THE INTERNATIONAL JOURNAL OF
SCIENCE & TECHNOLEDGE

Reliable Multicasting for Power Optimization in Data Center Networks

Prabakaran S.
M.Tech. Student, Department of Computer Science and Engineering, SRM University, Chennai, India
Deeban Chakravarthy V.
Assistant Professor, Department of Computer Science and Engineering, SRM University, Chennai, India

Abstract:
In Multicast Data Center network, Power optimization is a challenge since Data Center network is highly energy concerned as well as cost conscious. The Energy efficiency of a Data Center network must always be high for reliable performance. Reliable Multicast in Data Center Networks, which minimize the packet loss and optimize the power consumption, is proposed. Energy Efficient Routing is proposed for power optimization. Energy Efficient Routing is done by shutting down the unused links during off-peak traffic times in the data center network. The backup Overlay algorithm is suggested to minimize the packet loss in Data Center networks. During high speed data transmission, the Backup Overlay handles packet loss and also responds to link failure and receiver failure causing curtailed CPU overhead. This RDCM is basically designed for generic data center topologies. BCube tree is presented and simulation is carried out. But they can be easily extended to other data center networks. The proposed RDCM causes reduced Power consumption and reduced CPU overhead to datacenter servers.

Keywords: *RDCM, Backup Overlay, packet loss, BCube tree, CPU overhead*

1. Introduction

In Data Center Networks, Reliable multicast is very essential. Reliable Multicasting provides a reliable sequence of packets to multiple recipients simultaneously, making it suitable for applications like multi-receiver file transfer or streaming media. So in order to guarantee the successful packets delivery to multicast receivers, reliable multicasting is important. Network devices consume 20 - 30 % energy of the whole data center's power consumption. This large amount of power which is wasted must be reduced. Traffic in Data Center Networks vary greatly between daytime and night, which means, traffic peaks during the day and falls at night. This variation in traffic can be used to reduce the power consumed by DCN.

There are many challenges in data center networks. Some of them are:

Traditional networks are closed, and they separate the network layer from the application layer. In the cloud computing era, the data center network needs to respond to upper-layer application requests immediately, provide dynamic QoS and security policies, and enable real time status monitoring. An intelligent data center network that supports service collaboration is critical for Internet enterprises.

The Internet provides many different applications and is open. Internet enterprises are required to provide flexible and fast service deployment for third-party partners. An open platform meets the requirements of frequently changing Internet services and enhances enterprise competitiveness. Server virtualization is one of the core cloud computing technologies that support an open platform. The data center network must allow VM migration and flexible service deployment.

10GE server interfaces have been in use for over 10 years, but in the cloud computing era, virtual servers require higher bandwidth. In the next 10 years, 10GE/40GE/100GE interfaces will coexist on data center networks, and if a data center network cannot support GE, 10GE, 40GE, and 100GE servers, the network will have to be upgraded each time a faster server is deployed.

2. Architectural Design

In the Architecture of Reliable multicast system,

- Single acknowledgment is used for multiple repair packet, thereby reducing the overall repair traffic.
- Repair technique is used to minimize the repair traffic. The repair technique varies with the number of receivers. (unicast or multicast)
- Power optimization is done by using Energy Efficient Routing

The Architecture of Reliable Multicasting system consists of four major components.

i. Multicast manager
ii. Traffic Analyser
iii. Consolidator
iv. Flow Path

The actual user will interact with the system through application via interface available for the data center. The data center network is a controlled environment; here all the operations are controlled and coordinated by the multicast manager. The Data center consists of tens of thousands of servers which is connected to each other in various architectural structures like Fat tree, BCube, VL2. The coordination of this structure can be done through by administrator when operated manually or by software defined networking in case of automated process.

This architecture consists of traffic analyzer and consolidator which will help the multicast manager to perform energy efficient routing operation. The multicast manager will get the detail report on the current status about the network from the traffic analyzer which reference with the status multicast manager instruct the consolidator to invoke the algorithm to perform the consolidation among the available link into a set of link for demand traffic and shutdown the unused network devices. Then the flow path will help the consolidator to arrange the flow efficiently in a way that it avoids violation of adoption more no of flow into a single link to perform efficient scheduling.

