

Resolving Packet Loss in a Computer Centre Applications

M. Rajalakshmi

Assistant Professor

P. A. College of Engineering
and Technology
Pollachi, Tamil Nadu

C. Angel

Assistant Professor

P. A. College of Engineering
and Technology
Pollachi, Tamil Nadu

K. M. Brindha Shree

Assistant Professor

P. A. College of Engineering
and Technology
Pollachi, Tamil Nadu

ABSTRACT

The modern data centre consists of tens of thousands of servers to interconnect distributed computations. It is a centralized repository for storage, management and distribution of data and information. Data centre applications have group communication. Multicast routing supports group communication by both saving network traffic and improving application throughput. Receiver driven multicast routing protocol is inefficient for number of links in a formed multicast tree and not support for largely connected networks and their routing entries. The results show that there is a severe link waste compared to efficient one. To build efficient multicast tree ESM uses source to receiver expansion approach. It eliminates unnecessary intermediate switches used in receiver driven multicast routing. For scalability ESM uses and combines both in packet bloom filter and in-switch entries to make the transaction between number of multicast groups. During packet forwarding there may be a chance to loss of packets. To reduce the packet loss in a multicast group by combining packet bloom filter and contention window adaptation (CWA). Results can be analysed and executed by Network Simulator.

KEYWORDS

Multicast routing, Data centre network, Packet loss

1. INTRODUCTION

Data centre is the main key infrastructure for cloud computing. Servers, with a scale of hundreds of thousands, are interconnected to run distributed computations for cloud service applications. Data centres are not only run on cloud based applications, but also works on back-end infrastructural computations that includes distributed file systems [2], [3]. Multicast routing supports group communication by both saving network traffic and improving application throughput. Though network- level multicast group communication has many issues such as congestion control, cost and security concern, recently there is a noticeable recovery of it.

This paper explores network level multicast routing, to build the multicast delivery tree in modern computer centre networks. The modern data centre poses new challenges for bandwidth hungry, cost, large-scale and efficient multicast routing. The data centre topology usually represents a high link density. For example Bcube [4], consist of both servers And switches which contains multiple ports for interconnections. Connecting a large population of servers, Fat Tree [5] and VL2 [6] uses number of switches and tens of ports. Pair of servers or switches consist equal-cost paths. In

this scenario receiver-driven multicast routing can result in severe link waste compared to efficient one.

In most data centre low commodity switches used to design for scalability and economic considerations. Memory space of the routing table is relatively narrow. So it is pretty challenging to support a several number of multicast groups in data centre networks.

The ESM will discover network level multicast routing, for building the multicast delivery tree, in modern data centres. There are two forwarding entries in multicast. First without any topological information the multicast addresses are logical identifiers. Second, each multicast forwarding entry maintains an outgoing interface.

The containerized data centre is the basic units for building modern data centre applications. This containerized data centre packs 1k to 2k (approximately) servers and switches into a 20 or 40 feet shipping container. It has the advantages in easy wiring, low cooling cost and high power density. The DCN (Data Centre Network) architecture consists of two categories namely switch-centric and server-centric. The switch-centric is used to improve the low bisection bandwidth. And server-centric the server uses multiple NIC ports to interconnect a networking architecture and participates in packet forwarding. In a server centric the server not only acts as an end hosts it can act as a relay nodes for each others.

By interconnecting containers it uses an existing structure such as Bcube and fat tree and it treat each container as a unit. The goal of designing a data centre application has three major categories. They are Scalable interconnection bandwidth, Economics of scale and Bandwidth compatibility. The Bcube network architecture specially designed for shipping container based data centres. Bcube is robust and high performance network architecture. Bcube is designed to support the bandwidth-intensive and one-much communications [7].

2. RELATED WORKS

Multicast routing supports group communications by saving network traffic and improving application throughput. The receiver driven multicast routing is inefficient for number of links, densely connected networks and it uses a low end commodity switches. It is well not supported for densely connected networks and their routing entries for massive groups of multicasting. The data centre network uses ESM an efficient and scalable multicast routing. Multicast is supported for routers and end-host. The data centre environments also give a good chance for multicast operation. ESM address the challenges of data centre applications by improving the throughput and bandwidth.

ESM uses source to receiver expansion approach and it depends on multicast manager for a multicast routing environment to control the data centre environments. By developing the multistage graph, the multicast manager builds multicast tree using source to receiver expansion approach.

