

Efficient Data Access in Disruption Tolerant Networks Using Cooperative Caching With Drama

Dinesh.E, Akshaya.S, Ashini Meena.P,Bhuvaneswari.S, Deepak.A

Abstract—Disruption Tolerant Networks (DTNs) is characterized by low node density and unpredictable node mobility. The current research efforts in DTN focus on data forwarding and reduce absence of end-to-end path between source and destination, but only limited work has been done on providing efficient data accesses to mobile users. In this paper, we propose a greatly improved energy efficiency strategy named DRAMA aimed to improve storing and fast accessing of data in cache and also it supports cooperative caching in DTNs, which makes the sharing and coordination of cached data among multiple nodes and reduces data access delay. Our idea is to cache data at a set of nodes as network central locations (NCLs), which can be easily accessed by other nodes in the network. We propose an efficient scheme to select appropriate NCL based on probabilistic selection metric and coordinates multiple caching nodes to optimize the trade-off between data accessibility and caching overhead.

Keywords— Cooperative caching, DTN, Network central Location, Cache replacement, DRAMA.

I. INTRODUCTION

Disruption Tolerant Network is one of the mobile ad-hoc networks that changes its topology frequently. Only intermittent connectivity exists in DTN and the difficulty of maintaining end to end communication path makes it necessary to use “store-and-forward” methods for data transfer. In DTN all nodes does not need large storage capacity to store and forward the data in the network. Examples of such networks are military fields, disaster recovery area, extreme terrestrial environments and VANET. The problem is how to determine the appropriate relay selection strategy. Although forwarding schemes have been proposed in DTNs^[1], there is limited research on providing efficient data access to mobile users, despite the importance of data accessibility in many mobile applications. The mobile users can only request the data whenever needed and they do not know data locations in advance. The destination of data is, hence, unknown when data are generated.

Appropriate network design is needed to ensure that data can be promptly accessed by requesters in such cases.

Dinesh.E is Assistant Professor, Department of Computer Science and Engineering, Narasu's Sarathy Institute of Technology, Salem-636 305 (email : dineshelayaperumal@gmail.com)

Akshaya.S, Ashini Meena.P,Bhuvaneswari.S, Deepak.A are UG scholars B.E – Computer Science and Engineering , Narasu's Sarathy Institute of Technology, Salem-636 305 (email : akshayaselvareaj08@gmail.com, ashinimeena@gmail.com, sbhuvaneswari1993@gmail.com, deepaknsit21@gmail.com)

A general technique used to improve the performance of data access is caching i.e., to cache data at appropriate network central locations (NCLs) based on query history, so that queries in the future can be responded with less delay. Although cooperative caching has been used for both web-based application and wireless ad-hoc networks, to allow sharing and coordination among multiple caching nodes, it is difficult to be realized in DTNs due to lack of continuous network connectivity. First, the opportunistic network connectivity complicates the data transmission delay, and furthermore makes it difficult to determine appropriate caching locations for reducing data access delay. This difficulty is raised by the incomplete information at individual nodes about query history. Second, due to the uncertainty of data transmission, multiple data copies are needed to be cached at different locations to ensure data accessibility. The difficulty in coordinating multiple caching nodes makes it hard to optimize the tradeoff between data accessibility and caching overhead.

In this paper, we propose a scheme to address the difficulty and to increase the efficiency of energy consumption and data accessibility using DRAMA architecture, to support cooperative caching in DTN. Our basic idea is that a 3D stack coarse-grain reconfigurable accelerators (CGRAs) atop off-chip DRAM devices, which is used to improve and speed up the storing, reduce data transfer across conventional processor memory hierarchy (i.e., from off-chip DRAM to on-chip cache, register files) which makes data access from memory faster and thus reduces energy consumption.

- Its main aim is to intentionally cache data at set of network central locations (NCLs), so that other nodes in the network can be access it easily. Each NCL in the network has high popularity, can be represented as central nodes and is prioritized for caching data. Due to the limited caching buffer of central nodes, multiple nodes near a central node may be involved for caching. We ensure that popular data are always cached nearer to the central nodes via dynamic cache replacement based on query history. We use an efficient DRAMA architecture to reduce the energy consumption to transfer data across conventional memory hierarchy up to 65-95% while achieving speedups of up to 18 x commodity processor.