2.1. Multicast Manager

Multicast Manager manages the multicast traffic within your network environment. This includes the identification and monitoring of all multicast groups, sources, receivers and other critical sources like rendezvous point (RP) routers. Multicast Manager provides several views which enable us to monitor the connectivity and performance of these Multicast elements. These views offer details relating to Multicast groups, sources, receivers, routers.

The information provided in the Multicast Manager views makes troubleshooting Multicast-related outages much more efficient. One can isolate a problem quickly because the group a receiver belongs to, the RP of the group, the source for the group, the routers and interfaces that a group uses, and the configuration and bandwidth limits of these interfaces can be determined. Based on this information, troubleshooting efforts for multiple problems can be prioritized. In Multicast Manager, membership of multicast groups, layout of multicast trees, active sessions, and protocol entities can be viewed, and how these relate to the underlying IP network can be understood. These capabilities help in delivering the highest levels of service assurance to the business as well as users of multicast services, plus significantly speed repair time when problems arise.

2.2. Traffic Analyser

A packet analyzer which is also known as a network analyzer, protocol analyzer or packet sniffer, or for particular types of networks, an Ethernet sniffer or wireless sniffer is a computer program or a piece of computer hardware that can intercept and log traffic passing over a digital network or part of a network. As data streams flow across the network, the sniffer captures each packet and, if needed, decodes the packet's raw data, showing the values of various fields in the packet, and analyzes its content according to the appropriate RFC or other specifications

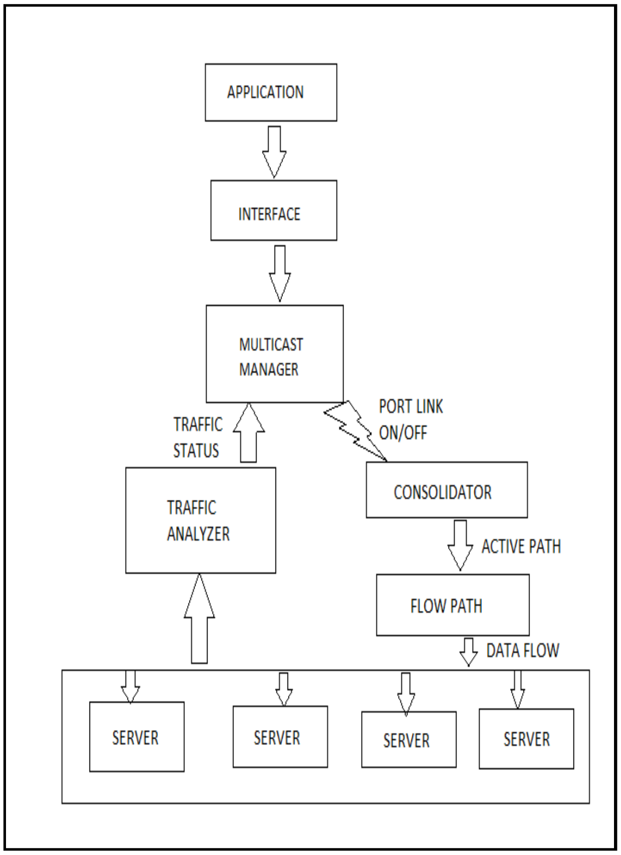

Figure 1: Architectural Design of Reliable Multicast System

On wired broadcast LANs, depending on the network structure (hub or switch), one can capture traffic on all or just parts of the network from a single machine within the network; however, there are some methods to avoid traffic narrowing by switches to gain access to traffic from other systems on the network. For network monitoring purposes, it may also be desirable to monitor all data packets in a LAN by using a network switch with a so-called monitoring port, whose purpose is to mirror all packets passing through all ports of the switch when systems are connected to a switch port. To use a network tap is an even more reliable solution than to use a monitoring port, since taps are less likely to drop packets during high traffic load.

