A Scaleable Event Infrastructure for Peer to Peer Grids

Geoffrey Fox, Shrideep Pallickara, Xi Rao and Qinglin Pei
Community Grid Labs, Department of Computer Science
Indiana University

The peer-to-peer (abbreviated as P2P) style interaction [10] model facilitates sophisticated resource sharing environments between "consenting" peers over the "edges" of the internet; the "disruptive" [11] impact of which has resulted in a slew of powerful applications built around this model. Resources shared could be anything – from CPU cycles, exemplified by SETI@home (extraterrestrial life) [14] and Folding@home (protein folding) [15], to files (Napster and Gnutella [17]). Resources in the form of direct human presence include collaborative systems (Groove [18]) and Instant Messengers (Jabber [16]). Peer "interactions" involves advertising resources, search and subsequent discovery of resources, request for access to these resources, responses to these requests and exchange of messages between peers. An overview of P2P systems and their deployments in distributed computing and collaboration can be found in [9]. Systems tuned towards large-scale P2P systems include *Pastry* [19] from Microsoft, which provides an efficient location and routing substrate for wide-area P2P applications. Pastry provides a self-stabilizing infrastructure that adapts to the arrival, departure and failure of nodes. The JXTA [12] (from *juxta*position) project at Sun Microsystems is another research effort that seeks to provide such large-scale P2P infrastructures. Discussion pertaining to the adoption of event services as a key building block supporting P2P systems can be found in [8,9]. We propose a peer-to-peer (P2P) grid comprising resources such as relatively static clients, high-end resources and a dynamic collection of multiple P2P subsystems. We investigate the architecture, comprising a distributed event service that will support such a hybrid environment.

Narada [3,4,5,6,7,24] is an event brokering system designed to run on a large network of cooperating broker nodes. Communication within Narada is asynchronous and the system can be used to support different interactions by encapsulating them in specialized events. The Narada scheme of automating broker additions within a distributed cluster architecture, while resulting in the creation of "small world networks" [1,2], allows support for large heterogeneous client configurations that scale to arbitrary size. Narada guarantees delivery of events in the presence of failures and prolonged client disconnects, and ensures fast dissemination of events within the system. Narada is JMS [21] compliant and provides support not only for JMS clients, but also for replacing single/limited server JMS systems transparently [24] with a distributed Narada broker network. Since JMS clients are vendor [22,23] agnostic, this JMS integration has provided Narada with access to a plethora of applications built around JMS, while the integrated Narada-JMS solution provides these applications with scaling, availability and dynamic real time load balancing. Among the applications ported to this solution is the Anabas distance education conferencing system [25] and the Online Knowledge Center (OKC) portal [26] being developed at the IU Grid labs.

JXTA is a set of open, generalized protocols to support peer-to-peer interactions and core P2P capabilities including peer grouping and security. The JXTA peers, and rendezvous peers – specialized routers, rely on a simple forwarding of interactions for disseminations and rely on time-to-live (TTL) indicators and peer traces to attenuate interaction propagations. JXTA provides features such as dynamic discovery and a rich search mechanism while allowing peers to communicate through firewalls using a relay service. However JXTA interactions are unreliable, tend to be very localized and are based on simple forwarding. We intend to use the Narada system as an intelligent delivery mechanism for JXTA interactions while also providing optimizations for certain interactions. Furthermore, Narada could improve communication latencies associated with JXTA peer interactions. It is expected that existing P2P systems would either support JXTA or have bridges initiated to it from JXTA. Support for JXTA would thus enable us to leverage other existing P2P systems along with applications built around these systems. Narada's support for JXTA in addition to the support for JMS would result in interactions that are robust and dynamic while being supported by a scalable and highly available system. There are thus several benefits that can be accrued, by both JXTA and Narada, while providing support for JXTA. This paper describes the integration of P2P style interactions within Narada and optimizations/benefits that can be used by both systems. We discuss further research including the use of Narada's topic queues to reliably cache JXTA advertisements and to act as a hub for JXTA search and other services. We also discuss the role of Web Services [27,28,29,30] in our architecture.

1.0 NARADA JXTA Integration Objectives:

In our strategy for providing support for P2P interactions within Narada, we start out with the following objectives –

- Minimal or zero changes to the Narada system core and the associated protocol suites.
- We also make no changes to the JXTA core and the associated protocols. We make additions to the rendezvous layer for integration purposes. Peers do not communicate directly with the Narada and continue to interact with other peers and rendezvous peers just as they presently do.

