



A Review on Event Matching Service for Content-Based Publish/Subscribe Systems

Prof. Sapna Jain Choudhary , Bhawna Kol

Assistant Professor, Dept. of CSE, Shri Ram Group of Institutions, Jabalpur, MP, India

M.Tech Scholar, Dept. Of CSE, Shri Ram Group of Institutions, Jabalpur, MP, India

ABSTRACT: Internet is changing the world of distributed computing. For making a loosely coupled infrastructure, peer-to-peer communication system preparing system more complex in distributed environment. This loosely coupled infrastructure can be provided by Publish/Subscribe system. For establishing a reliable and scalable publish/subscribe system is demanding task to perform in cloud computing environment. The publish/subscribe system is kind of a messaging system which has three elements publisher, subscriber and event notification system. In Publish/subscribe system, publishers send the information and subscribers receive the information based on their subscriptions. Publishers and subscribers are loosely coupled and they both are aware of each other. Publishers do not know about the subscribers and likewise, subscribers do not know about the publishers, the event notification system maintains the communication between publisher and subscriber. This paper introduces a survey about the approaches used in the publish/subscribe environment. The main aim of this paper is to examine how these approaches works and what are their effects on the system after using it.

KEYWORDS: Cloud computing, data dissemination, publish/subscribe

I. INTRODUCTION

Today millions of users are subscribing to information content over a cloud computing environment through applying filtering rules. The web applications, like Instagram, Facebook, Twitter and RSS feeds, enables users to apply their interests and subscribe to information provided by information producer (publishers). Figure 1 shows overview of Publish/Subscribe system. Publish/Subscribe (pub/sub) system, builds a connection between publishers and subscribers of events, acting as a middleware between publishers and subscribers. The publishers and subscribers decoupled from each other, they do not trust on each other. Publishers (information producers) submit events to the pub/sub system which is responsible for notifying the interested subscribers (information consumers). Subscribers make subscriptions for the events they are interested in to the pub/sub system. The aim of the pub/sub system matching is to notify subscribers about matching events quickly and efficiently. In these systems, the matching element becomes the obstacle of pub/ sub system. The pub/sub system is used in data dissemination due to of its efficiency and flexibility in complex matching.

In the given Figure 1, the publish/subscribe system has three main elements. First is the publishers are the information producers means who produces the information for requested users or maybe we can say consumers. Second is the subscribers are the information consumers means who consumes the information which he was requested from the event notification system. And third is the event notification system which acts like a filter between publisher and subscriber, this event notification system contains the brokers. In this publish/subscribe system, the publishers does

not need to know who are consuming the information (subscribers) and the subscribers does not need to know who is providing the information (publishers) that is the collecting the results of the subscriptions. Information are sent from the publishers to the event notification system (brokers), subscriptions are sent from the subscribers to the event notification system (brokers) and the event notification systems which contains brokers they forwards the information

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 6, June 2017

to the relevant subscribers. The brokers in the event notification system make secure that the information distributes to the relevant subscribers.