On wireless LANs, one can capture traffic on a particular channel, or on several channels when using multiple adapters.On wired broadcast and wireless LANs, to capture traffic other than uncast traffic sent to the machine running the sniffer software, multicast traffic sent to a multicast group to which that machine is listening, and broadcast traffic, the network adapter being used to capture the traffic must be put into promiscuous mode; some sniffers support this, others do not. On wireless LANs, even if the adapter is in promiscuous mode, packets not for the service set for which the adapter is configured will usually be ignored. To see those packets, the adapter must be in mode. When traffic is captured, either the entire contents of packets can be recorded, or the headers can be recorded without recording the total content of the packet. This can reduce storage requirements, and avoid legal problems, but yet have enough data to reveal the essential information required for problem diagnosis.

The captured information is decoded from raw digital form into a human-readable format that permits users of the protocol analyzer to easily review the exchanged information. Protocol analyzers vary in their abilities to display data in multiple views, automatically detect errors, determine the root causes of errors generate timing diagrams, reconstruct TCP and UDP data streams, etc.Some protocol analyzers can also generate traffic and thus act as the reference device; these can act as protocol testers. Such testers generate protocol-correct traffic for functional testing, and may also have the ability to deliberately introduce errors to test for the DUT’s ability to deal with error conditions.

Protocol analyzers can also be hardware-based, either in probe format or, as is increasingly more common, combined with a disk array. These devices record packets (or a slice of the packet) to a disk array. This allows historical forensic analysis of packets without the users having to recreate any fault.A packet sniffer, sometimes referred to as a network monitor or network analyzer, can be used legitimately by a network or system administrator to monitor and troubleshoot network traffic. Using the information captured by the packet sniffer an administrator can identify erroneous packets and use the data to pinpoint bottlenecks and help maintain efficient network data transmission.

In its simple form a packet sniffer simply captures all of the packets of data that pass through a given network interface. Typically, the packet sniffer would only capture packets that were intended for the machine in question. However, if placed into promiscuous mode, the packet sniffer is also capable of capturing ALL packets traversing the network regardless of destination.

By placing a packet sniffer on a network in promiscuous mode, a malicious intruder can capture and analyze all of the network traffic. Within a given network, username and password information is generally transmitted in clear text which means that the information would be viewable by analyzing the packets being transmitted.

A packet sniffer can only capture packet information within a given subnet. So, its not possible for a malicious attacker to place a packet sniffer on their home ISP network and capture network traffic from inside your corporate network (although there are ways that exist to more or less “hijack” services running on your internal network to effectively perform packet sniffing from a remote location). In order to do so, the packet sniffer needs to be running on a computer that is inside the corporate network as well. However, if one machine on the internal network becomes compromised through a Trojan or other security breach, the intruder could run a packet sniffer from that machine and use the captured username and password information to compromise other machines on the network.

Detecting rogue packet sniffers on your network is not an easy task. By its very nature the packet sniffer is passive. It simply captures the packets that are traveling to the network interface it is monitoring. That means there is generally no signature or erroneous traffic to look for that would identify a machine running a packet sniffer. There are ways to identify network interfaces on your network that are running in promiscuous mode though and this might be used as a means for locating rogue packet sniffers.

The versatility of packet sniffers means they can be used to:

- Analyze network problems
- Detect network intrusion attempts
- Detect network misuse by internal and external users
- Documenting regulatory compliance through logging all perimeter and endpoint traffic
- Gain information for effecting a network intrusion
- Isolate exploited systems
- Monitor WAN bandwidth utilization
- Monitor network usage (including internal and external users and systems)
- Monitor data-in-motion
- Monitor WAN and endpoint security status
- Gather and report network statistics
- Filter suspect content from network traffic
- Serve as primary data source for day-to-day network monitoring and management
- Spy on other network users and collect sensitive information such as login details or user cookies (depending on any content encryption methods that may be in use)
- Reverse engineer proprietary protocols used over the network

- Debug client/server communications
- Debug network protocol implementations
- Verify adds, moves and changes
- Verify the internal control system effectiveness (firewalls, access control, Web filter, spam filter, proxy)

2.3. Consolidator

Traffic flows in a DCN can be consolidated onto a small set of links and switches, which are sufficient to serve the bandwidth demands for most of the time

While traffic consolidation has been demonstrated to be a highly effective way to achieve energy proportionality in DCNs by shutting down unused network devices, existing work consolidates traffic flows in a greedy way and assumes that the bandwidth demand of each data flow can be approximated as a constant during the consolidation process. This is in contrast to the fact that the bandwidth demand of a traffic flow can vary over time. The variations can be significant because the consolidation period normally cannot be very short due to overhead considerations Therefore, existing work has to use either estimated maximum or average demands to perform consolidation, which can result in either unnecessarily high power consumption or undesired link capacity violations, respectively.

