

Policy-based Management of Semantic Clustering

Dominic Jones, John Keeney, David Lewis, Declan O'Sullivan

Knowledge & Data Engineering Group (KDEG) – Trinity College Dublin, Ireland

{Jonesdh | John.Keeney | Dave.Lewis | Declan.OSullivan}@cs.tcd.ie

1. INTRODUCTION

We introduce the concept of a Knowledge-based Network KBN [1] as an extended Content-based Network (CBN), where semantically rich messages co-exist with the more traditional CBN messages format. One of the main advantages this new technology, a KBN, is in its ability to define subscription filters across the semantics of the contents of messages, in addition to the commonly available, traditional, CBN filters.

Knowledge-based Networks (KBN) support the routing of semantically enriched messages between interested parties across a common network of message brokers. Knowledge-based Networks introduce an unprecedented level of semantic richness, which allows semantic clusters of publisher/subscriber to form within a KBN creating pockets of focused interest. When this is exploited direct performance increases can be seen [1]. We propose a flexible policy-driven mechanism to manage the future semantic clustering of Knowledge-based Networks.

2. KBN CLUSTERING

Within Knowledge-based Networks publishers and subscribers direct their publications or subscriptions towards single, or multiple brokers. Clusters of publishers, subscribers and brokers are formed around groups of users interested in the same content. Baldoni et al [2] present an “architecture based on clustering peers subscribed to the same topic” as well as the work of Anceaume et al [3] in which “subscribers self-organize according to similarity relationships among their subscriptions”. These works support the argument that a generalized clustering technique is of benefit within a Distributed Event-based System. It has been shown that the overall network and broker performance of a Knowledge-based Network can be increased through semantic clustering [1].

The provision of semantically rich publications and subscriptions allows for the introduction of an even stronger level of semantic clustering than shown in current work. For example, basing the network around an ontology in which academic conferences are represented enables the clustering of academics interested in research areas, conferences, locations, dates and the *relationships between these concepts*, as opposed to static references to possible publications. When a user's interest change, the ontology changes, and thus the clusters which are formed around the users and their ontologies should also change. This allows for the natural representation of a drift in interests to be represented semantically and must be represented in the Clusters themselves, seen as an important change in the underlying structure of the KBN.

3. MOTIVATION

In a clustered Knowledge-based Network some brokers deal with a focused range of semantics within the set of messages they can

process. This both categorizes the messages passing across the broker/network and moves towards a loose guarantee that a message arriving at a broker is of interest to that broker (or group of brokers) and the subscribers / publishers connected to that broker.

Through the clustering proposed within this paper we see three main benefits beyond a normal semantically enhanced publish/subscribe system, these benefits allow KBNs to: 1) Reduce and optimize the routing and subscription information held each node through the reduction of the possible set of interests applicable to that node; 2) Reduce the number of hops between related producers and consumers by semantically clustering both around relevant brokers within the network; 3) Increase the number of subscription-aggregation occurrences, which decreases and optimizes the number of subscriptions held in any particular subscription table. A smaller and more ordered subscription table decreases the time taken to match a publication to a possible set of subscriptions. Additionally given that each broker within the network holds an ontological representation of the knowledge-base on which it may be required to reason, and through the introduction of clustering, we can reduce the size of this knowledge-base and the subsequent reasoning overhead at each broker. This methodology is still dependent upon an upper-level broker network which knows of all topics, at a high level, such that an unknown message can be passed up the router chain until a router knows of the knowledge presented in the message and can correctly forward the message, at least towards the correct cluster.

Semantic Clustering demands the natural grouping of clients across a network of brokers which pushes towards there being a refinement and movement away from the publish/subscribe “anywhere” methodology. Although a message inserted anywhere into the network must still be routed towards any interested subscriber we must try to group related clients and brokers. This aligns with the vision of using an overlay network, where clients (and sub-brokers) may not necessarily connect to a geographically close broker, but rather offset the decreased network performance with more optimized, semantic, application-level concerns. Outlining that scalability and accessibility need to be central in a semantically enriched operational environment it becomes important that decisions are made which are both relevant and representative of the networks operation as prescribed by service-level-agreements and managerial decisions made using as high a level of governance as possible.

4. MANAGEABILITY

Having established the driving force behind clustering, the semantics associated with those clusters, and the performance gains shown through the coupling of the two [1], it is possible to address the future direction of this research. By visualizing a

network of brokers on which producers and consumers communicate it is easily seen that change and fluidity within this trio is as important as the purpose of the network itself. This supports the argument that human designed clusters are only as efficient as the last time they were designed. Therefore changes in interests across the consumers, producers and brokers within the network must be represented in the topology and dynamics of the clusters themselves. It is planned, and important, that the cluster management system currently in design, will allow for this change to be recognized, acted upon and represented in the number, meaning and groupings of clusters within the system.

