Knowledge-based Networking

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
http://kdeg.cs.tcd.ie/KBN

1. INTRODUCTION

Content-based networks (CBN), such as [1-3], have formed around the necessity to match a varying subscriber base to that of a networks publication creators. This "de-coupling" of the parties involved in the communication process allows for message routing to be conducted based on who *is interested* in a particular message, through a routing table based on subscriber's filters applied to each message, rather than flooding the network in the search for interested possible parties. The underlying message routing structure allows for a message inserted anywhere in the network to propagate across the network based on positive matches to filters (subscriptions) until each and every client interested in the message has been notified in an efficient event-based manner.

Content-based Filtering, which is the core message matching algorithm within Siena [4] is described be Mühl et al in [5] as a set of "Filters [which] are evaluated against the whole contents of notifications" in this case notifications can be thought of as publications. Notifications are only forwarded to a user when the contents of the filter, which is made up from a set of constraints, matches against the messages contents. This allows for a flexible message format and increases the desired separation between the publishers and subscribers within the network. De-coupling in this sense is increased through the ability of single parts of the pub/sub message to match as opposed to the whole messages matching. This makes more likely the possibility that a consumer and producer, using separate formats may match based on a sub-set of their publication and subscription. The subscription filters which match these publication messages are constructed using a set of operators, where the range of these operators determines the expressivity and scalability of the network through its support for subscription aggregation.

The work within this paper does not however focus on Content-based Networks, rather focuses on Knowledge-Based Networks (KBN) [6-8]. Where CBNs and KBN differ is in the rich set of ontological operators available within the KBN which extend the original Siena [1] CBN. The KBN used for this work provides ontological concepts as an additional message attribute type, onto which subsumption relationships, equivalence, type queries and arbitrary ontological relationships can be applied. It also provides for a bag type to be used that allows bag equivalence, sub-bag and super-bag relationships to be used in subscription filters, composed with any of the Siena subscription operators or the ontological operators previously mentioned. These additional operators and types fundamentally change the way in which messages are formed, the semantic scope of those messages, the matching of publications to subscriptions and most importantly the use of ontologies to describe the semantics used in the messages, all with the single aim of increasing expressiveness and flexibility of subscriptions and publications. The driving force behind any pub/sub network is the messages that flow between the producers and consumers - within a KBN this flow is as much more semantically expressive.

With semantic-web ontologies we are able to describes concepts and types, instances of those types. We can also map the relationships between concepts, add separations, similarities and importantly classify and group concepts in a taxonomical fashion. This, combined with a distributed event-based system, allows us to embed rich semantics within publications, in addition to the traditional cont-based publications used previously. This allows the ontological-based semantically-rich operators to play a major part in the core message matching and routing across a network, along-side content-based filters. Semantics allows Knowledge-based Networks to begin the grouping of users, where the categorization of generalized interests, can be represented through the formation of common interest groups, allowing for a level of "traffic refinement" [9] which we term "semantic clusters".

2. IPTV AS AN APPLICATION DOMAIN

Traditional analog-based television offered the user a limited number of channels across a limited spectrum of content. With the introduction of Internet-based Television broadcasts both the range and diversity of the offered content has dramatically increased. In the United Kingdom both the BBC and Channel 4 offer Internet-based Television across the BBC iPlayer [10] and Channel 4 On-Demand Services [11]. In the United States ABC [12] and Fox [13], to name a few, offer an Internet-based TV service. With the additional broadband available with such TV transmissions comes the ability to provide an additional range of fine-grained semantically-rich program specific information. This information can include; a textual synopsis of the program, the categorization of the program, the ability to tag the program as well as the more mundane start and end times, regionally set.

Joost [14] has recently revolutionized the IPTV market and describes itself as "The new way of watching free, full-screen, high-quality TV. All the things you love about TV, fused with all the fun and interactive power of the internet." YouTube is the phenomenally popular video sharing site, based on a user search algorithm in which videos are sought rather than delivered to a particular user. Where the Knowledge-based Network technology adds to both of these services is in its ability to asynchronously inform users of broadcasts that are of real interest to the user. Here the users' interests are based on a set of long-lived, pre-defined, semantically-rich subscriptions.

