Special Issue | August 2017 ISSN: 2320-9984 (Online)

International Journal of Modern Engineering & Management Research

Website: www.ijmemr.org

# All India Seminar on

## Futuristic Trends in Telecommunication Engineering & Telecom Panorama – Fundamentals and Evolving Technology, with Particular Reference to Smart City on 5th – 6th August 2017

## Organized by

## The Institution of Engineers (India) Jabalpur Local Centre

# A Cloud Aided Content Distribution and Registration Service

**Bhawna Kol**
*M.Tech Research Scholar*
*Shri Ram Group of Institutions*
*Jabalpur (M.P.), [INDIA]*
*Email:bhawnakol@gmail.com*

**Sapna Choudhary**
*Associate Professor*
*Department of Computer Science*
*Shri Ram Group of Institutions*
*Jabalpur (M.P.), [INDIA]*
*Email: choudharysapnajain@gmail.com*

***Abstract*****—***The huge data term is represented by the growing arrival ratio of live content crisis applications stand a great question: how to distribute data with various sizes for relevant clients in a reliable and scalable manner. The distribution/registration model is broadly uses for the data distribution due to its capacity of smoothly and continuously enlarging the system to big size. The existing pub/sub systems provides low matching throughput when matching the skewed subscriptions in a large amount, and interfere in data distribution when a large amount of servers fail. In this paper, we propose CADAR, a cloud aided content distribution and registration service in cloud computing environment. We propose a distributed overlay to arrange brokers, to obtain low routing latency and reliable connections between brokers. To obtain high matching latency and multiple candidate servers for each event, we introduce a hybrid space partitioning technique HPartition-8. Compared with the existing approaches, CADAR gives the better achievements.*

***Keywords:—****Cloud Computing, Data Dissemination, Distribution/Registration, HPartition.*

## 1. INTRODUCTION

In recent years, data distribution has evolved into the essential element in the sense-response applications like disaster weather warning [1], stock information update, smart transportation systems [2] and earthquake monitoring [3]. Modern data distribution presents two trends. One is the increasing arrival rate of large-scale live content. For instance, in every second Instagram, Facebook, Twitter users contribute to about 5 million bits of content and Apple, Android users download just about 50 000 apps. The other is the variations or maybe we can say that the differences in live content. For instance, Wall posts, comments, photos, audios, videos and software's are jumping up in the internet environment and the sizes of the live content range from tens of kilobytes to hundreds of megabytes. In sense-response systems, the aim of the data

dissemination is to timely inform user's up to date live content with user's interests. First, with large-scale of live content, it is essential to give a matching service to filter out unsuitable live content. Alternatively, the large amount of live content has to cross a large amount of unrelated users, and results to a high data disseminating latency and bulky memory cost. Second, due to various sizes of live content and the bandwidth throughput comes to be a main issue to spreading of data. There is a huge disparity in upload and download capacities of the clients, the upload capacities become the barrier. To solve this problem, there are has been two solutions: One is the application layer multicast structure and other is the distribution/registration structure. For extent content data dissemination, the application layer multicast systems are eMule [4] and BitTorrent [5]. These both approaches lead to increasing data dissemination latency.

The distribution/registration system is used in data dissemination due to of its efficiency and flexibility in complex matching. In the distribution/registration systems, receivers (subscribers) registers as per their need or interests in the form of subscriptions and senders (publishers) sends (publish) the live content in the form of messages (events). So distribution/registration system matches with the events against the subscriptions and dispatch them to the interested candidates.

For extent content dissemination, another distribution/registration is designed, which is peer-assisted technique like Publicly [6] and PAPAS [7]. In these techniques, systems use a number of helpers to behave, such as trackers of BitTorrent. Clients who are interested in the live content can use the helpers and data dissemination done quickly. The scalability of this design is still not suitable for live content dissemination. It creates problem of scalability, high matching latency and the limited upload bandwidth of capacities of clients.

Recently, most of the distribution/ registration systems are designed such as Move [8], BlueDove [9], SEMAS [10], GSEC [11], GSEM [12], SETO [13] and SREM [14]. These all systems use the one-hop lookup scheme to get scalable event matching service. In these systems simply the events are sends to the interested clients. These systems approaches out heavy workloads on brokers because of data dissemination of bulk amount of content. These system approaches has drawbacks such as small matching latency high memory overhead, traffic overhead, scalability, consistency get high routing latency and well-timed report to the clients about the dissemination of live content. In the distribution/registration systems, each subscriptions represents a subspace of the complete content space and the different clients with the same matching content overlapping the interests along a high possibility as the number of subscriptions increase. This shows that each client may join a large number of groups, which gives a large memory overhead and traffic overhead.