The bandwidth demands of different flows usually do not peak at exactly the same time. As a result, if the correlations among flows are considered in consolidation, more power savings can be achieved. Another important observation is that the 90-percentile bandwidth demands are usually half or less of the peak demands. Therefore, if we could avoid consolidating traffic flows that are positively correlated (e.g., peak at the same time) based on 90-percentile demands instead of peak demands, we may further improve the energy efficiency of traffic consolidation. Power optimization scheme that consolidates traffic flows based on correlation analysis among flows in a DCN. DCN network flows usually have weak pair-wise correlations and thus do not peak at the same time. Based on the observation, consolidation of traffic flows according to their correlation is done.

The first approach is traffic consolidation – consolidate traffic flows in a DCN onto a small set of links and switches such that unused network equipment’s can be turned off for power saving. Although their consolidation approach is developed based on a real data center traces, they assume that the traffic rate of each data flow is approximately a constant, which may not be valid for many of the current production data centers due to their high variations in workloads.

The correlations among flows and consolidates traffic flows based on the 90-percentile of each traffic flow’s peak demand, such that only a small set of the network devices is needed to hold all the traffic flows while the rest of the network devices can be shut down for power savings.

2.4. Flow Path

A central scheduler, possibly replicated for fail-over and scalability, manipulates the forwarding tables of the edge and aggregation switches dynamically, based on regular updates of current network-wide communication demands. The scheduler aims to assign flows to non-conflicting paths; more specifically, it tries to not place multiple flows on a link that cannot accommodate their combined natural bandwidth demands. Whenever a flow persists for some time and its bandwidth demand grows beyond a defined limit, we assign it a path using one of the scheduling algorithms

Depending on this chosen path, the scheduler inserts flow entries into the edge and aggregation switches of the source pod for that flow; these entries redirect the flow on its newly chosen path. The flow entries expire after a timeout once the flow terminates. The state maintained by the scheduler is only soft-state and does not have to be synchronized with any replicas to handle failures.

Scheduler state is not required for correctness (connectivity); rather it aids as a performance optimization to make intelligent flow placement decisions, we need to know the flows’ max-min fair bandwidth allocation as if they are limited only by the sender or receiver NIC. When network limited, a sender will try to distribute its available bandwidth fairly among all its outgoing flows. TCP’s AIMD behavior combined with fair queuing in the network tries to achieve max-min fairness. Note that when there are multiple flows from a host A to another host B, each of the flows will have the same steady state demand. The demand estimator performs repeated iterations of increasing the flow capacities from the sources and decreasing exceeded capacity at the receivers until the flow capacities converge.

2.5. Advantages

- Reduced overhead for repair packet traffic.
- Energy saving without impacting network performance

3. Algorithm for Consolidator

- Input: Flow list F *U {f_i}* with *M* flows, correlation threshold ,link capacity and m *PATH* .
- Output: Final Path List

1. Form the correlation matrix [M][M] using the Flow list F.
2. Set Date rate for each Flow (i.e.) rate [M].
3. Start of Loop1, loops till the Flow list become NULL.
4. Start of Loop2, for each value of j-> 1 to m.
5. Start of Loop3, loops for all value of fi that can take the path[j].
6. If the link capacity ≥ the Flow rate[i]

```
7. Then go to step 7
8. Else step 5.
9. If Correlation b/w (fi) and (fi') ≤ Correlation threshold
10. Then go to step 8
11. Else step 5.
12. Add the current path[j] to Final Path List.
13. Reduce the allocated Flow(fi)from the Flow list F.
14. Update the Capacity of the Path[j].
15. End IF.
16. End IF.
17. End Loop3.
18. End Loop2.
19. End Loop1.
20. Return Final Path List.
```