The source to receiver expansion approach overcome the problems in receiver driven multicast routing given that it eliminates needless intermediate switches in a formed multicast tree. The tree links are dynamically added or removed by using dynamic receiver join/leave without affecting the whole tree. And it avoids the out of order packets during multicast delivery.

Achieving the scalability, ESM combines both in-packet bloom filter and in switch entries. By using this approach there is a transaction between number of multicast groups to be supported. The in-packet bloom filter is mainly used for small sized groups by saving the routing space in switches. The small sized groups are most common in data centre applications.

2.1 Background of data centre multicast

The group communications is most widely used in computer centre applications. Multicast is the technology that supports this kind of group communication. This communication purposes are both saving network bandwidth and reducing the load on the sender. Multicast can be supported by most network devices such as routers and switches. Due to the several technological causes switches and routers are not commonly arranged in the internet.

The data centre network is usually managed and used by single authority. The servers are treated as a secret one. The protocol which was implemented at switches and servers are network oriented. it have to investigate whether the implemented protocol can hold the data centre environment.

For efficiency look for to build multicast trees take up as small number of links which was used to save the network bandwidth. For Scalability, try to support a large number of multicast groups.

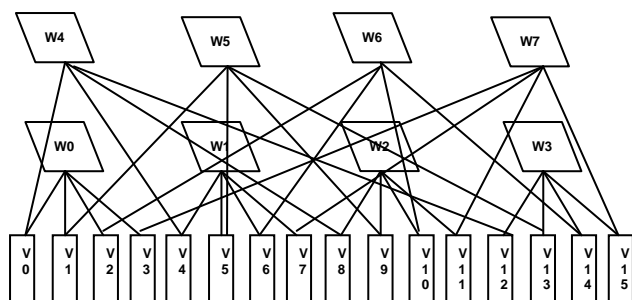


Figure:1 Bcube(4,1) architecture

2.2 Architecture of data centre network

Data centre servers are connected in the form of hierarchical manner of Ethernet switches, with commodity one at the first level and other expensive ones at higher level. This kind of hierarchical structure may cause many problems.

Bandwidth bottleneck can be occurred in top-level switches so high-end high speed switches have to be used. By using a high level switch there may be a single point failure in a sub tree branch. The problem cannot be solved by using unused switches, but it earns higher cost. The new data centre networks need a high bisection bandwidth to replace a tree structure, so Bcube and Fat tree for these architectures. **Bcube:** Bcube is a server-centric interconnection topology, which focussed for shipping-container based, modular data

centres, naturally sized of 1 k-4 k servers. It is constructed in recursive way. To speeding up one-to-one, one-to-several and one-to-all traffic patterns. Bcube supports various bandwidth-intensive applications. Bcube exhibits stylish show degradation as the server failure rate increases. The importance for shipping-container data centres, while once the container is preserved and prepared, it becomes very complex to repair its components. A Bcube (n, 1) is build from n n n-port switches. Bcube gives a multiple parallel short paths between any pair of servers.

By using Bcube is not only providing a high bandwidth and also it improves the fault tolerance and load balancing. It is mainly used to exchange the large amount of data that is exist server pairs such as disk backup. This Bcube network architecture supports server-centric approach than switch-centric approach. And it provides a high network capacity for all-all traffic in map reduce.

Fat-Tree: There are three levels in fat tree that can be represented in fig (2) with 4 port switches. And there is an n-port which contains the two levels of n/2 switches that is aggregation level and edge level. In edge level n-port switches can uses n/2 ports to connect the n/2 servers and also n/2 aggregation level switches used to connect with a remaining n/2 ports.

In the core level, there are (n/2)² n-port switches and every switch has single port connecting to single shell. So, the total number of servers supported in fat tree is n³/4. First low end switches are used to interconnect the server instead of high-end expensive. Second high link density survive in these networks since a more number of switches use several ports for interconnection.