Television previously offered an audio and visual feed, nothing else. With the greater bandwidth and with the correct use of the semantics associated with the TV feeds' metadata streams, dynamic personalized content discovery becomes a real possibility. Current technologies allow a user with a broadband connection to rewind and pause previously aired programs with no more technological knowledge required that needed to program a VCR. With such a vast wealth of content available to the user, originating from a rapidly growing set of content producers, requires the provision of a content discovery system that operates in a more efficient manner than currently offered through traditional program listings. This can be made possible through the use of Knowledge-based Networks (KBN). Publications in KBNs, in line with Content-based Networks, are formed around the concatenation of separate sub-sections of a greater message. This allows, as shown through this application demo, for the URL of a IPTV feed (channel or program) to be packaged into a semantically-simple CBN message with metadata including: a description, date originally aired, program creator, copyright and technical specs. However with such a vast range of possible scenarios it becomes unmanageable over a large time frame, in such an application scenario the user is no longer being offered matching content, the user begins to search for matching content according to more semantically rich searches.

With the introduction of both Ontologies and Knowledge-based Networks the ability to be offered content based on a set of semantically defined interests refines both the possible matches for a subscriber and the ease in which a subscription can be formed. By offering to the user a categorization system for all the possible content and by using the ontological "subsumes" and "subsumed-by" operators opens up a multitude of additional content discovery possibilities. The addition of these new operators allows for the discovery and delivery of previously unmatchable messages.

It is important to note that in this application demo we are enhancing the discoverability of the feeds associated with an occurring presentation. The publications associated with a particular show will include:

- title
- feed encoding and definition
- description
- time, location and date
- ontological categorization
- a set of keywords
- The link to the feed

By using semantically enriched subscriptions, the above publications become readily matched using a combination of the content-based and knowledge-based subscriptions. Ontological Categorization and the use of sets of keywords present the ability to represent a user's interests in a category and keyword basis. For example, a subscription for all documentaries regarding news-based topics, along with all British and Australian soap-operas, Irish news and no reality TV programs can be easily represented in this new semantically enriched subscription format.

3. SCENARIOS

The Implementation represented within this Demo paper operates around an extended Siena Knowledge-based Network. The demo uses a number of message brokers arranged hierarchically which act as both routers and brokers in the delivery of publications to subscribers. The brokers match the filters, the routers forward the messages. Publishing clients attach to randomly chosen brokers and publish semantically diverse messages regarding various IPTV feeds and programs. Users form subscriptions from a range of unique categories. For example, the ontological sub-class (subsumed-by) operator allows the user to subscribe to every category below the category: Italy, Rome. This subscription would return every IPTV feed below this, whether this be a news program from Rome, a Papal Speech from the Vatican, a DEBS Video Cast or Soccer game played in a stadium in Rome. Another example subscription could form around the desire to receive everything that is published by a specific operator, the BBC for example. Or everything that is less specific than a specific program. Taking "The Simpson" cartoon it becomes possible to subscribe to everything that is less specific than "The Simpson's" providing matches to all "Cartoons" for example.

Each publisher, server, and subscriber operates around a centrally loaded and reasoned ontology. This ontology allows for a harmonized view, across all routers, of a common set of semantics from which subscriptions can be formed. The ontology in this sense can be seen as the driving force behind the representation of the complete set of knowledge, from which the publications and subscriptions can be formed.

Keywords form an important part of a message, using Sets, as introduced in [15] we are able to create an associated bag of key Words which are attached to each publication. "Tagging" and the use of keywords has become an important part of the human classification process. In being able to tag IPTV feeds, it becomes possible to add another layer of classification to the content itself. As outlined within this paper, given a single feed were able to attach an ontological categorization of the content, a bag of keywords and traditional Content-based Network attributes, such strings, Integers and Booleans. Subscription covering, central to Siena [1], has been assured with these new ontological operators

4. A TEST-BED DEPLOYMENT

The demonstration accompanying this paper offers a local example of Knowledge-based Network technology, involving a number of brokers, running on the same machine, using different ports. However in the validation of the service and in driving to meet a desired Internet scale architecture, the technology presented within this paper has been deployed globally the on PlanetLab network [16]. PlanetLab offers a distributed test-bed, on which currently there are over 600 research projects operating across around 700 machines distributed in research institutions and academic globally. This walled garden of servers allows for large-scale experimentation to occur across the same long-haul backbones on which the current Internet operates. PlanetLab offers researchers, the chance to test deployment on a large scale in comparison to local networks.