- We develop an approach to select appropriate NCL in DTN based on probabilistic selection metric. The selected NCLs achieve high chance to respond user queries with low

overhead and high data access speed in network storage and transmission.

- We propose a data access scheme to coordinate multiple caching nodes for responding to user queries and optimize the tradeoff between data accessibility and caching overhead, to minimize the average number of cached data copies in the network.
- We propose a utility-based cache replacement technique to dynamically adjust cache location (i.e., NCL) based on query history.

The rest of the paper is organized as follows: In section 2 brief descriptions about the existing work. Section 3 provides an overview of intentional caching in DTN. Section 4 describes about appropriate NCL selection in DTN. Section 5 describes about the proposed DRAMA architecture, and section 6 proposes load balancing technique among NCLs.

II. RELATED WORK

Research on data forwarding in DTNs originates from Epidemic routing^[4], which floods the entire network. Later focuses on proposing efficient relay selection metrics to approach the performance of Epidemic routing with lower forwarding cost, based on prediction of node contacts in the future. Some schemes do such prediction based on their mobility patterns, which are characterized by Kalman filter or semi-Markov chains^[2]. The aforementioned metrics for relay selection can be applied to various forwarding strategies, which differ in the number of data copies created in the network. While the most conservative strategy always keeps a single data copy and Spray-and-Wait^[3] holds a fixed number of data copies, most schemes dynamically determine the number of data copies. In Compare-and-Forward, a relay forwards data to another node whose metric value is higher than itself. Delegation forwarding reduces forwarding cost by only forwarding data to nodes with the highest metric. In other schemes^[4], without brokers, data items are grouped into predefined channels, and are disseminated based on users' subscriptions to these channels.

Caching is another way to provide data access. Cooperative caching in wireless ad hoc networks, in which each node caches pass-by data based on data popularity, so that queries in the future can be responded with less delay. Caching locations are selected incidentally among all the network nodes. Some research efforts^{[5], [9]} have been made for caching in DTNs, but they only improve data accessibility from infrastructure network such as Wi-Fi access points (APs) or Internet. Data are intentionally cached at appropriate network locations with generic data and query models, but these caching locations are determined based on global network knowledge. In this paper, we support cooperative caching in a fully distributed manner in DTNs with heterogeneous node contact patterns and behaviors.

III. OVERVIEW

A. DRAMA Architecture

The DRAMA architecture stacks CGRA on top of DRAM devices, which is connected to internal DRAM I/O through TSVs (Through silicon vias). Each CGRA is connected to its DRAM devices and operate on data that contained in that device, independently of CGRAs on the other DRAM devices. DRAMA can be easily used in existing processors to accelerate DRAMA enhanced applications. The processor communicates with CGRA through a memory mapped I/O interface that operates similarly to mode registers in conventional DRAM systems.

The CGRA layer reads or writes data through TSV's connected to the existing GIO buses without changing the architecture of DRAM device. The device has eight banks divided into two groups of four banks. Each bank group shares 128 bit inter-bank global I/O (GIO) bus comprised of upper and lower 64-bits.

TSV transfers command and address bits from CGRA layer to DRAM device to indicate the type of operation involved. The processor and CGRA do not operate on same data sets. When CGRAs are working, the processor busy-waits until CGRA computations are completed to avoid frequent concurrent accesses.

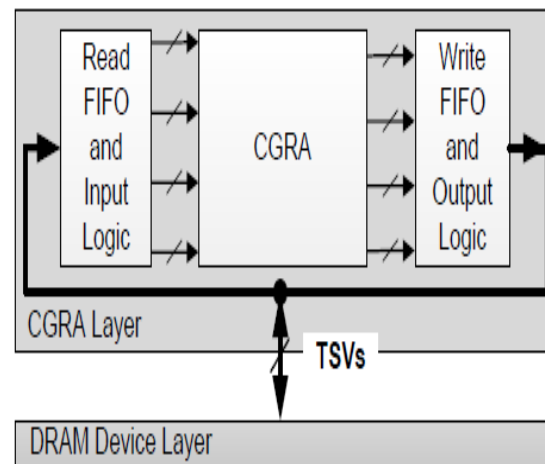


Fig 1: DRAMA CGRA stacked on top of DRAM Device

Thus, no changes to memory consistency are needed. If the processor side MC (memory controller) requires sending memory commands, it first need to halt CGRA side MCs by writing to the mode registers. CGRA side MCs then close pages before the processor side MCs takes control back. Next the processor side MC activates the required page and access data.

