Improving Scalability in Pub-Sub Knowledge-Based Networking by Semantic Clustering

John Keeney, Dominic Jones, Dominik Roblek, David Lewis, and Declan O'Sullivan

Knowledge & Data Engineering Group (KDEG) - Trinity College Dublin, Ireland
{John.Keeney,jonesdh,roblekd,Dave.Lewis,
Declan.OSullivan}@cs.tcd.ie
http://kdeg.cs.tcd.ie/technologies/kbn

The three main pub/sub systems – type-based, content-based and topic-based networks allow for normalised subscriptions and publications using a combination of push/pull message delivery. Knowledge-based Networks (KBN) extends content-based networks by allowing subscriptions to be matched not only on the contents of messages, but also on some semantics of the message contents [1,2,3]. This creates a fuller, richer and more meaningful system whereby publishers and subscribers can be matched using more a expressive subscription mechanism. Here we discuss how a Knowledge-Based Network implementation was extended to support node clustering based on subscription semantics, thereby improving performance and scalability.

In extending the Knowledge-Based Network to incorporate semantic-based clustering, this research aims to provide a network environment in which routing nodes, publishers and subscribers are clustered based on their semantic footprint and interests. The benefits of this are twofold: Firstly, this reduces the processing time involved in making routing decisions based on messages content. Its takes fewer hops to get from source to destination, as these are already closely linked based on the likelihood of there being a match between the two. Secondly, this allows for natural grouping of likeminded publishers and subscribers as seen in traditional web forums / newsgroups. The cluster-based approach to pub/sub networks turns the normal user-based search paradigm full circle as network data is passed from node to node towards those who are most likely to be interested in the data as opposed to those users searching out that same data.

An implementation of a KBN, based on the Siena CBN, has been implemented, and enables the efficient routing of distributed heterogeneous knowledge to, and only to, nodes that have expressed a specific interest in that knowledge. This KBN implementation currently operates on distributed PlanetLab Nodes.

Initially clusters are statically designed and operated. In this sense nodes are assigned to clusters without the possibility of changing clusters once they have joined; later users will be able to join and leave clusters independently. Clusters will then be seen as organic structures in which users join and leave as their own personal interests change, grow, reform and are refined.

Preliminary evaluations show the importance of semantic clustering for efficient performance and network scalability. These evaluations demonstrate how even inflexible and static clustering can have a substantial positive effect. Ongoing research is focusing on how clustering can be performed dynamically as the semantics of the data in the network changes. Such clustering may also form the basis for a viable means for forming KBN sub-domains, thereby sharing the load of development and supporting incremental deployment.

R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM 2007 Ws, Part I, LNCS 4805, pp. 19–20, 2007.

Current work is also focusing on integrating policy-based cluster management for the KBN [1,3] to support much more sophisticated cluster schemes, where sub-cluster intercommunicate over a super-peer network. This will support overlapping clusters and hierarchies of clusters under separate administrative control. Policy-driven clustering enables the size of the super-peer network and the size and granularity of peer clusters to reflect different application domain needs. For example, the clustering policy may be specified in terms of accuracy and latency as well as the semantic spread of the query-able knowledge-base, or in terms of queries across a peer population and subscription and notification load across that population. This use of clustering policies supports innovation in clustering strategies by allowing peers to introduce new policy elements and the supporting super-peer matching capabilities.

The scalability and flexibility of the KBN under high load of heterogeneity will be evaluated using a combination of small-scale and large-scale simulations and deployments in order to test the message overhead involved in clustering and the effectiveness of semantic load sharing. To facilitate these investigations a tool has recently been completed to allow the parameterised generation of arbitrary ontologies with defined characteristics, suitable for a diverse set of application scenarios.

Ultimately we aim to design and validate differing clustering policies, and tune a range of semantic distance calculations to make semantic clustering more effective. We will also assess the impact of policies on the coexistence of different reasoning capabilities in KBN nodes. In addition, the effect of semantic interoperability in node matching functions and in inter-cluster communications is being assessed [2]. This requires evaluation of different schemes for injecting newly discovered semantic interoperability mappings into the ontological corpus held by any given cluster, as well as how these mapping are shared between clusters [2]. We expect that any practical system will need to adapt its clustering to reflect the constantly changing profile of semantics being send and subscribed to via the KBN. This also raises interesting possibilities of the KBN operating as a knowledge discovery tool.

One of the main questions that surround the use of ontologies deep in the network at the routing layer remains the evaluation of the resulting performance overhead. Previous small scale studies in this area [1,3] shows a definite performance penalty, but this may be acceptable when offset against the increased flexibility and expressiveness of the KBN subscription mechanism. This research will continue evaluating how the performance of off-the-shelf ontology tools will affect the scalability of the KBN at larger scales.

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

[1] Lewis, D., Keeney, J., O'Sullivan, D., Guo, S.: Towards a Managed Extensible Control Plane for Knowledge-Based Networking. In: State, R., van der Meer, S., O'Sullivan, D., Pfeifer, T. (eds.) DSOM 2006. LNCS, vol. 4269, Springer, Heidelberg (2006)

[2] Guo, S., Keeney, J., O'Sullivan, D., Lewis, D.: Adaptive Semantic Interoperability Strategies for Knowledge Based Networking. In: Workshop On Scalable Semantic Web Knowledge Based Systems at OTM 2007, Vilamoura, Algarve, Portugal (November 2007)

[3] Keeney, J., Lewis, D., O'Sullivan, D.: Ontological Semantics for Distributing Contextual Knowledge in Highly Distributed Autonomic Systems. Journal of Network and System Management (JNSM) 15(1) (March 2007)