The algorithm we designed is based on the greedy-bin packing algorithm. We greedily assign as many traffic flows as possible to a single path. The algorithm takes the flow list F, the link list l, link capacity c, path link list PATHL for each available path, and the correlation threshold Corthas input. Each entry in PATHL is a link list of one path. The path list is ordered from left to right based on the network topology. In the Algorithm, the correlation value (Cor) between all flow pairs, and the data rate (rate) of each flow are 91nitialized. Then each flow to a path is assigned. The correlation between flow f_i and the flows existing on path j meets the correlation requirements is checked. If both of these two requirements are satisfied, flow f_i is assigned to path j and the available link capacity of each link along path j is updated. Then the available link capacity is updated with the 90-percentile link utilization value of the aggregated traffic after the new flow is assigned in each step. Program terminates when all the flows are assigned.

4. Conclusion

In Multicast Data Center network Power optimization has always been a challenge. Reliable multicast is important for data center network to guarantee the successful packets delivery to multicast receiver. But network devices consume 20 - 30 % energy in whole data center. In this project, Reliable Multicast in Data Center Networks minimizes the packet loss and optimizes the power consumption. Traffic in data center network varies greatly between daytime and night (i.e) traffic peaks during day and falls at night. Thus, for Power Optimization, Energy Efficient Routing is done by shutting down the unused links during off-peak traffic times in data center network. The modules present in RDCM which helps in this power optimization and packet loss minimization are Multicast Manager, Traffic Analyzer, Consolidator and Flow Path. The proposed RDCM causes reduced CPU overhead to data center servers. Power consumption is also expected to have a remarkable reduction.

5. References

1. A.Greenberg, J. Hamilton, N. Jain et al., "VL2 a scalable and flexible data center network," in Proceeding. ACM SIGCOMM, pp. 51–62, Aug. 2009.
2. A.Greenberg, J. Hamilton, D. Maltz et al., "The cost of cloud research problems in data center networks," in Proceeding. SIGCOMM Computer Communication Review, vol. 39, no. 1, pp. 68–73, 2009.
3. B. Adamson, C. Bormann, M. Handley et al., "NACK-oriented reliable multicast transport protocol," RFC3208, 2009.
4. C. Guo, G. Lu, D. Li et al., "BCube a high performance, server centric network architecture for modular data centers," in Proceeding. ACM SIGCOMM, pp. 63–74, Aug. 2009.
5. C. Hanle and M. Hofmann, "Performance Comparison of Reliable Multicast Protocols using the Network Simulator ns-2," Proceedings of IEEE Conference on Local Computer Networks, Boston, MA, USA, October 11-14, 1998.
6. C. Guo, G. Lu, Y. Xiong et al., "Datacast a scalable and efficient group data delivery service for data centers," Microsoft Tech. Rep.MSR-TR-2011–76, Jun. 2011.
7. D. Li, J. Yu, J. Yu et al., "Exploring efficient and scalable multicast routing in future data center networks," in Proceeding. IEEE Conference on Computer Communications, pp. 1368–1376, Apr. 2011.
8. D. Li, C. Guo, H. Wu et al., "Scalable and cost-effective interconnection of data center servers using dual server ports," IEEE/ACM Transaction on Networks, vol. 19, no. 1, pp. 102–114, Feb. 2011.
9. H. Holbrook, S. Singhal, and D. Cheriton, "Log-based receiver reliable multicast for distributed interactive simulation," in Proceeding. ACM SIGCOMM, pp. 328–341, Oct. 1995.
10. IEEE Standards for Local and Metropolitan Area Networks Virtual Bridged Local Area Networks, IEEE Standard 802.1Q, 2005.
11. J. Griffioen and M. Sudan, "A reliable dissemination protocol for interactive collaborative applications," in Proceeding.
12. J. Griffioen and M. Sudan, "A reliable dissemination protocol for interactive collaborative applications," in Proceeding. ACM Multimedia, pp. 333–344, Nov. 1995.
13. J. Chang and N. Maxemchuk, "Reliable broadcast protocols," AC Transaction on Computer Systems, vol. 2, no. 3, pp. 251–273, 1984.
14. K. Obraczka, "Multicast transport mechanisms a survey and taxonomy," IEEE Communications Magazine, vol. 36, no. 1, pp. 94–102, Jan. 1998.

