IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661 Volume 2, Issue 6 (July-Aug. 2012), PP 43-48
www.iosrjournals.org

An Approach to Continuous Queries in Unstructured Overlays (Coquos)

S.R.Srikanth, M.K.Madialagan
Asst. Prof, SVPCET

Abstract: *The current peer-to-peer (P2P) content distribution systems are running by their simple on-demand content discovery mechanism. The utility of these systems can be improved by incorporating two capabilities, first a mechanism through which peers can register their long term interests with the network so that they can be continuously notified of new data items, and second one is the peers to advertise their contents. Although researchers have proposed a few unstructured overlay-based publish-subscribe systems that provide the above capabilities, most of these systems require complex indexing and routing schemes, which not only make them highly complex but also gives less flexible on propagate the queries around transient peers.*

This paper argues that for many P2P applications, implementing full-fledged publish-subscribe systems is an complex task. For these applications, we study the alternate continuous query paradigm, which is a best-effort service providing the above two capabilities. we present a scalable and effective middleware, called CoQUOS stands for Continious queries in Unstructured Overlays for supporting continuous queries in unstructured overlay networks. Besides being independent of the overlay topology, CoQUOS preserves the simplicity and flexibility of the unstructured P2P network.

***Index Terms**: Peer-to-peer networks, Continuous queries, Publish-subscribe systems, Random walk.*

I. Introduction:

UNSTRUCTURED peer-to-peer (P2P)-based content/resource sharing platforms such as Gnutella and Kazaa have experienced excellent growths in the past days. The popularity of unstructured P2P networks can be based on the simplicity of their designs and their flexibility towards transient node population. Searching in these networks is essentially performed by circulating query messages in an ad-hoc manner and probing individual peer nodes.

The current unstructured P2P content distribution systems suffer from certain serious limitations. One such limitation is their simple-on demand mechanism for content discovery. Peers in these systems discover data items by circulating queries within the overlay network. A peer receiving a query responds back to the initiating node if it has any matching content. Upon processing a query, the recipient node removes it from its local buffers1. Thus, a query *expires* after it completes its circulation within the network. In other words, the network *forgets* the queries once they have completed their circulation. For clarity purpose called this as a ad hock query model , and we refer to the queries as *ad hoc queries*.

The ad hoc query model suffers from **two** main short comings **First,** an ad hoc query is only capable of searching and retrieving content that exists in the P2P network at the time the query was issued. In this scenario, the only way for a peer to discover newly added data items would be to repeatedly issue the same query. This is not desirable, since it creates unnecessary traffic within the network. **Second**, P2P systems that are purely based on the ad hoc query -model provide no support for peers to advertise or announce the data items they own to other interested peers. Advertisements are important for P2P applications where peers advertise their content. The above are shortcomings render the ad hoc query model .The many advanced P2P applications suffered from above two drawbacks. For handling the above situation we use a naive approach for tackling this problem would be to send advertisements to large subsets of peers through Flooding. However, this approach is unviable.

When heavy messaging overheads, this scheme could gives the peers with unwanted advertisements. To address the above limitations we use well-studied paradigm of publish-subscribe (pub-sub) Systems is a possible approach to address the above limitations. The pub-sub involves intricate indexing and routing mechanisms . To overcome the above problem we present a scalable and effective middleware, called CoQUOS stands for Continious queries in Unstructured Overlays for supporting continuous queries in unstructured overlay.

1.1.) Paper Contributions:

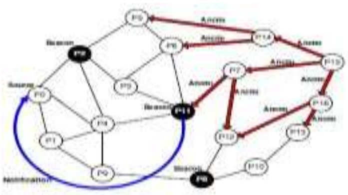
The diffculties in implementing pub-sub systems on top of unstructured overlays can be attributed to the inherent mismatch between the design requirements of pub-sub systems and the very nature of unstructured P2P systems. *The pub-sub model provides weaker guarantees and* individual implementations of may not provide and notification guarantees are weaker due to system failures or churn.

In this paper, we focus on an alternate notification paradigm called the *continuous query model*. Similar to content-based pub-sub systems this model provides a mechanism through which peers can register their queries, which are maintained in the network for extended durations of time. In this paper ,we develop lightweight middleware called *CoQUOS* (**co**ntinous **q**ueries in **u**nstructured **o**verlay**s**) for supporting continuous queries and advertisements in unstructured P2P networks. One of our goals in designing the CoQUOS system has been to preserve the design simplicity of the underlying unstructured P2P networks, and their flexibility towards network churn. In this regard, the design of the CoQUOS system exhibits two unique features. First, the CoQUOS system does not impose any topological constraints on the underlying P2P network, and it can be implemented as an independent module in any unstructured overlay network. Second, the CoQUOS system does not require complex index structures or routing mechanisms. Instead, our design is based upon very lightweight P2P primitive.