B. Network Model

The nodes in the DTN are represented by contact graph $G(V, E)$. The graph is undirected and the node contacts are symmetric i.e., node i contacts j , whenever j contacts i . The edges between the nodes are determined by inter contact time.

The basic idea of this paper is to cache data at a set of Network Central Locations (NCL), which is easily accessed

by all other nodes in the network. Due to limited memory in the cache, multiple nodes nearer to the NCL act as a cache. When the current NCL is full, then the nearby node can be selected as a new NCL in the network which act as caching node. NCLs are prioritized for caching data so that the data which has highest popularity is cached nearer to the central node. The data source S pushes the generated data towards the NCL and the cache C1 and C2 of NCL are prioritized for storing the data

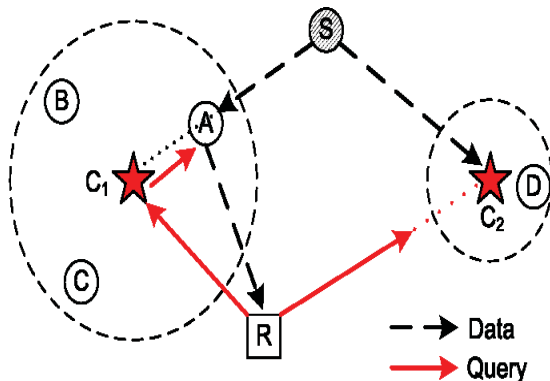


Fig 2: The picture of intentional caching

If C1 is full, data are cached at another node near C1. A requester R pulls data by querying NCL and data copies from multiple NCLs are returned to ensure data access. The central node closer to the requester, its queries are responded sooner by the NCL.

C. NCL Selection Metric

The multi-nodes connection in the network is represented by the contact graph $G = (V_p, E_p)$. The path P between the node A and B is defined as $P_{AB} = (V, E)$ where $V_p = \{A, N1, N2, \dots, B\}$ and edge $E_p = \{E1, E2, \dots, Er\}$ with edge weight $\{\lambda1, \lambda2, \dots, \lambda_r\}$. $P_A(T)$ is the probability that the data are transmitted from A to B along the path P_{AB} within time T.

$$P_{AB}(T) = \int_0^T p y(x) dx = \sum_{i=0}^n c^{(i)}_k \cdot (1 - e^{-\lambda_k T})$$

The data transmission delay between two nodes A and B is indicated by random variable Y, which is measured by the weight of the shortest path between the two nodes.

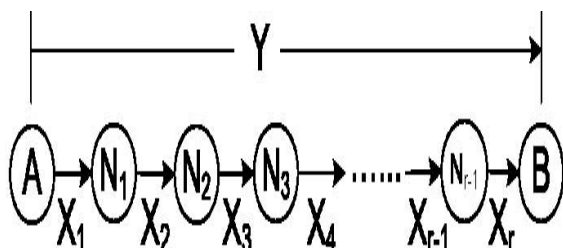


Fig 3: Opportunistic path

D. Caching Scheme

According to opportunistic path strategy each forwarding reduces the remaining delay for data to be delivered to the

central node. For newly generated node the initial caching locations are automatically determined during forwarding process based on buffer condition. The cache locations are adjusted by cache replacement.

The forwarding process only stops when caching buffer is full and data are cached at the current relay.