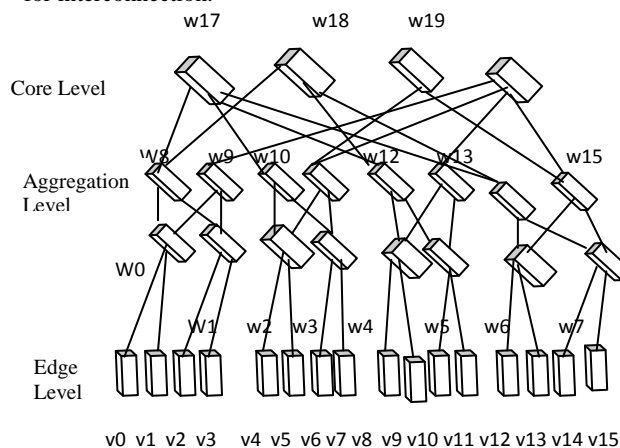


Figure: 2 Fat tree architecture with 4-port switches

Third the data centre is constructing in a hierarchical and normal way on the more population of servers/switches.

2.3 Design Basis

For efficient and scalable multicast routing the modern data centre network architecture create a new challenges to obtain it. First the receiver driven multicast routing protocol is inefficient in terms of redundant links in data centre networks. Second the spaces of the routing table entries are somewhat limited in low end commodity switches.

In heavily connected data centre networks there are frequently a more number of applicants for a group. If receiver-driven multicast routing, then the resultant tree have 14 links but instead use efficient and scalable multicast routing have only 9 links. Comparing with unicasts due to two reasons it is especially is not easy to aggregate multicast route entries.

The BCube topology which can be shown in Fig. 1. Assume the receiver set is {v5, v6, v9, v10} and the sender will be v0.

v0 → w0 → v1 → w5 → v5
 v0 → w4 → v4 → w1 → v6
 v0 → w4 → v8 → w2 → v9
 v0 → w0 → v2 → w6 → v10

In ESM, construct in following way:

v0 → w0 → v1 → w5 → v5
 v0 → w0 → v2 → w6 → v6
 v0 → w0 → v1 → w5 → v9
 v0 → w0 → v2 → w6 → v10

Assume that multicast manager exists in data centre network, to control the multicast routing management, allocate the multicast addresses, constructing the multicast trees and distributing the switches and server to the networks. For example consider the Fat-tree; a fabric manger is responsible for managing the network environment. To calculate multicast tree, configure the multicast routing rules on switches and send the tree information to the source server Bloom-Filter – based multicast routing is used. Single commodity machine should be work well in online tree building in ESM.

In the multicast manager is helping for more data centre networks, a group of servers can work together to play the role.

Second use the normal data centre network topology for multicast routing design. Later it will show that the Steiner-tree algorithm [7] is too slow in the large size of network.

To accelerate the tree calculation process, ESM leverages hierarchical regular data centre topology. To design a loop-free Bloom-Filter-based multicast forwarding engine in ESM depends on the topological characteristics of data centre networks. Third by examining that the small-sized groups are governing data centre and large- sized groups take significant bandwidth waste in in-packet Bloom-Filter-based multicast routing. For limited number of large groups, to avoid the traffic leakage multicast routing entries are installed in the switches. For the large volume of small groups, to eliminate the requirement of in-switch booting table space in-packet bloom filter is used.

3. COMPETENT MULTICAST TREE BUILDING

To save network traffic and also to reduce the task finish time of data centre applications, it requires ESM multicast tree.

3.1 Problem

ESM build a multicast tree to eliminate the unnecessary switches used in independent receiver-driven multicast routing. Based on the data centre topology the group membership distribution, the multicast manager calculates the multicast tree.

The problem is then transformed to calculate an efficient multicast tree on the multicast manager. Notice that the building of Steiner tree in general graphs is NP-Hard. The multicast tree building in Fat-Tree is polynomial for the recently proposed data centre networks.

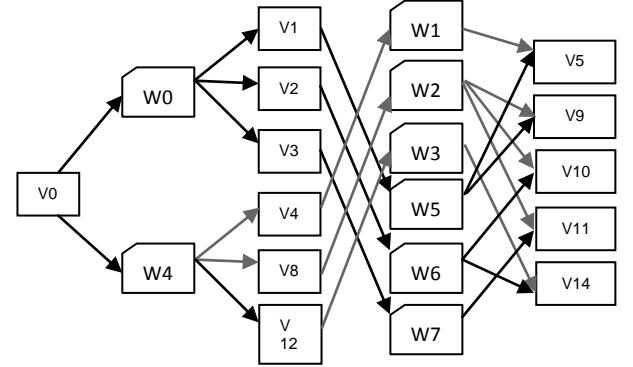
Theorem 1: In Bcube network and the Steiner tree problem is NP-Hard.

Proof: Please refer to [5].

