity: B bytes in a period T

I : a static collection of content items (files), all of size ‘s’.



The requests that cannot be satisfied locally, i.e., by any SBS, are routed to the MBS.

The optimal routing and caching strategy is the one that maximizes the requests satisfied by all the SBSs combined. I.e, minimize the requests routed to MBS.

But with constraints:

- Storage limits of SBS
- Transmission capacity
- SBS cannot serve requests located outside their coverage areas
- All the requests generated must be served

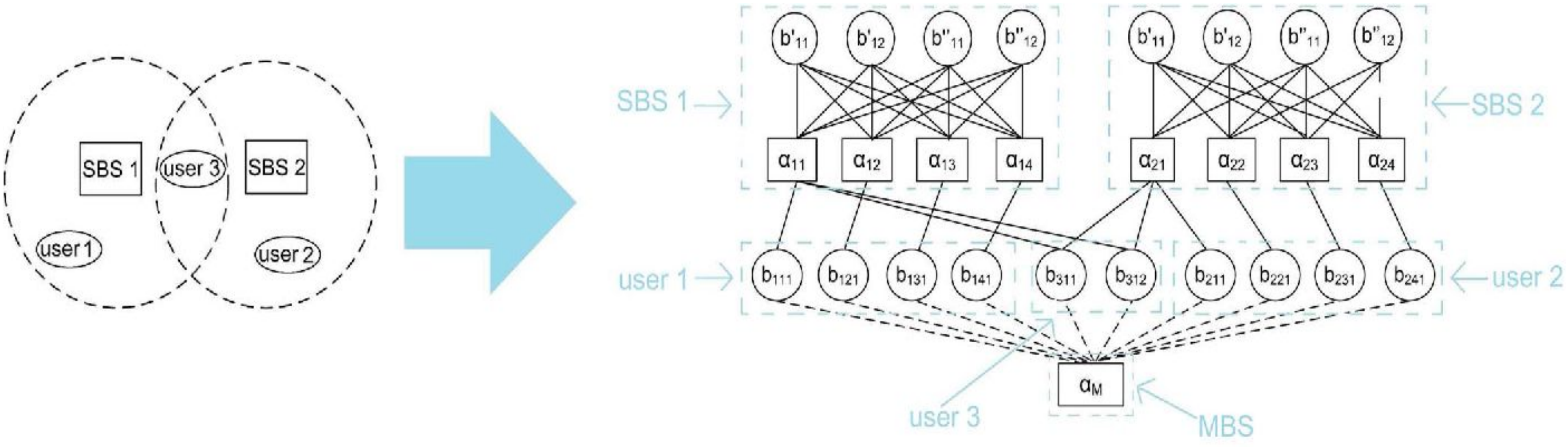
The Facility Location problem

We have a set of facilities A, and a set of clients B, (overlapping).
Every client has a different demand, and each facility has an opening cost, a servicing cost (per unit demand) and a Capacity.
Unsplittable: The entire demand must be assigned to a single open facility.
Capacity: limits the total sum of demands served by a facility.
The costs form a metric

Problem: to determine the facilities A* needed and the clients B* each one needs to serve so as to *minimize the aggregate opening and servicing costs.*

The JRC-UR problem is reduced to the facility location problem, in *polynomial time*.

This reduction is an important step, and a not-so intuitive one.



JRC-UR instance

UHCMFL instance

Now, using the literature available for the facility location problem, pre-existing Approximate Algorithms can be modified and used to solve the JRC-UR problem

The solution is not the optimal one, but an approximate one, with the approximation ratio dependent on :