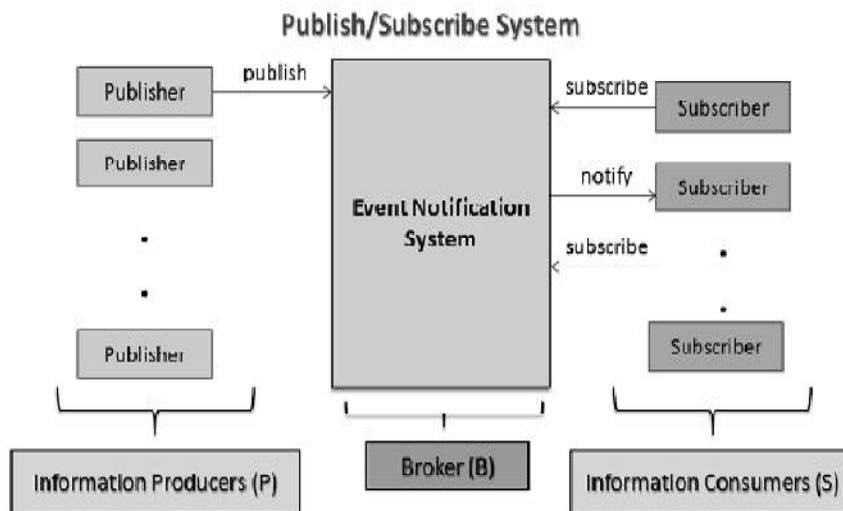


Figure1. Publish/Subscribe System

For extent content dissemination, another publish/subscribe is designed, which is peer-assisted technique like Publiy [6] and PAPA_S [7]. In these techniques, systems uses a number of helpers where clients (i.e. subscribers) 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 publish/subscribe system are designed such as Move [19], BlueDove [8], SEMAS [9], GSEC [12], GSEM [10], SETO [18], SREM [11] and CAPS [12]. These all systems use the one-hop lookup scheme to achieve scalable event matching service. In this system 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 low matching latency high memory overhead, traffic overhead, scalability, reliability achieve high routing latency and timely inform the clients about the dissemination of live content. In the publish/subscribe systems, each subscriptions represents a subspace of the entire content space and the different clients with the same matching content overlapping the interests along a high probability 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.

II. RELATED WORK

In this section, we give a brief introduction to the existing content-based pub/sub systems for time-critical data dissemination. According to the data volume of events, current content-based publish/subscribe can be categorized into 2 classes: publish/subscribe for narrow-volume content and publish/subscribe for extent content.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

A. *Publish/subscribe for narrow-volume content*

A large number of content-based publish/subscribe are designed for narrow-volume content. In this paper Siena [1] broadcasts each subscription to brokers in a tree-based overlay, and each broker reserves subscriptions that come from their neighbor brokers. Broadcasting subscriptions in Siena brings large memory overhead and matching latency. After this in Gryphon [2] 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 [3] uses the CAN [14] 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 [4] 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 [5] propose to build a routed spanning tree for each publisher based on Chord overlay [16]. 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 [8] has proposed a 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 lose a few messages after the failure of the server. After this, SEMAS [9] 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 [10] uses a hybrid content space partitioning scheme to split large subscriptions into the multiple clusters. It provides low matching latency among users. In GSEC [12] 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.

B. *Publish/subscribe for extent content*

A large number of content-based publish/subscribe are designed for extent content. The peer-assisted technique was first introduced by Wong [17] to distribute extent content. The peer-assisted technique uses the free upload capacities of uninterested nodes know as helpers to enhance the total upload bandwidth. Publiy [6] 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 this, it may lead to high routing latency with the increasing number of brokers. In PAPA_S [7] 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 SETO [18], it uses a two-layer pub/sub framework for decoupling the event matching service and total order service. Preceding graph techniques was proposed for scalability and a performance aware provisioning technique was proposed for elasticity. In SREM [11], it proposes a distributed overlay Skipcloud to achieve low routing latency and reliable links. It uses HPartition scheme for the bulk amount of content and the scheme splits the data into subspaces.

C. *Reviews of all the existing schemes.*

According to our survey analysis over proposed schemes, we have found that review about the existing schemes. There are drawbacks of each of the existing approach. In Table1, we have discussed about the proposed schemes. The reviews of each and every proposed scheme are mentioned in Table1.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 6, June 2017

Table1. Reviews of proposed schemes

S.NO.	PROPOSED SCHEME	REVIEWS
1.	BLUEDOVE	It is a Cloud-based pub/subs system. It uses one-hop overlay. It uses single partition or all partition. The drawback of the scheme was because of the delay in the failure detection, the system lose some messages after the failure of the server.
2.	MOVE	It uses an optimal filter allocation approach for scalable match. The drawback of the scheme was the approach is scalable only for the limited servers.
3.	GSEC	It uses the two- layer overlay, one-hop overlay and root-disjoint tree overlay. It uses the helper-based content distribution approach. It achieves scalability performance as the number of system increases. It doesn't provide reliability.
4.	GSEM	It uses one-hop overlay. It uses hybrid content space partitioning technique. It achieves scalability matching throughput as the scale of server increases. Elasticity of the servers maintains. But they didn't work on reliability of the servers.
5.	SEMAS	It uses one-hop lookup overlay. This scheme for attribute-based pub/sub systems. The drawback of the scheme was the memory overhead increases as the number of attributes increases in the system and not for the more complicated content based pub/sub systems.
6.	SETO	It uses a two-layer pub/sub framework. This scheme supports content-based pub/sub systems. It uses the preceding graph building technique. This scheme not for the casual order in content-based pub/sub systems.
7.	SREM	It uses a distributed overlay. This scheme for the scalability and reliability of the servers. it doesn't provide elastic resource provisioning strategies for obtaining good performance ratio and not guarantee that the brokers spread the large-scale live content with diverse sizes.
8.	CAPS	They proposed a helper-based content distribution for improving the upload bandwidth for spreading of data. The drawback of this paper was not supportable for large-scale data stream.