The integration is based on the proxy model, which essentially acts as the go-between the Narada system and JXTA. This proxy, the Narada-JXTA Proxy, also functions exactly as a rendezvous peer. Narada could be viewed as a service by JXTA. The discovery of this service is automatic and instantaneous due to the Narada-JXTA proxy's integration inside the rendezvous layer. Any peer can utilize Narada as a service so long as it is connected to a Narada-JXTA proxy. Nevertheless, peers do not know that the Narada broker network is routing some of their interactions.

2.0 The interaction model:

Figure 1 outlines the protocols that comprise the JXTA protocol suite [13]. Different interactions are queued at the queues associated with the relevant layers. Each layer performs some operations including the addition of additional information. The rendezvous layer processes information arriving at its input queues from the peer-resolving layer and the pipe-binding layer. The Narada-JXTA proxy operates inside the rendezvous layer. Since the payload structure associated with different interactions is different we can easily identify the interaction types associated with the payloads.

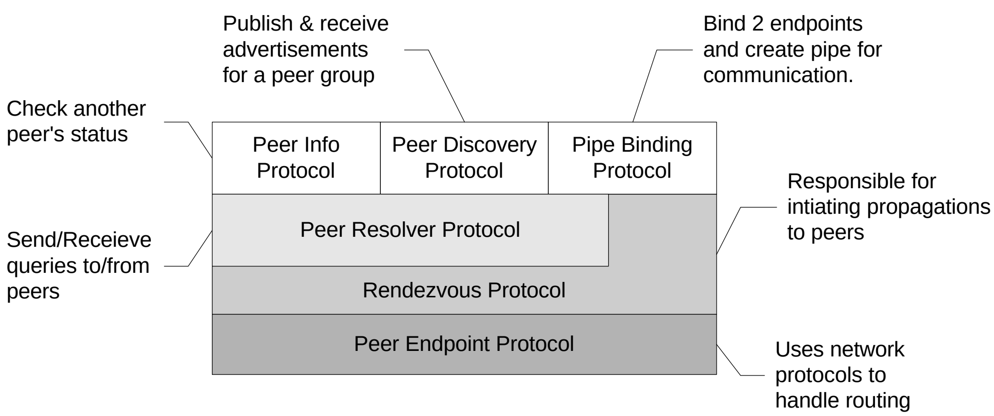

Figure 1: The JXTA protocol suite

JXTA interactions that would be routed by the Narada system are fed through the Narada-JXTA proxy, which also serves as a rendezvous peer. JXTA peers can continue to interact with each other and of course some of these peers can be connected to pure JXTA rendezvous peers. These Narada-JXTA proxies, since they are configured as clients within the Narada system, inherit all the guarantees that are provided to clients within the Narada system. Interactions pertaining to discovery/search or communications within a peer group would be serviced both by JXTA rendezvous peers and also by Narada-JXTA proxies. Interactions that peers have with the Narada-JXTA proxies are what are routed through the Narada system. The Narada-JXTA proxy serves in a dual role as both a rendezvous peer and as a Narada client providing a bridge between Narada and JXTA. Peers have multiple routes to reach each other and some of these could include the Narada system and some of them need not. Such peers can interact directly with each other during the request/response interactions. Figure 2 outlines the Narada JXTA interaction model.