In this paper, we a cloud aided content distribution and registration service, called CADAR. We mainly target the two problems: one is how to well arrange brokers for reaching scalable and reliable routing service in the cloud computing atmosphere and the other is how to disseminate large live content with a variety of sizes in a scalable and reliable behaviour.

## 2. RELATED WORK

A large number of content-based publish/ subscribe are designed for narrow-volume content. In this paper Siena [15] broadcasts each subscription to brokers in a tree-based overlay, and each broker reserves subscriptions that come from their neighbour brokers. Broadcasting subscriptions in Siena brings large memory overhead and matching latency. After this in Gryphon [16] builds and prunes a shortest path spanning tree rooted at the publisher's node to only keep the nodes with subscriptions matching the events. To compute the dynamic spanning tree, each node in Gryphon buffers a complete description of the nodes and subscriptions, which leads to large traffic overhead. In this paper, Meghdoot [17] uses the CAN [18] overlay to map

subscriptions to a 2k-Euclidean space, where k is the number of dimensions of the content space. It doesn't provide elasticity among servers. Now we have recommended Pastry Strings [19] uses single partition or all partition schemes. It provides content-based event filtering and routing. It doesn't provide elasticity among servers. In Terpstra et al [20] propose to build a routed spanning tree for each publisher based on Chord overlay [21]. In this the bandwidth of the brokers may give result in load imbalance among the brokers that is unfriendly to the event distribution and filtration. BlueDove has proposed an approach in which it uses a multidimensional space partitioning technique to split subscriptions to help the attribute based pub/sub model among servers in the cloud environment. The drawback of the scheme is the failure detection delay and it loses a few messages after the failure of the server. After this, SEMAS uses a hierarchical multi-attribute space partition technique to split the subscriptions and to achieve high matching throughput. The drawback of the scheme is the as the number of content and segment increases, the scheme obtain memory overhead. Several schemes and approaches followed by one by one GSEM uses a hybrid content space partitioning scheme to split large subscriptions into the multiple clusters. It provides low matching latency among users. In GSEC uses a hybrid space portioning technique to split large subscriptions into multiple clusters in a hierarchical manner. It provides low matching latency and leads to a high latency for disseminating extent content.

A large number of content-based publish/subscribe are designed for extent content. Publiy propose hybrid 2-layer architecture to organize both brokers and clients. Its total upload bandwidth is limited by the uplink capacities of the clients, which makes it inappropriate to time-critical data dissemination. In PAPaS system, a centralized broker system and a decentralized gossip-based P2P protocol are used for event matching and distribution, respectively. The spreading latency of the gossip-based P2P protocol in this system is impacted by the limited upload capacities of users.

In SREM, it proposes a distributed overlay Skipcloud to achieve small routing latency and reliable links. It uses HPartiton scheme for the bulk amount of content and the scheme splits the data into subspaces.

## 3. PROPOSED SCHEME

The proposed cloud aided content distribution and registration service can comfortably check out and then separate to refine the irrelevant users from the large amount of data. In this proposed system, we propose a distributed overlay, to arrange brokers at the cloud computing environment. Distributed overlay activates the subscriptions to be sent between brokers in a reliable and scalable at real-time manner. It becomes easy for implementing and maintaining. Moreover, we propose a hybrid space partitioning technique called HPartition-8 for achieving reliability and scalability for event matching between multiple brokers. This HPartition-8 scheme allows the same type of subscriptions to be partitioned into the multiple small clusters and it leads to multiple clients matching brokers for each event. It alleviates the hotspots and keeping the workload balanced for all the brokers.

Through this proposed system the problems of the existing system will be recovered. This system provides the facility or solution for disseminating data with various sizes for interested clients in reliable and scalable at real-time manner. The proposed system uses a cloud-aided technique for time-critical dissemination of extent content.

### *Advantages*

We propose a hybrid multidimensional space partitioning technique, called HPartition-8 for partitioning the large scale subscriptions into the multiple small clusters of subscription. It overcomes the problem of low matching latency.

The distribution of data with various sizes for interested clients in a reliable and scalable at real-time manner.

It solves the problem for time-critical dissemination of live extent content.