The fundamental idea of the CoQUOS system is to register the continuous query on a set of peers that are located in various topological regions of the overlay network.Towards addressing these challenges, this paper makes four novel contributions.

- **First**, we present a novel query propagation technique called *Cluster Resilient Random Walk (CRW).* CRW favours neighbors that more likely-
  to send messages deeper into the network there
  by enabling the continuous queries to reach differ-
  rent topological regions of the overlay network.
- **Second**, a *dynamic probability scheme* is proposed for enabling the recipients of a continuous query to make independent decisions on whether to register the query. In this scheme, a query that has not been registered in the past several hops has a higher chance of getting registered in its next hop, which ensures that registrations are well distributed along the path of a query message.
- **Third**, we discuss a *passive replication*-based scheme for preserving high notification effectiveness of the system even when the underlying P2P network experiences significant *churn*.
- **Finally**, we propose a local load re- distribution strategy to achieve fair distribution of notification loads among the participating peers.

The CoQUOS approach provides a registration and advertisement in unstructured overlays .The below figure 1 describes the functionality of the CoQUOS system.



II. The Coquos System:-

In this section, we are discussing about a high-level description of the CoQUOS system architecture.

2.1) Concepts and Notations:

The CoQUOS approach provides two features. They are registration and advertisement on peers in unstructured overlays. First one is, continuous query is the means through which a peer can register its long term interests with the Co-QUOS system. A continuous query, represented as:

Q =(SID; Predicate; V Time), is essentially a tuple of three components, namely, *source ID* (SID) , *query predicate* (Predicate) and *validity time* (V Time). The source ID uniquely identifies the peer issuing the query. The query predicate is the matching condition of the query, and is used by the source peer to specify its interests.

Second, Peers announce their new data items through *announcements*. An announcement is represented as

Ad = (AID; MData). The *announcing peer ID* (AID) identifies the advertising peer and the metadata (MData) is the metadata of the content being advertised.

2.2) Design Overview:

We focus on an alternate notification paradigm called the *continuous query model*.

- ✓ Which are maintained in the network for extended durations of time.
- ✓ Peers in these systems discover data items by circulating queries within the overlay network.
- ✓ The network *forgets* the queries once they have completed their circulation.

Design of the CoQUOS middleware strives for a stronger notion of notification effectiveness, wherein the goal is not only to achieve high overall success rate, but also reasonably high success rates for each individual query.

This approach works with high performance by maintaining each continuous query at one or more peers of the overlay by using beacon node.

III. Selecting Beacon Nodes Of A Query:

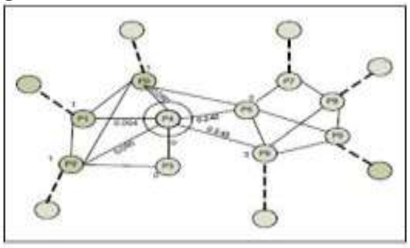
The beacon nodes plays very crucial role in notifying the source peer of matching data items. So, the choice of beacon nodes would have a significant impact on the notification success rates of a continuous query. The characteristics of a good beacon set are: **First**, the beacon nodes of a query should be distributed in every major region of the overlay network. **Second**, the beacon nodes of a query should not be located very close to one another. If there are many registrations in close proximity, a single announcement would reach multiple beacon nodes of the same query, thus generating duplicate notifications. The beacon node must satisfies the above two properties at selection stage.

The CoQUOS system incorporates a completely decentralized technique for beacon node selection. In this scheme, a continuous query is circulated in the network (by neighbour forwarding), and each peer that receives the continuous query decides independently whether to register and store the query. An important feature of our scheme is that although each peer makes a local decision regarding query registration ,the resulting beacon node sets manifest the above two important characteristics to a very high degree. For register the queries and circulating them in overlay we use two techniques

3.1.) Cluster Resilient Random Walk:

In the existing system use Flooding-based broadcast is for circulating continuous queries. But, in this scheme, messages do not go deep into the network. Random walk is another message propagation paradigm that propagates the message to that peer. Since, at each step the message is forwarded to only one neighbour, the message load imposed by random walk is very low. To overcoming the above drawback, we have designed a novel query dissemination scheme called ***cluster resilient random walk.***

Figure: The below figure shows the illustration of cluster resilient random walk



This scheme also plays important crucial role: *Two peers belonging to the same cluster generally have large numbers of common neighbors Random walk corresponds to a depth-first traversal of the network, and a message propagated through random walks has a higher probability of reaching remote regions of the network than flooding-based and pub-sub schemes.* The above property of the random walk makes it an *a*ttractive paradigm for propagating continuous queries.