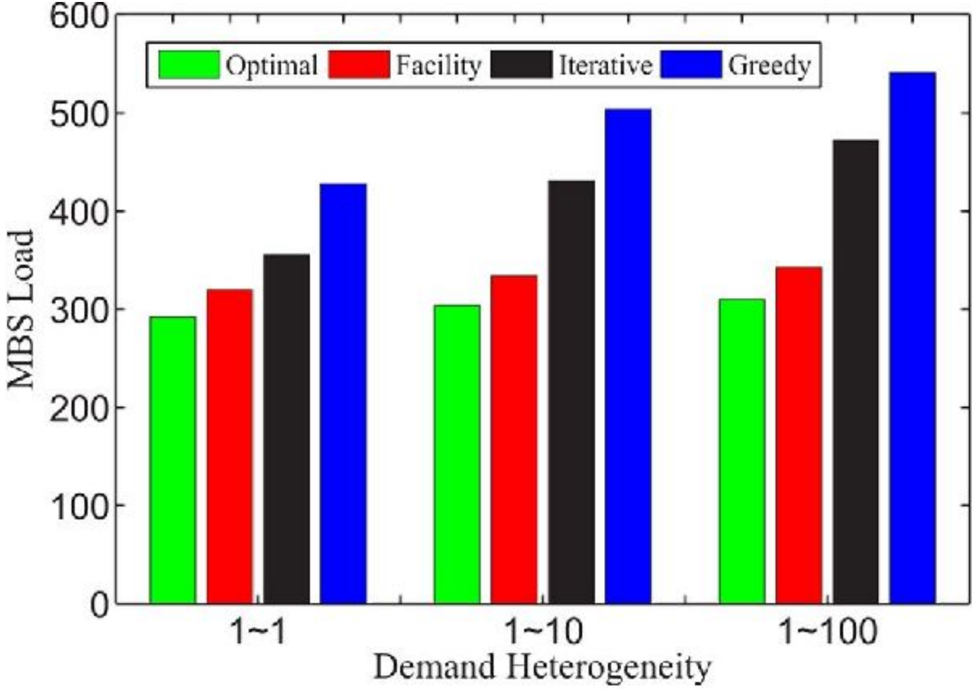
- cache size
- Transmission capacity
- File request pattern
- Demand heterogeneity