A semantic clustering experiment, detailed in [8], was deployed and evaluated across PlanetLab. In actuality this meant the experiment, similar to the domain experiment represented within this paper, was based around 77 KBN routers/brokers/publishers and subscribers deployed across 77 random PlanetLab servers in Europe, North America, South America, Asia and Australasia. There are a number of technical factors which affected this deployment including, to name a few, the actual upload to and deployment from 77 remote machines, the high level of unreliability in the machines, no guarantee of remote file storage (machines can be wiped, and are wiped, regularly) and the collection of a remote set of results. These issues were addressed in a number of ways, removing any reliance on locally contained settings with regard to the execution of servers. In other words nodes began to look to a central authority for execution configuration and the hierarchical topology required to operate. In using PlanetLab not only did this take the experiment to a point as close to real world deployment as possible it provides invaluable local and remote deployment results, which can be compared, strengthening the set of overall results supporting the Knowledge-based Networks deployment.

5. ADDITIONAL APPLICATION DOMAINS

In allowing for semantically rich publications and subscriptions the domain underlying the KBN changes, dramatically. The core of the KBN becomes an ontology or a set of combined ontologies. This opens up a once semantically sparse network to rich semantic messages based around the format of the ontological backbone. Semantic richness creates many exciting possibilities for the application of a KBN driven publish / subscribe paradigm.

The domain exposed within this paper, IPTV, is one of a small set of possible application domains for use with Knowledge-based Networks. Some additional domains which are being explored include: Scientific Calls for Papers (CFPs), Hospital Equipment and Patient Monitoring, Large Scale Ordering / Stock Control (JIT), Environmental Controls (Building Management), Commodities / Currencies and Network and Fault Management. A key component to future domain choice is that with ontologies no longer does each client need to know the full set of pre-decided operators or topics. By using Knowledge-based Network Technology the ontology becomes the source of subscriptions, and thus the domain becomes an interchangeable part of the network as a whole.

6. FURTHUR READING

There are a number of publications associated with Knowledge-based Networking which equip the reader with a fuller overview of the domain. [6] presents an overview of the initial ideas behind Knowledge-based networks and an initial implementation. This is supported by a detailed technical paper at this conference (DEBS2008) [17] which outlines a detailed evaluation of the ontological extensions to the Siena CBN, with this implementation forming the base of this technology demonstration. Additionally [17] presents the reader with an introduction to the concept of bags (similar to sets), as created in [15]. The Clustering of Knowledge-based Network servers, publishers and subscribers is evaluated in detail in [8], which supports the accompanying Fast Abstract paper to be presented at this conference (DEBS2008) [18] where policy-based management is introduced to the reader as a component in the formation, merging and deletion of semantic clusters in a KBN. The URL located in the address section of this paper directs readers to a full list of publications associated with Knowledge-based Networks and any questions can be directed to the papers authors.

7. FUTURE WORK

Research, currently under review, is investigating the move away from a static hierarchy in the extended Siena KBN allowing for subscribers to move from broker to broker in a dynamic manner. This research is focusing on the clustering of like-minded users, using policy, within Knowledge-based Networks, as introduced in [18]. Being able to move strands of subscribers from server to server within a hierarchical topology is an important part of the proposed re-clustering. The policy-based clustering of Knowledge-based Networks will also allow for the use of a standardized benchmark which will be created from an extension to the work in [19] in which the performance of Knowledge-based networks can be evaluated against changes in the overall network operation.

Ongoing research is also investigating the use of semantically enhanced RSS feeds as a source of publication for use with KBNs. Once complete the enriched RSS feeds will be coupled with a newly designed user interface used to form user subscriptions.

8. CONCLUSION

Knowledge discovery is central to the success of the current Internet. Being able to search and find relevant information has to a great extent given Google the vast share of the web-search market. Users want information, and they want to find it fast, a service which offers this to the user is often a great success. However traditional and current techniques involved in knowledge-discovery involve the user searching for the content which he/she desires. With the introduction of Knowledge-based Networks this methodology is reversed, whereby knowledge is pushed towards the user, not pulled by the user. This shift from traditional search and find techniques was pivotal in the creation of initial Content-based Networks. Knowledge-based networks extend this technique by using the semantics of the message to direct proposed matches towards interested parties, which greatly increase the reach and target of any given message.

