ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Content Based Publish/Subscribe Matching Service with Complex Computing and Reliable Communication

Kalyani Yele
M.Tech,
Dept of CSE,
Teegala Krishna Reddy Engineering College.

Ravinder Reddy Ch
Assistant Professor,
Dept of CSE,
Teegala Krishna Reddy Engineering College.

Abstract:

Characterized by the increasing arrival rate of live content, the emergency applications pose a great challenge: how to disseminate large-scale live content to interested users in a scalable and reliable manner. The publish/subscribe (pub/sub) model is widely used for data dissemination because of its capacity of seamlessly expanding the system to massive size. However, most event matching services of existing pub/sub systems either lead to low matching throughput when matching a large number of skewed subscriptions, or interrupt dissemination when a large number of servers fail. The cloud computing provides great opportunities for the requirements of complex computing and reliable communication. In this paper, we propose SREM, a scalable and reliable event matching service for content-based pub/sub systems in cloud computing environment.

To achieve low routing latency and reliable links among servers, we propose a distributed overlay Skip Cloud to organize servers of SREM. Through a hybrid space partitioning technique HPartition, large-scale skewed subscriptions are mapped into multiple subspaces, which ensures high matching throughput and provides multiple candidate servers for each event. Moreover, a series of dynamics maintenance mechanisms are extensively studied. To evaluate the performance of SREM, 64 servers are deployed and millions of live content items are tested in a CloudStack testbed. Under various parameter settings, the experimental results demonstrate that the traffic overhead of routing events in SkipCloud is at least 60 percent smaller than in Chord overlay, the matching rate in SREM is at least 3.7 times and at most 40.4 times larger than the single-dimensional partitioning technique of BlueDove. Besides, SREM enables the event loss rate to drop back to 0 in tens of seconds even if a large number of servers fail simultaneously.

Keywords:

Publish/subscribe, event matching, overlay construction, content space partitioning, cloud computing.

Introduction:

Common requirement for any system is security. The need for security must be extremely high. It is one of the major requirements to protect or control any sort of failures. There are number of mechanisms which are available to provide security. In that one of the most important mechanisms is encryption. In cryptography encryption is the process of converting plain text to cipher text which is unreadable from unauthorized users. The cryptography mechanism is required in publish/subscribe system. In publish/subscribe system publisher is one who publishes his content without specifying a particular destination to reach publisher will not program the documents to be delivered to a particular subscriber. Publisher will classify publishing documents based on different criteria and release it and subscriber will show interest on one or more documents and subscribe to that particular one in order to have access over it. This publish/subscribe system is traditionally carried out in broker-less [12] content based routing which forwards or routes the message based on the content of the message instead of clearly routing to an

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

specified destination. Content based routing applies some set of rules to Its content to find the users who are interested in its content. Its different nature is helpful for huge-level scattered applications and also provides a high range of flexibility and adaptability to change. Authorized publisher have permission to publish events in the network and similarly subscribers who likes the content can gets subscribed to a particular published content and have access over it by which high level access control [7] can be achieved. Here published content should not be exposed to routing infrastructure and subscribers should receive content without leaking subscription identity to the system, which is a highly challenging task which needs to be carried out in content-based pub/sub system. Publisher and subscriber are the two entities and they do not trust each other. Even though authorized publisher publish events, nasty publisher pretend to be the real publisher and may spam the network with fake and duplicate contents similarly subscribers are very much eager to find other users and publishers which are challenging tasks.

Finally, Transport Layer Security (TLS) or Secure Socket Layer (SSL) is secure channels for distributing keys from key server to the required. Existing security approach deals with traditional network and security is based on restricted manner which tells about key word matching [8]. Key management was the challenging task in the existing approach, so to overcome all these, we use new approach called pairing-based cryptography mechanism, which helps in mapping between to end parties so called cryptographic groups. Here, Identity Based Encryption Technique (IBE) [9] is used under this mechanism. New approach IBE provide greater concern towards authentication and confidentiality in the network. Our approach permit users to preserve credentials based on their subscriptions. Secret keys provided to the users are labeled with the credentials. In Identity-based encryption (IBE) mechanisms

1) key can be used to decrypt only if there is match between credentials with the content and the key; and 2) to permit subscribers to check the validity of received contents.

Moreover, this approach helps in providing fine-grained key management, effective encryption, decryption operations and routing is carried out in the order of subscribed attributes.

Related work:

There are two entities in the System publishers and subscribers. Both the entities are computationally bounded and do not trust each other. Moreover, all the peers (publishers or subscribers) participating in the pub/sub overlay network are honest and do not deviate from the designed protocol. Likewise, authorized publishers only allow valid events in the system. However, malicious publishers may masquerade the authorized publishers and spam the overlay network with fake and duplicate events. We do not intend to solve the digital copyright problem; therefore, authorized subscribers do not reveal the content of successfully decrypted events to other subscribers.