An initial metric identified in the success of the cluster management system is the validity and freshness of any cluster at any point in time. In the term freshness we identify that a cluster that has been updated, used, refined and is relevant, is fresh. We define *Validity* as being supported through the ability to introduce management techniques, protocols and policies for use within a cluster-able semantically enriched KBN which assure the increases in performance shown through initial investigation are integral in the future of KBN technology. In this model the *ontology is central to the semantic meaning and purpose of the network*. We therefore propose that the management system should fit the ontology and not that the ontology should fit the management system.

5. POLICY-BASED MANAGEMENT

Policy-based Management (PBM) is an administrative approach used to simplify the management of a system by declaring operating rules which deal with situations that are likely to occur. Informally, policy rules can be regarded as an declarative instruction or authority for a manager to execute actions on a managed target to achieve an objective or execute a change. Supported by this premise, PBM is being used within this research to provide a set of external controls which, in certain configurations, allow management level decisions to be filtered down into the operational characteristics, and nature of a clustered KBN. The goal behind using PBM in the clustering of those brokers and clients is to control the behavior of brokers and clients by employing well defined policy rules so that an administrator can manage the network as an entity in-itself. PBM allows this to be realized through its ability to implement changes across the network as a whole, in comparison to managing individual network entities and actions.

The current clustering capabilities of the KBN are relatively static in the sense that rules regarding overall operation are hard-coded into the core of the application itself. In introducing policy-based management, *operational diversity* is encoded in the rules and reactions of the policies as opposed to being prescribed in the KBN itself. This new management plane for the KBN, directed by the policy system will dramatically increase the flexibility and functionality of the cluster management system. This research aims to prove that the policy-driven KBN is as flexible as the underlying policy system itself, and that policy rules can be readily enforced within a Distributed Event-Based System. There are three main cluster management techniques currently in development applying specifically to message brokers, which include: Dynamic Cluster Creation, Deletion and Merging. With regard to the publishers and subscribers attached to these brokers the policy system aims to direct and redirect existing and new clients and sub-brokers to brokers which match their semantic

characteristics and interests. This creates an organic structure where captured clusters drift over time, managed using PBM, while continually assessing network operation and behavior. These characteristics are predicated to grow into a fuller set of KBN management tools, which will be addressed in future publications and will aim to show future increases in performance. The authors will present a Policy-based Clustering, Merging and Deletion management plane in which audits occur both regularly and efficiently in order to assure a high level of freshness and validity in the cluster hierarchy, ultimately improving performance. As part of the evaluation of the introduced system an extended semantic benchmark [5], based on the original Siena Specific benchmark [6], will be extended so that performance increases/decreases can be clearly identified and documented.

6. CONCLUSIONS

The approach taken within this paper aims to provide an initial area for discussion regarding the relationship between KBN technology, Clustering and Policy-based management. An extension to previous work on clustering [1], is proposed as a starting point, from which the benefits arising from semantic clustering are carried forward using Policy-based Management. The achievement of this starts with the necessity to outline initial high-level operational targets and goals, as has been done within this paper. This will be followed by initial results which are planned for late fall 2008 and this paper sets the scene for the detailed evaluation of those results. Most importantly this paper aims to provide the arena for initial early feedback and critical contribution from the DEBS community.

7. REFERENCES

- [1] J. Keeney, D. Jones, D. Roblek, D. Lewis, D. O'Sullivan, "Knowledge-based Semantic Clustering," *Symposium on Applied Computing*, Fortaleza, Brazil, March 16 - 20, 2008.
- [2] R. Baldoni, R. Beraldi, V. Quema, L. Querzoni, S. Tucci-Piergiovanni, "TERA: Topic-based Event Routing for peer-to-peer Architectures," *Distributed event-based systems (DEBS 2007)*, New York, NY, USA, June 20-22, 2007.
- [3] E. Anceaume, M. Gradinariu, A. K. Datta, G. Simon, A. Virgillito, "A Semantic Overlay for Self-Peer-to-Peer Publish/Subscribe," *Distributed Computing Systems (ICDCS 2006)*, Lisboa, Portugal, July 4-7, 2006.
- [4] R. Boutaba, I. Aib, "Policy-based Management: A Historical Perspective," *J. Netw. Syst. Manage.*, vol. 15, pp. 447-480, 2007.
- [5] J. Keeney, D. Lewis, and D. O'Sullivan, "Benchmarking Knowledge-based Context Delivery Systems," *International Conference on Autonomic and Autonomous Systems (ICAS 2006)*, Silicon Valley, California, USA, July 16-21, 2006.
- [6] A. Carzaniga and A. Wolf "A Benchmark Suite for Distributed Publish/Subscribe Systems," *Dept. of Computer Science, University of Colorado*. 2002.

This material is based upon works supported by the Science Foundation Ireland under Grant No 05/RFP/CMS014.