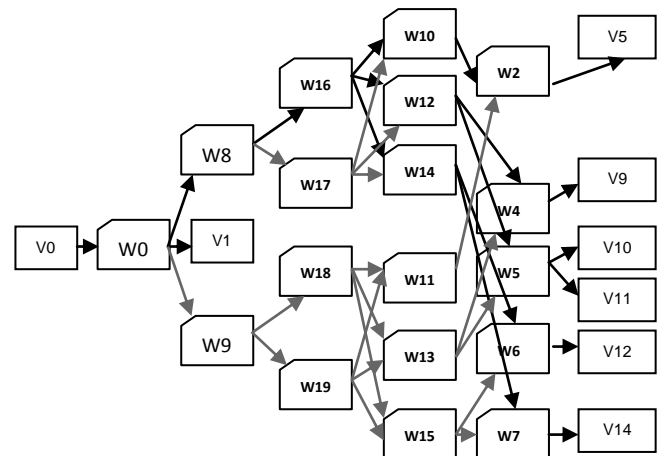
ESM focus at building efficient multicast tree in common data centre networks. Due to low bandwidth and low through put there is an occurrence of packet loss [1].

3.2 Source Driven Tree Building

Data centre architecture use more than few levels of switches for server interconnection, and there is no direct connection for switches within the same level and thus they are multistage graphs [9]. The paths that are from multicast source can be expanded to all receivers in the path as a multistage graph which is directed with d+1 stages where the d is a network topology. This will call as a *group spanning graph* for the multicast group. Stage 0 contains only the sender, and stage d is composed only with receivers. Any node should appear at most once in the group spanning graph.



(a)



(b)

Figure: 3 Demonstrated group spanning tree.

BCube: The selection of tree node in BCube network can be conducted on the group spanning graph in an iterative way.

The set of servers are selected from the stage 2 of a group spanning tree for a BCube (n, k) with a sender s. The process will be iterated until getting all BCube (n, 0). O(N) is the computation complexity of BCube, where N represents the total number of servers in BCube.

Fat-Tree: The number of hops in cluster straddling graph depends upon the span of the network. For example: if the diameter of the network is 8, then the number of hops is 8. The selection of nodes for the tree is carried out by following steps. From first stage to centre height stage, only single pathway is choose which covers most of the downstream receivers. From the stage of core-level to final stage, the path is set by some interconnection convention of fat tree.

3.3. Active Receiver Join/Leave

Multicast receivers actively join/leave a group the multicast tree should be rebuild to include group dynamics. ESM can

kindly hold this case because the active receiver join/leave cannot alter the source-end paths of the other receivers in the group. During the packet delivery it is important for avoiding the out-of-order packets.

BCube: While a new receiver r_j joins an accessible group in a BCube (n, k) , first have to extend the preceding cluster straddling graph to call up r_j . To check whether there is a BCube (n,k) present in the previous multicast tree in new cluster straddling graph that can contain r_j , then have to add r_j to the BCube $(n,0)$ [1]. This kind of process are not able change the source-to-end paths for new receivers.

Fat-Tree: When a receiver r_j joins a group in Fat-Tree, it should be added at the final stage of the group spanning graph. In the case of new tree, there is no need to do anything in the tree except adding the path from the early chosen core-level switch to r_j . Notice that the core-level switch to r_j are set, and they may overlies with a number of links in the prior multicast tree.

4. DESIRABLE MULTICAST ROUTING

A scalable multicast routing scheme not only constructs an efficient multicast group, and also scales to a bulky number of multicast groups.

4.1 Grouping Routing Scheme

It supports very big multicast groups using usual way of installing all the multicast routing entries in to switches, especially the low-end service switches frequently used in data centres. The alternative way to eliminate the requirement of in-switch entries used in the data centre by using an In-packet Bloom Filter. During multicast forwarding bandwidth waste can be caused by in-packet Bloom Filter.

Bandwidth Slide Relation: There are two aspects of bandwidth waste by using in-packet Bloom filter. First the network bandwidth cost will be takes out from Bloom Filter field in the packets. Second the traffic leakage can be caused false-positive by Bloom Filter. Likewise, switches in receipt of by false positive [10] forwarding can more forward packets to new switches, invite not only further traffic leakage, but also possible loops. By defining the bandwidth slide Relation of in-packet Bloom Filter, r , which is a ratio of additional traffic caused by in-packet Bloom Filter above the real payload traffic to hold.