III. CONCLUSION

This paper introduces a survey about the approaches used in the publish/subscribe environment. The main aim of this paper is to examine how these approaches works and what are their effects on the system after using it. There are various approaches for content-based publish/subscribe systems for achieving reliability, scalability and elasticity. We analyze the approaches detect the limitations and ignore them. We will propose a solution to overcome the problem the existing publish/subscribe system faces.

REFERENCES

1. Carzaniga A, "Architectures for an event notification service scalable to wide-area networks", Ph.D, Thesis, 1998.
2. 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, Vol. No, Issue no., pp. 262-272, 1999.
3. Gupta A, Sahin OD, Agrawal D, El Abbadi A, "Meghdoot: content-based publish/subscribe over P2P networks" , Middleware, Toronto, Ontario, Canada, Vol. No, Issue no., pp. 254-273, 2004.
4. Aekaterinidis I, Triantafillou P, "Pastrystrings: a comprehensive content-based publish/subscribe DHT network" , ICDCS, Lisboa, Portugal, Vol. 23, Issue 4., 2006.
5. 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, Vol. 33, Issue 9., pp. 1-8, 2003.
6. Kazemzadeh RS, Jacobsen H-A, "Publi+: a peer-assisted publish/subscribe service for timely dissemination of extent content" , ICDCS, Macau, China, Vol. 12, Issue 3., pp. 345-354, 2012.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

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, Issue 2., 2012.
8. Li M, Ye F, Kim M, Chen H, Lei H, "A scalable and elastic publish/subscribe service" , IEEE IPDPS, Alaska, USA, Vol. 20, Issue 5., pp. 1254–1265, 2011.
9. 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, Issue 3., pp. 102–119, 2014.
10. 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.
11. 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.
12. Ma X, Wang Y, Pei X, Xu F, "A cloud-assisted publish/subscribe service for time-critical dissemination of bulk content" , Concurr Comput Pract Experience, Vol. 29, Issue 8., 2016.
13. 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.
14. Lakshman A, Malik P, "Cassandra: a decentralized structured storage system" , Operating Syst Rev., Vol. 44, Issue 2., pp. 35-40, 2010.
15. Ratnasamy S, Handley M, Karp RM, Shenker S, "Application-level multicast using content-addressable networks" , Proceedings of the Third International COST264 Workshop on Networked Group Communication, Vol. 30, Issue 7., pp. 14-29, 2001.
16. Rowstron AIT, Druschel P, "Pastry: scalable, decentralized object location, and routing for large-scale peer-to-peer systems" , Middleware, Heidelberg, Germany, Vol. 2218, Issue 10., pp. 329-350, 2001.
17. Stoica I, Morris R, Karger DR, Kaashoek MF, Balakrishnan H, "Chord: a scalable peer-to-peer lookup service for internet applications" , SIGCOMM, San Diego, CA, USA, Vol. 11, Issue 1., pp.149-160, 2001.
18. Wong JHT, "Enhancing collaborative content delivery with helpers" , Ph.D, Thesis, 2004.
19. X. Ma et al., "Scalable and elastic total order in content-based publish/subscribe systems" , Vol. 83, Issue 7., pp. 297–314, 2015.
20. W. Rao, L. Chen, P. Hui, and S. Tarkoma, "Move: A large scale keyword-based content filtering and dissemination system" , Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst., Vol. 33, Issue 12., pp. 445-454, 2012.

BIOGRAPHY

Prof. Sapna Choudhary is an associate professor of Department of Computer Science and Engineering at Shri Ram Group of Institutions, Jabalpur, MP, India.

Bhawna Kol is a M.tech scholar in Department of Computer Science and Engineering, at Shri Ram Group of Institutions, Jabalpur, MP, India. She is pursuing Master of Technology (M.Tech) degree since 2014. Her research interests are Cloud Computing, Computer Networks etc.