3.2.) Dynamic Probability Based Query Registration:

The registration of queries at every node it visits gives the result as large numbers of unnecessary subscriptions with fixed probability, which affects the efficiency of the network. To address this problem, we have designed a novel *dynamic probability*-based technique *(DP scheme)*, for peers to decide whether to register a continuous query .In this, the registration probability of a query varies as the query traverses along its route. The central idea of the dynamic probability scheme can be summarized as follows: ***The probability of registering a query at a peer node -would be high if the query has not been registered at the nodes it visited in the recent past. In contrast, if the query has been registered at a node that visited in the past few hops, the probability of it getting registered at the current peer would be low.***

The number of beacon nodes of a query can be controlled through the initial probability and probability increment-parameters. Higher values of these parameters result in larger number of subscriptions and vice-versa.

IV. Enhancements And Discussion:

Here, we discuss about two issues that are of particular importance to the performance of the CoQUOS system, namely:

(a) ***Churn of the P2P overlay and***

(b) ***Load distribution among peers***.

4.1.) Overlay Churn:

P2P networks are, in general, highly dynamic systems, with nodes entering and exiting the system quite frequently. This *churn* of the overlay network can adversely impact the success of continuous queries and announcements. When a node Pi gracefully leaves the system, it asks one of its neighbors to handle all registered queries at Pi and also notifies all the beacon nodes with queries issued by Pi to remove the queries. However, when Pi exits the system unexpectedly, all the registrations are lost and the notification success rates of the respective queries and the matching announcements drop.

Thus, effective mechanisms are needed to alleviate the negative effects of churn in the overlay network. In order to counter the adverse effects of network churn, we have designed a low-cost technique where in the query registrations present on a peer are replicated on one or more of its neighbours. To handle that situation, we designed a low-cost technique wherein the query registrations present on a peer are replicated on one or more of its neighbours.

4.2.) Load Balancing:

The below figure shows the effective load balancing technique based on power law. Achieving good load distribution among peers is another important requirement for performance of the CoQUOS system. The number of queries and the numbers of notifications sent out per unit time by various nodes represent two key load metrics for CoQUOS system.

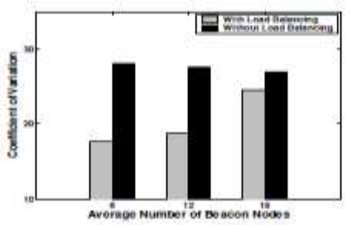


Fig. 14. Effectiveness of Load Balancing Technique (Power-law)

These load parameters can vary widely among the nodes of the CoQUOS system due to variety of reasons, including topological characteristics of the network, variation in the resource availabilities at peers or a combination of these factors. So it causes, load imbalances not only degrade the performance of the system, but may also cause overloaded peers to exit the network.

Ensuring good load balancing in decentralized, loosely coupled systems such as unstructured P2P overlays is challenging. To ensure the good load balancing, we use **Skip graphs** and ***passive load balancing***. *The* skip graphs require maintenance of circular linked lists consisting of all nodes in the system. Hence, they are not suitable for a large-scale, dynamic environment like unstructured P2P content -distribution platforms.

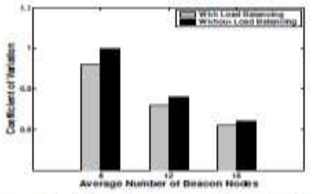


Fig. 15. Effectiveness of Load Balancing Technique (Random)

.

So we use the second strategy called the **passive load balancing scheme** is extremely light-weight. In this scheme, an overloaded peer avoids registering any new queries until the load becomes more balanced

V. Experimental Evaluation:-

The objectives of our experimental study of the CoQUOS system are four types:

1.) Evaluating the effectiveness of the CRW technique in propagating continuous queries.
2.) Studying the performance of the dynamic probability approach for query registrations,
3.) Evaluating the churn resilience and the load balancing mechanisms,
4.) Evaluating the communication costs.

6.1.) Performance Metrics:

We use two performance metrics for quantifying CoQUOS middleware's effectiveness, namely *mean notification success rate* and *minimum notification success rate*. Below table shows measuring the performance the CRW technique in propagating continuous queries.