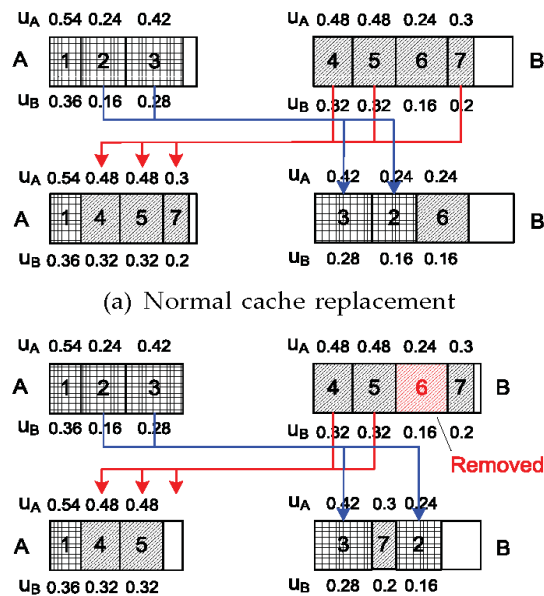
E. Cache Replacement

The cache replacement is based on data popularity. The cache replacement strategies such as LRU which removes least recent used data from the cache when new data are available are ineffective due to over-simplistic of data popularity. When data size is small and node buffer constraint is not tight, cache replacement will not be frequently completed. When data size becomes larger these strategies do not always select the most appropriate data to be cache.

We collect the cached data at both nodes into a selection pool $S = \{d_1, \dots, d_n\}$ and formulate cache replacement as

$$\text{Max} \sum_{i=1}^n x_i \cdot u_i + \sum_{j=1}^n y_j \cdot v_j$$

where $x_i, y_i \in [0,1]$ indicate whether data d_i are cached at node A and B after replacement, respectively. S_i indicates size of data d_i and S_A and S_B are buffer size of A and B. $u_i = w_i \cdot p_A$ and $v_i = w_i \cdot p_B$ indicates the utility of data d_i at A and B to cumulative cache performance, where w_i is the popularity of data d_i , p_A and p_B are the weights of shortest opportunistic path to the corresponding central node. Node A caches data d_1, d_2, d_3 and node B caches data d_1, d_5, d_6, d_7 . Since $P_A > P_B$, node A generally caches the popular data d_4, d_5, d_7 and leaves data d_2 and d_3 with lower popularity to node B. When size of caching buffer of node A and B decreases, A does not have enough buffer to cache data d_7 which is instead cache at node B.



(b) Possible data removal

Fig 4: Cache Replacement

Data d_6 with lower popularity will be removed from cache because neither node A nor B has enough space to cache it.

F. NCL Load Balancing

The central node plays a vital role in cooperative caching in DTNs. First the central node caches the most popular data in the network and respond to the frequent queries for these data. Second, the central nodes are responsible for broadcasting all the queries they receive to other caching nodes nearby. However, such functionality may quickly consume the local resources of central nodes that include their battery and their local memory.

G. Selecting Central Nodes

When a central nodes fails or its local resources are depleted, another node is selected as a new central node with highest NCL selection metric values among the current non central nodes in the network. When the local resources of central node C1 are depleted, its functionality is taken over by C3. Since C3 may be far away from C1,

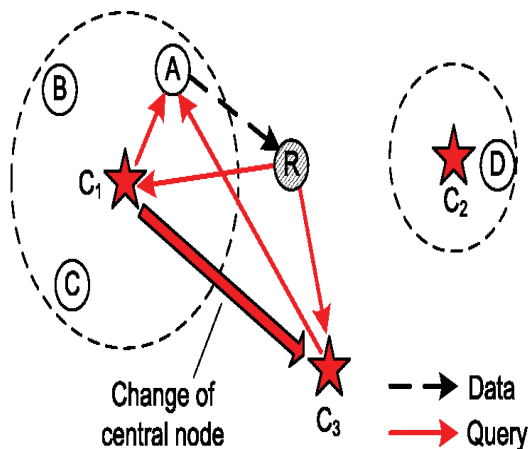


Fig 5: NCL load balancing

the queries broad casted form C3 may take a long time to reach the caching node A, and hence reduces the probability that the requester R receives data from A on time. The distance between the new central node and C1 should also be taken into account with respect to the original central node j, we define the metric C_i^j for the node i to be selected as the new central node as

$$C_i^j = C_i^i \cdot p_{ij}(T),$$

Where C_i^i is the original NCL selection metric and $p_{ij}(T)$ is the weight of shortest opportunistic path between the nodes i and j.