A. Publisher subscriber technique:

Publishers and subscribers interact with a key server. They provide credentials to the key server and in turn receive keys which fit the expressed capabilities in the credentials. Subsequently, those keys can be used to encrypt, decrypt, and sign relevant messages in the content based pub/sub system, i.e., the credential becomes authorized by the key server. A credential consists of two parts: 1) a binary string which describes the capability of a peer in publishing and receiving events, and 2) a proof of its identity [1].

B. Identity based encryption Identity(ID):

based public key cryptosystem, which enables any pair of users to communicate securely without exchanging public key certificates, without keeping a public key directory, and without using online service of a third party, as long as a trusted key generation center issues a private key to each user when he first joins the network [2].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

C. Identity Handling:

Identification provides an essential building block for a large number of services and functionalities in distributed Information systems. In its simplest form, identification Is used to uniquely denote computers on the Internet By IP addresses in combination with the Domain Name System (DNS) as a mapping service between symbolic Names and IP addresses. Thus, computers can conveniently Be referred to by their symbolic names, whereas, in The routing process, their IP addresses must be used.[3] Higher-level directories, such as X.500/LDAP, consistently Map properties to objects which are uniquely identified by Their distinguished name (DN), i.e., their position in the X.500 tree [4].

D. Content based publish/subscribe:

Content-based networking is a generali- zation of the content based publish/subscribe model. [4] In content-based networking, messages are no longer addressed to the communication end-points . Instead, they are published to a distributed information space and routed by the networking sub -strate to the "interested" communication end-points. In most cases, the same substrate is responsible for realizing naming, binding and the actual content delivery [5].

E. Secure Key Exchange:

A key-exchange (KE) protocol is run in a network of interconnected parties where each party can be activated to run an instance of the protocol called a session [6]. Within a session a party can be activated to initiate the session or to respond to an incoming message. As a result of these activations, and according to the specification of the protocol, the party creates and maintains a session state, generates outgoing messages, and eventually completes the session by outputting a session-key and erasing the session state [7].

EXISTING SYSTEM:

- In traditional data dissemination applications, the live content are generated by publishers at a low speed, which makes many pub/subs adopt the multi-hop routing techniques to disseminate events.
- A large body of broker-based pub/subs forward events and subscriptions through organizing nodes into diverse distributed overlays, such as tree based design, cluster-based design and DHT-based design.

DISADVANTAGES OF EXISTING SYSTEM:

- The system cannot scalable to support the large amount of live content.
- The Multihop routing techniques in these broker-based systems lead to a low matching throughput, which is inadequate to apply to current high arrival rate of live content.
- Most of them are inappropriate to the matching of live content with high data dimensionality due to the limitation of their subscription space partitioning techniques, which bring either low matching throughput or high memory overhead.

PROPOSED SYSTEM:

- Specifically, we mainly focus on two problems: one is how to organize servers in the cloud computing environment to achieve scalable and reliable routing. The other is how to manage subscriptions and events to achieve parallel matching among these servers.
- We propose a distributed overlay protocol, called SkipCloud, to organize servers in the cloud computing environment. SkipCloud enables subscriptions and events to be forwarded among brokers in a scalable and reliable manner. Also it is easy to implement and maintain.
- To achieve scalable and reliable event matching among multiple servers, we propose a hybrid multidimensional space partitioning technique, called HPartition. It allows similar subscriptions to be divided into the same server and provides multiple candidate

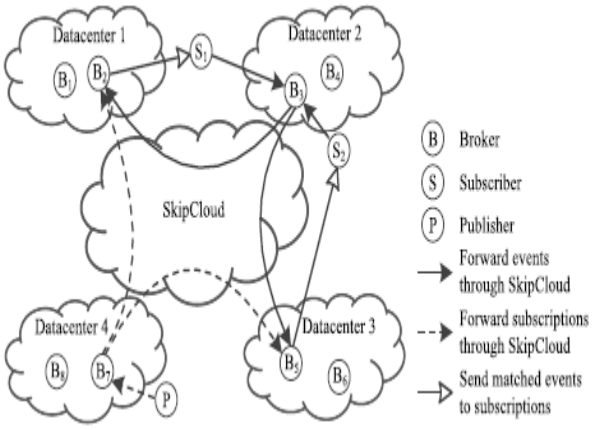
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

matching servers for each event. Moreover, it adaptively alleviates hot spots and keeps workload balance among all servers.