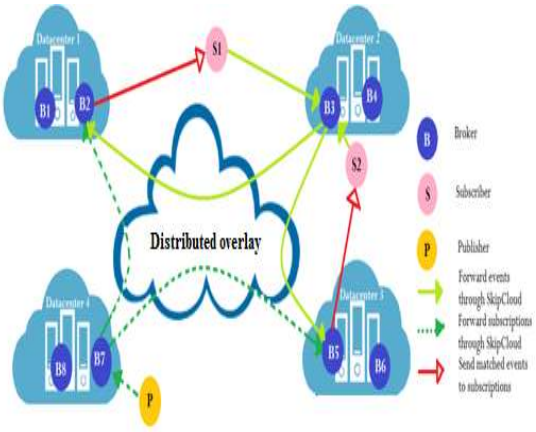


Figure 1. System Architecture

## 3. OVERVIEW OF CADAR

We are considering a cloud computing environment with an arranged geological distributed cloud datacenters by way of the Internet. In a cloud computing environment, each datacenter includes a big number of brokers. These brokers or servers are guided by a datacenter management service like Amazon EC2.

We represent system architecture for CADAR which is shown in Figure1 all the brokers in CADAR are made public to the Internet, so that any of the subscriber and publisher can directly associate with them for achieving reliable connectivity and low routing latency. These brokers of datacenters are connected by a distributed overlay. The complete content space is divided into the multiple subspaces; each subspace is guided through the brokers. The subscriptions and events are forwarded to those subspaces which are overlapping with them by the way of distributed overlay. Hence, the subscriptions and events falling at the same subspace that are matched in the same broker. After the completion of the matching procedure, events or messages are broadcasted to the interested users or subscribers. In Figure2, all the subscriptions produced by the subscribers S1 and S2 that are forwarded to brokers B2 and B5 in a respective way. After receiving events which is published by the publishers, brokers B2 and B5 will send matched events to the S1 and S2 in a respective way.

One may try to convince that the dependable for some of subset of the subscription because of the geological location, then we do not need a lot of cooperation among the brokers [15]. After all distribution/ registration system wants to catch all matched subscribers. It needs each of the event should be matched in all datacenters, that leads to higher traffic overhead as the number of datacenters increases and the high arrival rate of large-scale live content. Apart from, it is difficult to achieve workloads balanced among brokers of the datacenters because of the diverse data distributions of candidate's interests. We need a distributed overlay because with the increasing number of servers in all the datacenters, the failure of the node comes to be normal but not exceptional. This node failure creates unreliability and inefficient routing among brokers. For solving this problem, we attempt to arrange brokers in Distributed overlay for reducing the routing latency in a reliable and scalable way.

The proposed scheme shows a many benefits for reliable data dissemination at real time. Firstly, it permits the system to group up similar subscriptions at the right time into the same broker because of the large bandwidth among brokers, so that the local searching time may be reduced. Then it becomes urgently important for reaching to high matching throughput. Secondly, as we have mentioned above that each subspace is guided by the multiple brokers. This scheme is fault-tolerant even if a large number of brokers fail spontaneously. And thirdly, due to the datacenter management service supplies scalable and elastic brokers and then the system can be easily extended to the large internet-scale.

***Hybrid Content Space Partitioning Technique***

In this, we propose HPartition-8, a hybrid content space partitioning technique to improve matching candidates. To take benefit of multiple distributed brokers in cloud computing environment, CADAR divides the whole content space of subscriptions among the top clusters of Distributed overlay, due to this each cluster examines a subset of the whole content space and seeking for a small number of candidate subscriptions. CADAR apply HPartition-8 to achieve reliable and scalable content-based event matching at real-time manner. Our survey analysis speaking, HPartition-8 divides the whole content space into the multiple disjoint subspaces. After that, the subscriptions and events with the overlapping subspaces are forwarded and matched at the same top cluster of distributed overlay network. In addition, to maintain workload balance among servers and to improve matching throughput, HPartition-8 divides the hot spots into multiple cold spots in an adaptive way.

## 4. CONCLUSION AND FUTURE WORK

This paper introduces CADAR, a cloud aided content distribution and registration service. CADAR connects the brokers through a proposed distributed overlay, which guarantee scalable and reliable connectivity between brokers and it arrange brokers of CADAR, to obtain low routing latency. To obtain high matching latency and multiple candidate servers for each event, we introduce a hybrid space partitioning technique HPartition-8. Performance evaluation produces results which determine that CADAR is effective and practical, and also presents good workload balance, scalability and reliability under various specification settings.