15. K. Obraczka, “Multicast transport mechanisms: A survey and taxonomy,” IEEE Communications Magazine, vol. 36, no. 1, pp. 94–102, Jan. 1998.
16. K. Birman, M. Handley, O. Ozkasap et al., “Bimodal multicast,” ACM Transactions on Computer Systems, vol. 17, no. 2, pp. 41–88, 1999.
17. L. Lehman, S. Garland, and D. Tennenhouse, “Active reliable multicast,” in Proceeding. IEEE Conference on Computer Communications, Mar. 1998.
18. M. T. Lucas. “Efficient Data Distribution in Large -Scale Multicast Networks,” Ph.D. Dissertation, Department of Computer Science, University of Virginia, May 1998.
19. M. Al-Fares, A. Loukissas, and A. Vahdat, “A scalable, commodity data center network architecture,” in Proceeding. ACM SIGCOMM, pp. 63–74, Aug. 2008.
20. M. Balakrishnan, K. Birman, A. Phanishayee et al., “Ricochet: Lateral error correction for time-critical multicast,” in Proceeding. USENIX Symposium on Network System Design and Implementation, pp. 73–86, 2007.
21. M. Borella, D. Swider, S. Uludag et al., “Internet packet loss measurement and implications for end-to-end QoS,” in Proceeding. International Conference on Parallel Process, 1998.
22. OpenFlow [Online]. Available: http://www.openflowswitch.org/, accessed on 2012.
23. P. Eugster, R. Guerraoui, S. Handurukande et al., “Light weight probabilistic broadcast,” ACM Transaction on Computer Systems, vol. 21, no. 4, pp.341–374, 2003.
24. Telecom R&D, “Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences influence of the recent effect,” ITU Study Group12, contribution D.139, 2000.
25. R. Yavatkar, J. Griffioen, and M. Suda. “A Reliable Dissemination Protocol for Interactive Collaborative Application,” In Proceedings of the ACM Multimedia Conference, Nov. 1995.
26. S. Floyd, V. Jacobson, S. McCanne et al., “Reliable multicast Framework for light-weight sessions and application level framing,” in Proc. ACM SIGCOMM, pp. 342–356, Oct. 1995.
27. S. Ghemawat, H. Gobioff, and S. Leung, “The Google file system,” in Proceeding. Symposium on Operating System Principles, pp. 29–43, Oct. 2003.
28. S. Kandula, S. Sengupta, A. Greenberg et al., “The nature of datacenter traffic measurements and analysis,” in Proceeding. ACM SIGCOMM Internet Measurement Conference, pp. 202–208, Nov. 2009.
29. S. Paul, K. Sabnani, J. Lin et al., “Reliable multicast transport protocol,” IEEE J. Sel. Areas Communications, vol. 15, no. 3, pp. 1414–1424, Apr. 1997.
30. T. Speakman, J. Crowcroft, J. Gemmell et al., “PGM reliable transport protocol specification,” RFC3208, Dec. 2001.
31. T. Benson, A. Anand, A. Akella et al., “Understanding data center traffic characteristics,” in Proceeding. Workshop Research. Enterprise Network, pp. 65–72, 2009.
32. T. Speakman, J. Crowcroft, J. Gemmell et al., “PGM reliable transport protocol specification,” RFC3208, Dec. 2001.
33. Telecom R&D, “Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences: Influence of the recent effect,” ITU Study Group12, contribution D.139, 2000.
34. V. Markovski, “Simulation and Analysis of Loss in IP Networks – Simulation scenarios,” M. Sci. Thesis, Department of Engineering Science, Simon Fraser
35. University, pp. 24-30, Oct. 2000.
36. Y. Vigfusson, H. Abu-Libdeh, M. Balakrishnan et al., “Dr. Multicast Rx for datacenter communication scalability,” in Proceeding. ACM European Conference on Computer Systems, pp. 349–362, Apr. 2010.
37. Y. Yang and S. Lam, “Internet multicast congestion control a survey,” in Proceeding. International Conference on Telecommunication, 2000.