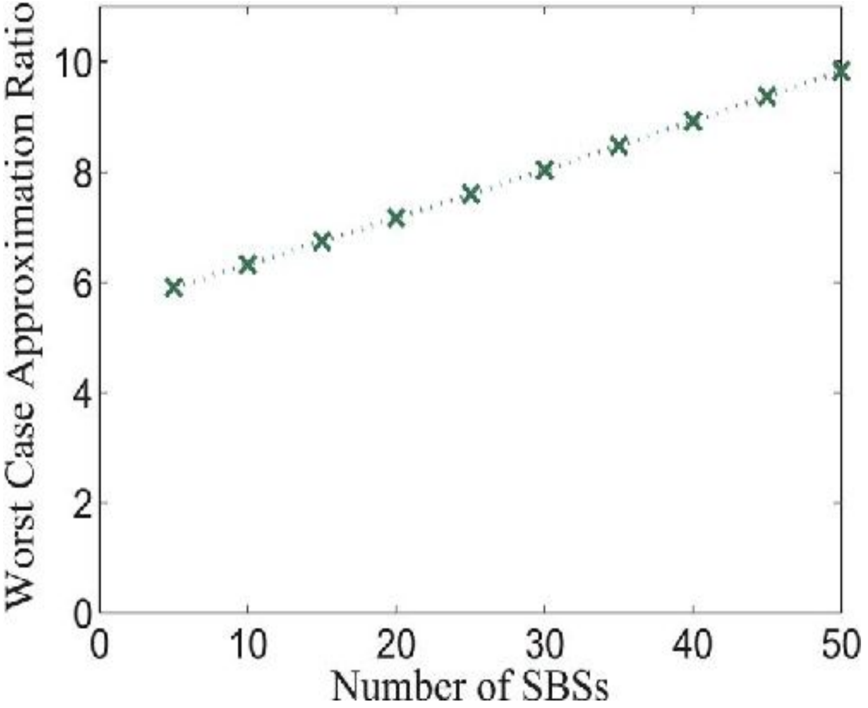
Performance comparison

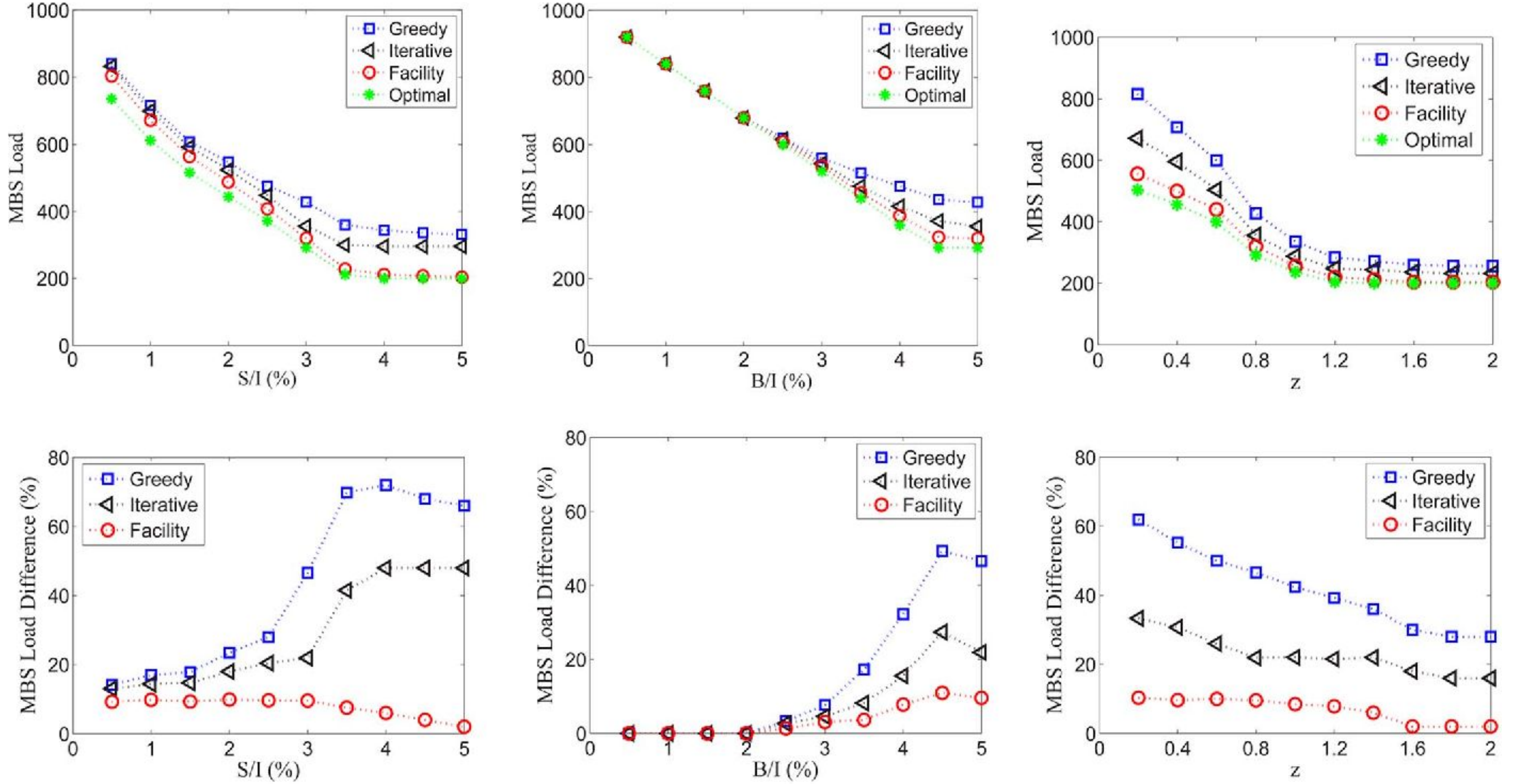
The Approximate Algorithm was compared to *Greedy, Iterative and Optimal* solutions.

It performed 31% better than the Greedy solution, 17% better than the Iterative Solution, and 10% worse than the optimal solution, in the worse case scenario

The result improves with increase in Cache size and transmission capacity


Optimal
Facility
Iterative
Greedy
MBS Load
600
500
400
300
200
100
0
1~1
1~10
1~100
Demand Heterogeneity


Worst Case Approximation Ratio
10
8
6
4
2
0
0
10
20
30
40
50
Number of SBSs


MBS Load
S/I (%)
B/I (%)
z
Greedy
Iterative
Facility
Optimal
MBS Load Difference (%)
0
200
400
600
800
1000
0.4
0.8
1.2
1.6
2
1
3
4
5
20
40
60
80

(a) (b) (c)