In recent years, data flow processing systems have become the prevailing concept to process massive data. Since data flows are probably free in size and processed in order, it leads big questions to real-time data dissemination. In future work, we will study how to use our proposed framework to design a time-critical pub/sub system for large-scale data flow.

## REFERENCES:

[1] Readywarn. http://www.readywarn.com/; 2016. Accessed February 10, 2016.

[2] Ganti RK, Pham N, Ahmadi H, Nangia S, Abdelzaher TF, "Greengps: a participatory sensing fuel-efficient maps application", Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, ACM, Breckenridge, CO, USA, pp. 151–164, 2010.

[3] Shakecast. http://earthquake.usgs.gov/research/software/shakecast/; 2016. Accessed January 10, 2016.

[4] Kulbak Y, Bickson D, "The eMule protocol specification", eMule project, https://www.gnunet.org/sites/default/files/10.1.1.60.7750_0.pdf 2005. Accessed March 1, 2016

[5] Cohen B, "Incentives build robustness in BitTorrent", Workshop on Econ Peer-to-Peer Syst., pp. 68–72, 2003.

[6] Kazemzadeh RS, Jacobsen H-A, "Publiy+: a peer-assisted publish/subscribe service for timely dissemination of bulk content", ICDCS, Macau, China, pp. 345–354, 2012.

[7] Ahmed N, Linderman M, Bryant J, "Papas: peer assisted publish and subscribe", Proceedings of the 7th Workshop on Middleware for Next Generation Internet Computing, ACM, Montreal, QC, Canada; Vol. 7, 2012.

[8] W. Rao, L. Chen, P. Hui, and S. Tarkoma, "Move: A large scale

keyword-based content filtering and dissemination system", in Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst., pp. 445–454, 2012.

[9] Li M, Ye F, Kim M, Chen H, Lei H, "A scalable and elastic publish/ subscribe service", IPDPS, Alaska, USA, pp. 1254–1265, 2011.

[10] Ma X, Wang Y, Qiu Q, Sun W, Pei X, "Scalable and elastic event matching for attribute-based publish/ subscribe systems", Future Gener Comput Syst., Vol. 36, pp. 102–119, 2014.

[11] Wang Y, Ma X, "A general scalable and elastic content-based publish/ subscribe service", IEEE Trans Parallel Distrib Syst., Vol. 26, Issue 8, pp. 2100–2113, 2015.

[12] Ma X, Wang Y, Pei X, Li X, "A general scalable and elastic matching service for content-based publish/ subscribe systems", Concurr Comput Pract Experience, Vol. 27, Issue 1, pp. 94–118, 2015.

[13] X. Ma et al., "Scalable and elastic total order in content-based publish/ subscribe systems", Future Gener. Comput. Syst., Vol. 83, Issue 7, pp. 297–314, 2015.

[14] Ma X, Wang Y, Pei X, "A scalable and reliable matching service for content-based publish/subscribe systems", IEEE Trans Cloud Comput., Vol. 3, Issue 1, pp. 1–13, 2015.

[15] Carzaniga A, "Architectures for an event notification service scalable to wide-area networks", Ph.D. Thesis, 1998.

[16] Banavar G, Chandra TD, Mukherjee B, Nagarajarao J, Strom RE, Sturman DC, "An efficient multicast protocol for content-based publish-subscribe systems", ICDCS, Austin, TX, USA, pp. 262–272, 1999.

[17] Gupta A, Sahin OD, Agrawal D, El Abbadi A, "Meghdoot: content-based publish/subscribe over P2P networks", Middleware, Toronto, Ontario, Canada, pp. 254–273, 2004.

[18] Ratnasamy S, Handley M, Karp RM, Shenker Shalini, "Application-level multi-cast using content-addressable networks", pp. 14–29, 2001.

[19] Aekaterinidis I, Triantafillou P, "Pastrystrings: a comprehensive content-based publish/subscribe DHT network", ICDCS, Lisboa, Portugal, Vol. 23, 2006.

[20] Terpstra WW, Behnel S, Fiege L, Zeidler A, Buchmann AP, "A peer-to -peer approach to content-based publish/subscribe", DEBS, San Diego, CA, USA, pp. 1–8, 2003.

[21] Stoica I, Morris R, Karger DR, Kaashoek MF, Balakrishnan H, "Chord: a scalable peer-to-peer lookup service for internet applications", SIG-COMM, San Diego, CA, USA, pp. 149–160, 2001.