Table1: The Performance of Query Dissemination Schemes at Various Message Loads

Per-Query Msg	Load Flooding	PRW	CRW
4	**22%**	**22%**	**31%**
27	**41%**	**72%**	**82%**

6.2.) Evaluating the Dynamic Probability Scheme:

To evaluate the query registration schemes, we have to be used in conjunction with a query propagation scheme. In our experiment we simulate the fixed probability and the dynamic probability schemes in conjunction with both the PRW and the CRW query propagation techniques to obtain four combinations, namely: *pure random walk with fixed probability (PRW-FP)*, *pure random walk with dynamic, probability (PRW-DP)*, *cluster resilient random walk with fixed probability (CRW-FP)* and *cluster resilient random walk with dynamic probability (CRW-DP).*The CoQUOS system uses the CRW-DP query registration scheme.

6.3.) Effectiveness of Churn Resilience and Load Balancing Mechanisms:

Figure: The below figure.12 shows Churn Resilience based on Power-law

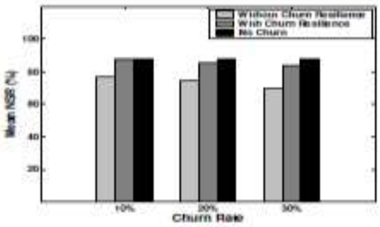


Fig. 12. Effectiveness of Churn Resilience Technique (Power-law)

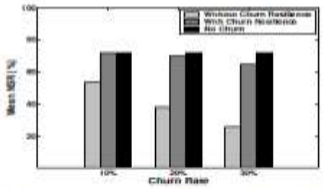


Fig. 13. Effectiveness of Churn Resilience Technique (Random)

Figure: The above figure.13 shows Churn Resilience based on random .However, the peers are no longer static; they can enter and exit the system at arbitrary points in time. We compare our churn resilience scheme to two other scenarios. Namely, a CoQUOS system with no node dynamics and a dynamic CoQUOS network but with no failure resilience.

6.4.) Messaging Costs of the CoQUOS System:

Finally, we study the communication costs of the CoQUOS system. We evaluate each of the four components of the cumulative message rate, namely query message rate, announcement message rate, notification message rate and-system maintenance message rate. The CoQUOS system is executed on the power-law and the random networks with all its capabilities enabled.

Table6: Communication Costs of the CoQUOS system

	Query Msg. Rate	Advt. Msg. Rate	Ntfy. Msg. Rate	Mtn. Msg. Rate	Cuml. Msg. Rate
Power-law N/W	2.5	11.92	0.26	0.38	15.06
Random N/W	2.5	6.01	0.38	60.97	9.85

The above table shows the Communication costs of the CoQUOS system.

VI. Related Work:-

In the past few years, both structured and unstructured P2P networks have experienced significant research. Searching through ad-hoc queries has been the predominant information discovery mechanism in P2P networks. Several researchers have studied efficient and scalable alternatives to the flooding-based searching in unstructured P2P networks .Random walk and its variants have been explored as alternatives to the broadcast strategy.

However, the ad-hoc query paradigm is inadequate for advanced P2P content sharing applications. A second research area that is closely related to work presented in this paper is that of pub-sub systems. The CoQUOS system differs from all the above techniques, it achives guaranteed notification and does not impose any restrictions on the topology of the overlay, employs lightweight techniques, yet it provides high success rates.

VII. Conclusion:-

Mechanisms that enable individual peers of unstructured P2P content sharing networks to register long standing queries and receive notification with pub-sub systems on unstructured overlays is often a very complex task. To overcome the above problem we presented the design and evaluation of a lightweight system, called CoQUOS, which supports continuous queries in unstructured P2P networks by incorporates several novel features such as cluster resilient random for query propagating, dynamic probability scheme for query registration and a lazy replication technique for countering network churn.

VIII. ACKNOWLEDGMENTS:

This work is partially supported by National Science
Foundation under Grant No. 0716357. An extended abstract
Of this paper appeared in the Proceedings of InternationaParallel and Distributed Processing Symposium 2007.

References:-

[1] Gnutella P2P Network. www.gnutella.com.
[2] Kazaa P2 P Network. www.kazaa.com.
[3] TIB/Rendezvous. White paper, 1999.
[4] S. Androutsellis Theotokis and D. Spinellis. A Survey of Peer to-Peer Content Distribution Technologies. *ACM Comput. Surv.*2004.
[5] B. Arai, G. Das, D. Gunopulos, and V. Kalogeraki. Approximating Aggregation Queries in Peer-to-Peer Networks. In *Proceedings of the* 22nd *International Conference on Data Engineering (ICDE)*, 2006.
[6] R. Baldoni, C. Marchetti, A. Virgillito, and R. Vitenberg. Contentbased Publish-Subscribe over Structured Overlay Networks. In *Proceedings of ICDCS,* 2005.