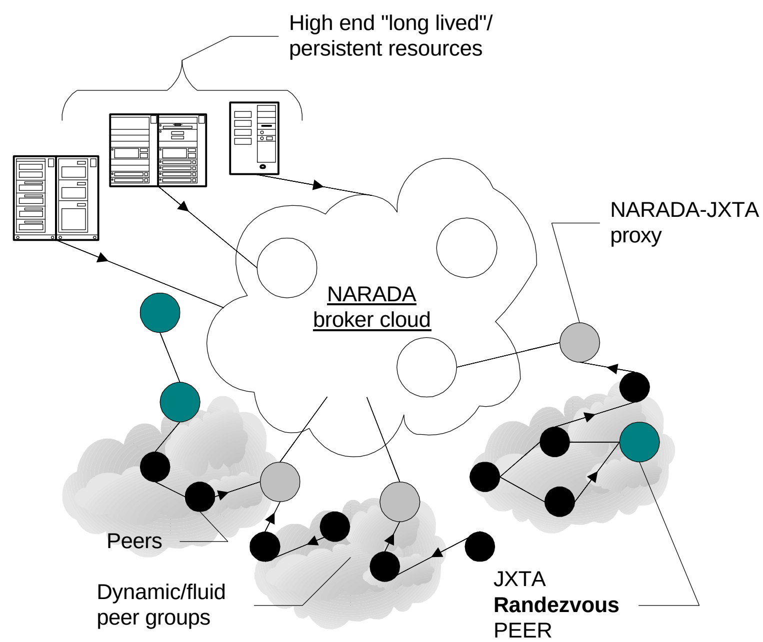

Figure 2: The Narada-JXTA Interaction Model

3.0 Interaction Disseminations

Peers can create a peer group; request to be part of a peer group; perform search/request/discovery all with respect to a specific targeted peer group. Peers always issue requests/responses to a specific peer group and sometimes to a specific peer. Any given peer would have multiple peer-id's assigned to it based on the number of peer groups that it is a part of. Each rendezvous peer keeps track of multiple peer groups through peer group advertisements that it receives, any given peer group advertisement could of course be received at multiple rendezvous peers. These rendezvous peers are then responsible for forwarding interactions; if it had received an advertisement for the peer group contained in these interactions.

To ensure the efficient dissemination of interactions, it is important to ensure that JXTA interactions that are routed by Narada are delivered only to those Narada-JXTA proxies that should receive them. This entails that the Narada-JXTA proxy perform a sequence of operations, based on the interactions that it receives, to ensure selective delivery. The set of operations that the Narada-JXTA proxy performs comprise gleaning relevant information from JXTA interactions, constructing an event based on the information gleaned and finally in its role as a Narada client subscribing (if it chooses to do so) to a topic to facilitate selective delivery. By subscribing to relevant topics, and creating events targeted to specific topics each proxy ensures that the broker network is not flooded with interactions routed by them. The events constructed by the proxies include the entire interaction as the event's payload. Upon receipt at a proxy, this payload is de-serialized and the interaction is propagated as outlined in the proxy's dual role as a rendezvous peer.

Narada-JXTA proxies are initialized first as rendezvous peers and then as Narada clients. During its initialization as a Narada client once a proxy is assigned a unique connection ID, the proxy subscribes to a topic identifying itself as a Narada-JXTA proxy. This enables the system to be aware of all the Narada-JXTA proxies that are present in the system. We now proceed to outline the sequence of operations associated with different JXTA interactions. The Narada-JXTA proxy in its role as a rendezvous peer to peers receives –

- Peer group advertisements
- Requests from peers to be part of a certain peer group
- Responses to requests from peers to be part of a certain peer group
- Messages sent to a certain peer group or a targeted peer.

3.1 Peer Group Advertisements

When peer group advertisements, propagated by a peer, are received at a Narada-JXTA proxy, the proxy creates the event depicted in figure 3. The peer group advertisement is the payload contained in this event. The proxy proceeds to initiate a subscription to the peer group with the subscription being registered to the connection that the proxy has into the Narada system. This enables the system to identify this proxy as a destination when certain interactions are targeted to that specific peer group. This peer group advertisement is delivered by the Narada system to all the Narada-JXTA proxies within the system.

Proxies that receive this advertisement do not initiate any actions and the proxy deals with this advertisement just as a JXTA rendezvous peer would.

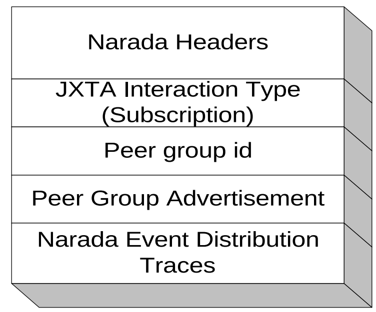

Figure 3: Event for Peer Group Advertisements

3.2 Peer Advertisements

When a peer issues a request to be part of a certain peer group, the event constructed by the proxy is depicted in figure 4.(a). The advertisement is contained in the payload, and the targeted topic contained in this event is the peer group id. The system thus propagates the event to the proxies that had subscribed to this topic. To ensure that responses to this advertisement are targeted to the proxy forwarding this request, the event also encapsulates its Narada connection information within this event. Narada-JXTA proxies receiving this event maintain the JXTA request and the connection associated with the request; to be used during propagation of responses. These Narada-JXTA proxies then behave as a normal rendezvous peer, processing the request as it normally would. This entails forwarding/routing of events en route to peers that are part of the peer group.