ADVANTAGES OF PROPOSED SYSTEM:

- ✓ We propose a scalable and reliable matching service for content-based pub/sub service in cloud computing environments, called SREM.
- ✓ We propose a hybrid multidimensional space partitioning technique, called HPartition SSPartition.
- ✓ To alleviate the hot spots whose subscriptions fall into a narrow space, we propose a subscription set partitioning,
- ✓ Through a hybrid multi-dimensional space partitioning technique, SREM reaches scalable and balanced clustering of high dimensional skewed subscriptions

SYSTEM ARCHITECTURE:



MODULES DESCSRIPTION:

Datacenter / Broker creation:

In the first module, we develop the Datacenter creation and Broker Creation. To support large-scale users, we consider a cloud computing environment with a set of geographically distributed datacenters. Each datacenter contains a large number of servers (brokers), which are managed by a datacenter management service. Our approach is suitable for large and reasonably stable environments such as that of an enterprise or a data center, where reliable publication delivery is desired in spite of failures. As future work, we would like to exploit our scheme to allow for multi-path load balancing, and support some of P/S optimization techniques such as subscription covering. It provides an abstract and high level interface for data producers (publishers) to publish messages and consumers (subscribers) to receive messages that match their interest.

Clustering Method:

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster. Suppose we are given a database of 'n' objects and the partitioning method constructs 'k' partition of data. Each partition will represent a cluster and $k \leq n$. It means that it will classify the data into k groups, which satisfy the following requirements:

- Each group contains at least one object.
- Each object must belong to exactly one group.

Content Space Partitioning:

The content space is partitioned into disjoint subspaces, each of which is managed by a number of brokers. Then each top cluster only handles a subset of the entire space and searches a small number of candidate subscriptions. The whole content space into non-overlapping zones based on the number of its brokers. After that, the brokers in different cliques who are responsible for similar zones are connected by a multicast tree.

Event Matching:

The data replication schemes are employed to ensure reliable event matching. For instance, it advertises subscriptions to the whole network. When receiving an event, each broker determines to forward the event to the corresponding broker according to its routing table. These approaches are inadequate to achieve scalable event matching.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Routing Method:

The routing process usually directs forwarding on the basis of routing tables, which maintain a record of the routes to various network destinations. Thus, constructing routing tables, which are held in the router's memory, is very important for efficient routing. Most routing algorithms use only one network path at a time. Multipath routing techniques enable the use of multiple alternative paths. Prefix routing in SkipCloud is mainly used to efficiently route subscriptions and events to the top clusters. Note that the cluster identifiers at level are generated by appending one b-ary to the corresponding clusters at level i. The relation of identifiers between clusters is the foundation of routing to target clusters. Briefly, when receiving a routing request to a specific cluster, a broker examines its neighbor lists of all levels and chooses the neighbor which shares the longest common prefix with the target ClusterID as the next hop. The routing operation repeats until a broker cannot find a neighbor whose identifier is more closer than itself.

CONCLUSION:

This paper introduces SREM, a scalable and reliable event matching service for content-based pub/sub systems in cloud computing environment. SREM connects the brokers through a distributed overlay SkipCloud, which ensures reliable connectivity among brokers through its multi-level clusters and brings a low routing latency through a prefix routing algorithm. Through a hybrid multi-dimensional space partitioning technique, SREM reaches scalable and balanced clustering of high dimensional skewed subscriptions, and each event is allowed to be matched on any of its candidate servers. Extensive experiments with real deployment based on a CloudStack testbed are conducted, producing results which demonstrate that SREM is effective and practical, and also presents good workload balance, scalability and reliability under various parameter settings.

Although our proposed event matching service can efficiently filter out irrelevant users from big data volume, there are still a number of problems we need to solve. Firstly, we do not provide elastic resource provisioning strategies in this paper to obtain a good performance price ratio. We plan to design and implement the elastic strategiesof adjusting the scale of servers based on the churn workloads. Secondly, it does not guarantee that the brokers disseminate large live content with various data sizes to the corresponding subscribers in a real-time manner. For the dissemination of bulk content, the upload capacity becomes the main bottleneck. Based on our proposed event matching service, we will consider utilizing a cloud-assisted technique to realize a general and scalable data dissemination service over live content with various data sizes.

REFERENCES:

[1]Dataperminite. (2014). [Online]. Available: http://www.domo. com/blog/2012/06/how-much-data-is-created-every-minute/

[2]F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing user behavior in online social networks,” in Proc. 9th ACM SIGCOMM Conf. Internet Meas. Conf., 2009, pp. 49–62.

[3]A. Carzaniga, “Architectures for an event notification service scalable to wide-area networks,” Ph.D. dissertation, Ingegneria Informatica e Automatica, Politecnico di Milano, Milan, Italy, 1998.