IV. SYSTEM MODEL

When the data is requested by the user then the source node sends data by initially processing the data and then selects the appropriate NCL as central node in the network to cache the data in it by NCL selection Mode. The NCL can cache the

data to be delivered faster when compared to the conventional memory storage. If the current NCL's buffer is completely full, then it is replaced with nearby high prioritized node as NCL by cache replacement technique. The cache replacement is based on popularity of the data in the network which can be identified by most user requested and received data.

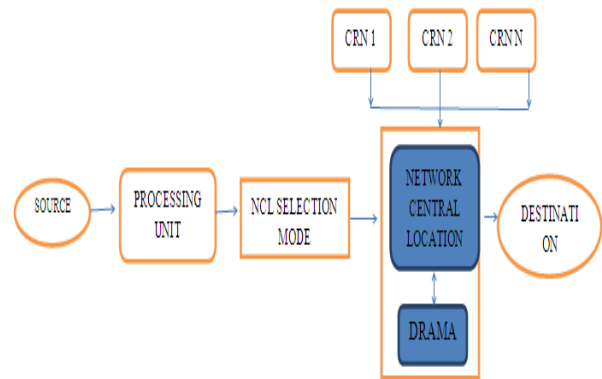


Fig 6: System architecture

Within the NCL low popular data is replaced to clear the storage space for further use. When the data has high popularity then the delivering of that data has been done faster rather than low popular data. Greedy dual size algorithm is used for cache replacement. After all these work has been completed data is delivered to destination without delay/loss with the help of NCL.

V. CONCLUSION

In this paper, we propose a novel scheme to reduce the data access delay and energy consumption by high performance computing using DRAMA architecture. In general, when the central node changes the existing caching location become inappropriate, and hence the successful ratio of data access is reduced up to 40 percent when the data life time is short, but will diminish significantly to 10 percent when there is a longer time for the queries to be forwarded to the caching nodes. And the caching overhead slightly increases by 10 percent. By using DRAMA architecture in central node the ratio of data access is reduced more when compared to the conventional technique i.e., up to 40 percent is reduced, thus increases the successful data access ratio.

REFERENCES

- [1] A. Balasubramaniam, B. Levine and A. Venkataramani, "DTN Routing as a Resource Allocation Problem", *proc, ACM SIGCOMM conf, Applications, Technologies, Architectures, and protocols for Computer Comm.*, pp. 373-384, 2007.
- [2] A. Basuet al., "Efficient Virtual Memory for Big memory servers," *in International Symposium on Computer Architecture, 2013, Vol. 41, no. 3*, pp. 237-248.

- [3] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age Matters: efficient Route Discovery in Mobile Ad Hoc Networks using Encounter Ages," *proc. ACM MobiHoc*, pp.257-266,2003.
- [4] A. vahat and D. Becker,"Epidemic Routing for Partially Connected Ad Hoc Networks", *Technical Report CS-200006,Duke Univ.,2000*.
- [5] M.J. Pitkanen and J. Ott, "Redundancy and Distributed Caching in Mobile DTNs,"*Proc. ACM/IEEE Second Workshop Mobility in theEvolving Internet Architecture (MobiArch)*, 2007.
- [6] Q. Yuan, I. Cardei, and J. Wu, "Predict and Relay: An Efficient Routing in Disruption-Tolerant Networks," *Proc. ACM MobiHoc*,pp. 95-104, 2009.
- [7] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," *Proc. ACM SIGCOMM Workshop Delay-TolerantNetworking*, pp. 252-259, 2005.
- [8] V. Lenders, G. Karlsson, and M. May, "Wireless Ad Hoc Podcasting," *Proc. IEEE Fourth Ann. Comm. Soc. Conf. Sensor, Meshand Ad Hoc Comm. and Networks (SECON)*, pp. 273-283, 2007.
- [9] Y. Huang, Y. Gao, K. Nahrstedt, and W. He, "Optimizing File Retrieval in Delay Tolerant Content Distribution Community," *Proc. IEEE Int'l Conf. Distributed Computing Systems (ICDCS)*,pp. 308-316, 2009