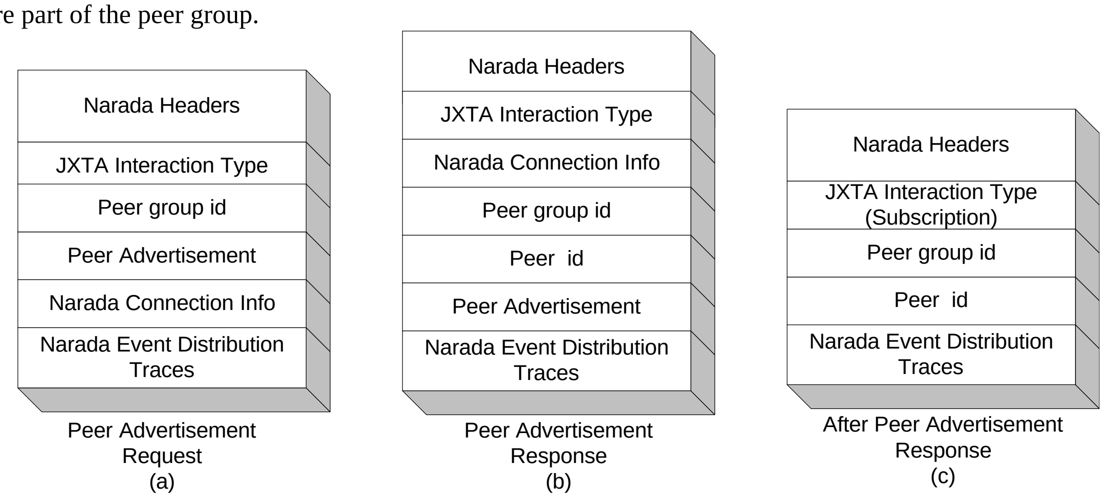

Figure 4: Dealing with requests and responses, to peers being part of a peer group

When responses, initiated by authorized peers, are received at the Narada-JXTA proxy, the proxy checks to see if there was a request associated with it and proceeds to forward the response encapsulated in a Narada event depicted in figure 4.(b). Responses include the peer-id that is assigned to the requesting peer, and this information is included in the event. The proxy also retrieves the target connection that this response should be sent to and includes it in the event. Upon receipt of this response at the initiating Narada-JXTA proxy, the proxy initiates operations in its role as a rendezvous to ensure propagation of the response to the requesting peer. The proxy then proceeds to subscribe to both the peergroup-id and the peer-id. This event is depicted in figure 4.(c), and as can be seen there is no payload contained in this event. The peergroup-id subscription ensures that interactions for the peer group that the serviced peer will be a part of are always received at the proxy; this includes advertisements to be part of that peer group. Furthermore, a lot of JXTA interactions are sometimes targeted to specific peers so we also subscribe to the peer-id contained in the received response. When events are sent to a specific peer in a peer group, the Narada system routes the event to the proxy (or proxies since a given peer can be attached to multiple Narada-JXTA Proxies/Rendezvous peers) that can route the event to the targeted peer.

3.3 JXTA Messages

When JXTA messages enter the Narada broker network, there are a few pieces of information that need to be gleaned from the message, these include

- The Peer group id/ peer id that this message is intended for.
- The unique id contained within this interaction based on the peer-id and the timestamp contained in the message.
- The interaction type.
- The hop count associated with the message.

These are the elements that the Narada system will operate on. The rest of the JXTA message needs to be serialized and would be the payload contained within the Narada event depicted in figure 5. Narada then routes this event to only those proxies that had subscribed to either the peergroup-id or the peer-id contained in this event. Upon receipt at the Narada-JXTA proxies the payload contained in the event is unmarshalled and the JXTA message is recreated. With the Narada-JXTA proxy now behaving as any other rendezvous peer, propagation from this point on proceeds as dictated by the JXTA reference implementation. Narada for its part deals with the efficient routing of the Narada events based on the topic and subscriptions (propagated during receipt of peer advertisements at the proxies).