The network size of the BCube $(8, 3)$ and 48 port switches of Fat-Tree, and we have to choose the number of hash function reducing the false positive ratio, and calculate the bandwidth slide relation for unusual group sizes along with raise of the in-packet Bloom Filter length. First, for a given group size, the “optimal” length of the in-packet Bloom Filter minimizes the bandwidth overhead ratio. Then the Bloom Filter length is smaller than the optimal value. The major factor for the bandwidth over head ratio is false-positive forwarding. While length grow is larger than the optimal value, the Bloom Filter itself control the bandwidth overhead.

4.2 Encoding the tree

There are node-based encoding and link-based encoding are the two ways to encode the tree information with in-packet bloom filter. In node-based bloom filters that includes both switches and servers are the elements of tree nodes. In node based bloom filters ambiguity may exist, that was called as tree ambiguity problem. But in link based bloom filter there is no tree ambiguity problem since tree link is clearly recognized.

Directed physical links are the elements of the link-based bloom filters. Instead of node based bloom filter, link based encoding is used in LIPSIN: Line Speed Publish/Subscribe Inter-Networking [8]. In node based bloom filter there is no ambiguity problem exist while it may satisfy the certain requirements.

Theorem 2: Given a tree T and topology G , the tree ambiguity problem can be avoided, except the links from T .

4.3 Loop Free Forwarding Scheme

The loop may occur due to the false positive forwarding, and there is no issue in node based and link based encoding scheme. For an example take fig2 the packets is forwarded in the path of $w_8 \rightarrow w_{17} \rightarrow w_{10} \rightarrow w_{16} \rightarrow w_8$, while w_8 forward the packets to w_{17} . Due to false positive forwarding the loop may forms.

Solving the false positive forwarding distance-based bloom filter forwarding scheme used in ESM. It classifies the distance between two nodes as the number of their shortest-path hops on the data centre topology.

Theorem 3: For free of looping distance-based Bloom filter forwarding scheme is used.

5. COMBINING IN PACKET BLOOM FILTER AND CONTENTION WINDOW ADAPTATION (CWA)

The data centre, as an interconnection is used to achieve the distributed processing tasks. For achieving the performance it has three key elements to determine, they are (1) Interconnection topology (2) Networking architecture (3) Routing scheme. To realize the scalable multicast routing in data centre networks by using both in-packet bloom filter and in-switch entries. Particularly, to save the routing spaces in switches in-packet bloom filter is used for a small size groups for a large group will improve the bandwidth overhead while the routing entries are installed into the switches.

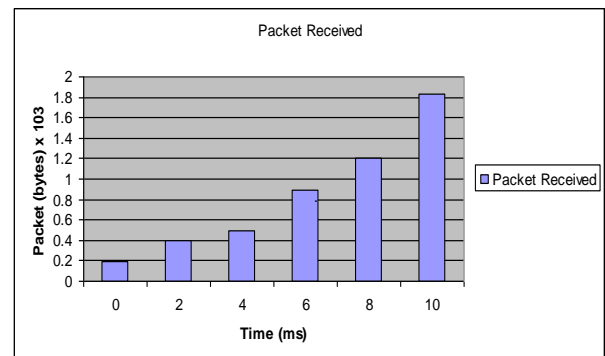


Figure 4: Calculation of packet receiver

The ESM will only achieves both scalability and efficient by increasing the throughput and network bandwidth in a data centre applications. In case of using ESM, due to packet forwarding there may be a possibility of packet delay and loss of packets. So we proposed to overcome the above problem by combining both in-packet bloom filter and contention window adaptation algorithm (CWA).