[4]P. Eugster and J. Stephen, “Universal cross-cloud communication,” IEEE Trans. Cloud Comput., vol. 2, no. 2, pp. 103–116, 2014. [5] R. S. Kazemzadeh and H.-A Jacobsen, “Reliable and highly available distributed publish/subscribe service,” in Proc. 28th IEEE Int. Symp. Reliable Distrib. Syst., 2009, pp. 41–50.

[6]Y. Zhao and J. Wu, “Building a reliable and high-performance content-based publish/subscribe system,”

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

J. Parallel Distrib. Comput., vol. 73, no. 4, pp. 371–382, 2013.

[7]F. Cao and J. P. Singh, "Efficient event routing in content-based publish/subscribe service network," in Proc. IEEE INFOCOM, 2004, pp. 929–940.

[8]S. Voulgaris, E. Riviere, A. Kermarrec, and M. Van Steen, "Sub-2- sub: Self-organizing content-based publish and subscribe for dynamic and large scale collaborative networks," Res. Rep. RR5772, INRIA, Rennes, France, 2005.

[9]I. Aekaterinidis and P. Triantafillou, "Pastrystrings: A comprehensive content-based publish/subscribe DHT network," in Proc. IEEE Int. Conf. Distrib. Comput. Syst., 2006, pp. 23–32.

[10]R. Ranjan, L. Chan, A. Harwood, S. Karunasekera, and R. Buyya, "Decentralised resource discovery service for large scale federated grids," in Proc. IEEE Int. Conf. e-Sci. Grid Comput., 2007, pp. 379– 387.

[11]A. Gupta, O. D. Sahin, D. Agrawal, and A. El Abbadi, "Meghdoot: Content-based publish/subscribe over p2p networks," in Proc. 5th ACM/IFIP/USENIX Int. Conf. Middleware, 2004, pp. 254–273.

[12]X. Lu, H. Wang, J. Wang, J. Xu, and D. Li, "Internet-based virtual computing environment: Beyond the data center as a computer," Future Gener. Comput. Syst., vol. 29, pp. 309–322, 2011.

[13] Y. Wang, X. Li, X. Li, and Y. Wang, "A survey of queries over uncertain data," Knowl. Inf. Syst., vol. 37, no. 3, pp. 485–530, 2013. [14] 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., 2012, pp. 445–454.

[15]M. Li, F. Ye, M. Kim, H. Chen, and H. Lei, "A scalable and elastic publish/subscribe service," in Proc. IEEE Int. Parallel Distrib. Process. Symp., 2011, pp. 1254–1265.

[16]X. Ma, Y. Wang, Q. Qiu, W. Sun, and X. Pei, "Scalable and elastic event matching for attribute-based publish/subscribe systems," Future Gener. Comput. Syst., vol. 36, pp. 102–119, 2013.

[17]A. Lakshman and P. Malik, "Cassandra: A decentralized structured storage system," Oper. Syst. Rev., vol. 44, no. 2, pp. 35–40, 2010.

[18]M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: Novel erasure codes for big data," in Proc. 39th Int. Conf. Very Large Data Bases, 2013, pp. 325–336.

[19]S. Voulgaris, D. Gavidia, and M. van Steen, "Cyclon: Inexpensive membership management for unstructured p2p overlays," J. Netw. Syst. Manage., vol. 13, no. 2, pp. 197–217, 2005.

[20]B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," IEEE J. Sel. Areas Commun., vol. 22, no. 1, pp. 41–53, Jan. 2004.

[21]Murmurhash. (2014). [Online]. Available: http://burtleburtle. net/bob/hash/doobs.html

[22]A.-M Kermarrec, L. Massoulie, and A. J. Ganesh, "Probabilistic reliable dissemination in large-scale systems," IEEE Trans. Parallel Distrib. Syst., vol. 14, no. 3, pp. 248–258, Mar. 2003.

[23]M. Jelasity, A. Montresor, and € O. Babaoglu, "Gossip-based aggregation in large dynamic networks," ACM Trans. Comput. Syst., vol. 23, no. 3, pp. 219–252, 2005.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[24]Cloudstack. (2014). [Online]. Available: (http://cloudstack. apache.org/)

[25]Zeroc. (2014). [Online]. Available: (http://www.zeroc.com/)

[26]B. He, D. Sun, and D. P. Agrawal, "Diffusion based distributed internet gateway load balancing in a wireless mesh network," in Proc. IEEE Global Telecommun. Conf., 2009, pp. 1–6.

[27]Y. Wang and S. Li, "Research and performance evaluation of data replication technology in distributed storage systems," Comput. Math. Appl., vol. 51, no. 11, pp. 1625–1632, 2006.