The demonstration system presented within this paper represents a knowledge discovery system, which aids the user in the delivery of relevant publications to previously defined subscribers. The technology behind Knowledge-based Networks aims to provide a cross-domain message delivery system, on which the advantages of publish / subscribe networks can be realized with the additional semantic representation of publications and subscriptions.

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

REFERENCES

[1] A. Carzaniga, D. S. Rosenblum, and A. L. Wolf, "Design and evaluation of a wide-area event notification service (Siena)," *ACM Trans. Comput. Syst.,* vol. 19, pp. 332-383, 2001.

[2] P. Pietzuch and J. Bacon, "Hermes: A Distributed Event-Based Middleware Architecture," in *Proceedings of the 22nd International Conference on Distributed Computing Systems*: IEEE Computer Society, 2002.

[3] The Gryphon Team. Achieving Scalability and Throughput in a Pub/Sub System. Research report, IBM, Feb. 2004.

[4] R. Baldoni, R. Beraldi, L. Querzoni, and A. Virgillito, "Efficient Publish Subscribe through a Self-Organizing Broker Overlay and its Application to SIENA," *The Computer Journal* 2007.

[5] G. Muhl, F. Fiege, and P. Pietzuch, *Distributed Event-Based Systems*: Springer-Verlag, 2006.

[6] D. Lewis, J. Keeney, D. O'Sullivan, and S. Guo, "Towards a Managed Extensible Control Plane for Knowledge-Based Networking.," *DSOM,* vol. 4269, pp. 98-111, 2006.

[7] J. Keeney, D. Jones, D. Roblek, D. Lewis, and D. O'Sullivan, "Improving Scalability in Pub-Sub Knowledge-

Based Networking by Semantic Clustering," in *The 6th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2007)* Vilamoura, Algarve, Portugal Nov 27-29, 2007.

[8] J. Keeney, D. Jones, D. Roblek, D. Lewis, and D. O'Sullivan, "Knowledge-based Semantic Clustering," in *The 23rd Annual ACM Symposium on Applied Computing* Fortaleza, Ceará, Brazil, March 16 - 20, 2008.

[9] R. Baldoni, R. Beraldi, V. Quema, L. Querzoni, and S. Tucci-Piergiovanni, "TERA: Topic-based Event Routing for peer-to-peer Architectures," in *DEBS 07 Proceedings of the 2007 inaugural international conference on Distributed event-based systems* New York, NY, USA: ACM Press, 2007, pp. 2-13.

[10] "http://www.bbc.co.uk/iplayer/," "Making the unmissable, unmissable", Accessed May 2008.

[11] "http://www.channel4.com/4od/index.html," "Channel 4's TV and Film on Demand Service", Accessed May 2008.

[12] "http://dynamic.abc.go.com/streaming/landing," "ABC.com Full Episode Player", Accessed May 2008.

[13] "http://www.fox.com/FOD/," "Fox on Demand", Accessed May 2008.

[14] "http://www.joost.com/," "Free online TV - Comedy, cartoons, sports, music and more." Accessed May 2008.

[15] D. Roblek, "Decentralized Discovery and Execution for Composite Semantic Web Services," M.Sc. Thesis, University of Dublin, Trinity College, Ireland., 2006.

[16] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman "PlanetLab: an overlay testbed for broad-coverage services," *SIGCOMM Comput. Commun. Rev.,* vol. 33, pp. 3-12, 2003.

[17] J. Keeney, D. Roblek, D. Jones, D. Lewis, and D. O'Sullivan, "Extending Siena to support more expressive and flexible subscriptions," in *The 2nd International Conference on Distributed Event-Based Systems (DEBS 2008)*, Rome, Italy., 2008.

[18] D. Jones, J. Keeney, D. Lewis, and D. O'Sullivan, "Policy-based Management of Semantic Clustering," in *The 2nd International Conference on Distributed Event-Based Systems (DEBS 2008)*, Rome, Italy, 2008.

[19] J. Keeney, D. Lewis, and D. O'Sullivan, "Benchmarking Knowledge-based Context Delivery Systems," in *Proceedings of the International Conference on Autonomic and Autonomous Systems*: IEEE Computer Society, 2006.