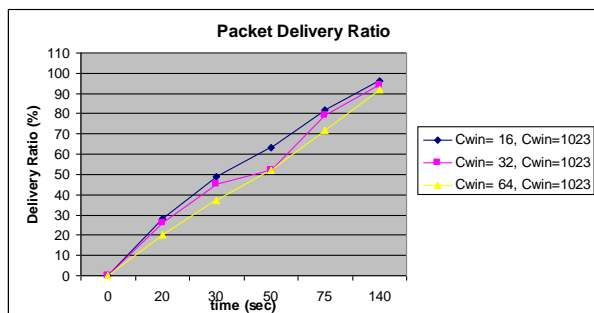


Figure 5: Packet-delivery-ratio

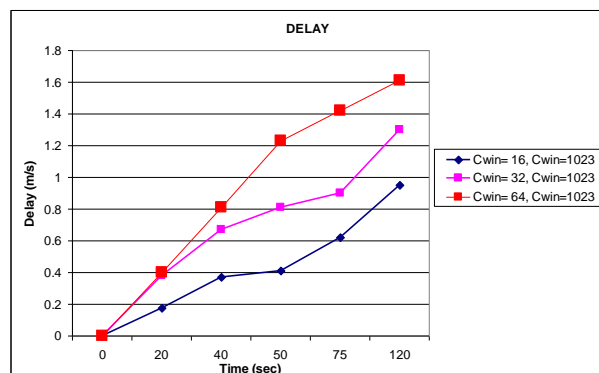


Figure 6: Delay of packets

In data centre applications the bandwidth is referred as a data transfer rate and throughput is referred as packet ratio that is average rate of packet delivery over a communication network. And throughput can be measured by using bps (bits per second). The CWA used to minimize the network traffic and improve the network capability. By combining both in-packet bloom filter and CWA we achieve better bandwidth utilization, throughput and minimizing the loss of packets [11]. We show the results in both wireless and wired applications by using network simulator.

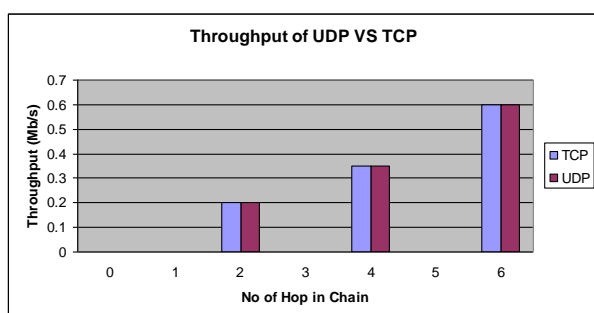


Figure 7: Traffic forwarding capability

6. CONCLUSION

By using of in-packet bloom filter and contention window adaptation, the delay of packets and loss of packets can be minimized. The throughput, Bandwidth also increased and energy also consumed. It saves the network traffic and also achieves the scalability and efficiency in data centre applications. The simulation results show that tree building algorithm, effectively reducing the packet loss, delay of packets and consumption of energy. Results can be analysed and executed by network simulator.

7. REFERENCES

- [1] Dan Li, Yuanjie Li, Jianping Wu, Sen Su, and Jiangwei Yu, "ESM: Efficient and Scalable Data Centre Multicast Routing". IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 20, NO. 3, JUNE 2012.
- [2] S. Ghemawat, H. Gobio, and S. Leungm, "The Google file system," in Proc. ACM SOSP., 2003, pp. 29–43.
- [3] "Hadoop," Apache Software Foundation, 2011 [Online]. Available: <http://hadoop.apache.org/>.
- [4] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, and C. Tian, "BCube: A high performance, server-centric network architecture for modular data centres," in Proc. ACM SIGCOMM, Aug. 2009, pp. 63–74.
- [5] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data centre network architecture," in Proc. ACM SIGCOMM, Aug. 2008, pp. 63–74.
- [6] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "VL2: A scalable and flexible data centre network," in Proc. ACM SIGCOMM, Aug. 2009, pp. 51–62.
- [7] D. Li, M. Xu, H. Zhao, and X. Fu, "Building mega data centre from heterogeneous containers," in Proc. IEEE ICNP, Oct. 2011, to be published.
- [8] P. Jokela, A. Zahemszky, C. E. Rothenberg, S. Arianfar, and P. Nikander, "LIPSIN: Line speed publish/subscribe inter-networking," in Proc. ACM SIGCOMM, Aug. 2009, pp. 195–206.
- [9] D. Li, J. Yu, J. Yu, and J. Wu, "Exploring efficient and scalable multicast routing in future data centre networks," in Proc. IEEE INFOCOM, Apr. 2011, pp. 1368–1376.
- [10] Bose, P., Guo, H., Kranakis, E., Maheshwari, A., Morin, P., Morrison, J., Smid, M., and Tang, Y. (2008). On the false-positive rate of Bloom filters. Information Processing Letters, 108(4):210–213.
- [11] Hui Ma, Sumit Roy., "Contention Window and Transmission Opportunity Adaptation" for Dense IEEE 802.11 WLAN Based on Loss Differentiation. WA 98195-2500.