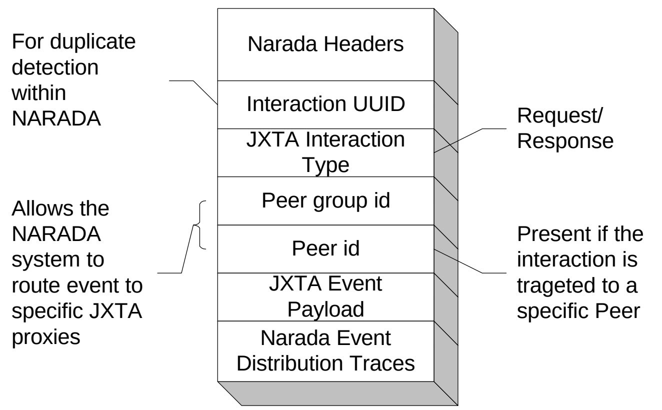

Figure 5: The Event for JXTA messages

4.0 The Duplicate detection problem:

Unlike Narada clients that interact with only one broker at any given time and never with any other client directly; peers can be connected to multiple peers and rendezvous peers. This situation leads to having the same interaction entering the Narada system through multiple points and also due to loops in peer connectivity. Under conditions of increasing loads and high peer concentrations the cumulative effects of these JXTA interactions entering the distributed broker network could entail prohibitive CPU/bandwidth costs. The most crucial thing is to ensure that the Narada broker network is not inundated with such duplicate interactions. The scheme that efficiently garbage collects interactions, which have already been serviced by the system, is described in Appendix A.

5.0 Narada JXTA Systems

5.1 Scaleable Systems

The Narada JXTA integration service scales naturally since Narada provides dynamic “long distance” support while JXTA provides for dynamic localized supports. In the combined global scale infrastructure Narada works best for “long lived” and persistent resources that would be efficiently replicated within Narada. This integrated model provides efficient search/discovery not only for static activity but also for dynamic activity while allowing JXTA interactions at the edges. The resultant system also scales with multiple JXTA Peer Groups linked by Narada, which can dynamically instantiate new brokers to balance the load. As opposed to the simple forwarding of interactions, the intelligent routing in Narada in tandem with the duplicate detection scheme in our solution ensures faster disseminations and improved communication latencies for peers. Furthermore, targeted peer interactions traversing along shortest paths within the broker network obviates the need for a peer to maintain dedicated connections to a lot of peers. Proxies, due to their initialization as clients within Narada, inherit all the guarantees accorded to clients within the Narada such as guaranteed delivery in the presence of failures and fast disseminations.

Discovery of rendezvous peers in JXTA is a slow process. A rendezvous peer generally downloads a list of other rendezvous peers from a server, not all of which would be up and running at that time. We allow for dynamic discovery of Narada-JXTA proxies, which need not be directly aware of each other, but do end up receiving interactions sent to a peer group if they had both received peer group advertisements earlier. The scheme also allows us to connect islands of peers and rendezvous peers together, allowing for a greater and richer set of interactions between these peers.

5.2 Enhanced Narada JXTA Integration

This integration of JXTA interactions within Narada leads us into strategies for further optimizing these interactions. Narada, since it is aware of advertisements, would act as a hub for services such as search and discovery. These advertisements when organized into "queryspaces" allow Narada to respond to JXTA search more efficiently. These services would be provided for resources whose advertisements Narada is aware of. P2P systems replicate resources in an ad hoc fashion, the availability of which is dependent on the peer's active digital presence. Narada could be used to manage these resources and the replication strategy within Narada ensures that the resources are "closer" and "available" at locations with a high interest in that resource. Furthermore, since peer interactions and resources in the JXTA model are traditionally unreliable, we could use Narada's guaranteed delivery properties in tandem with its JMS compliance to provide a standards based notion of reliable peers, interactions and resources.

5.3 Web Services and Applications

A typical application of the hybrid Narada/JXTA technology is distance education. This often consists of multiple linked classrooms where the participants in each classroom are individually linked to collaborative environment. Here a peer-to-peer model (such as JXTA) can be used in a classroom to give fast dynamic response to shared input control while the JMS style global Narada capability is used to multicast between classrooms. More generally this combination of global structured and local dynamic messaging scales to support general applications. We can package the JXTA/JMS/Narada hybrid as a event or messaging web service whose different modes could correspond to different ports in WSDL [27]. These ports trade off scaling, performance and reliability in a fashion that can be chosen by the user. Alternatively web services communicate via channels and we could use the technology of this paper to flexibly link different services together with dynamic choice of service provider. Other applications we are pursuing include workflow where administrative agents control the traffic between different web services. We have explained in more detail how one can base a P2P Grid architecture on a generalized publish/subscribe mechanism in [8].

6.0 Conclusion

In this paper, we presented our strategy for providing support for JXTA interactions. We also described the benefits that can be accrued by clients from these systems. It is our belief that this integration, to go along with the JMS integration, has added considerable value to Narada and that we are well positioned to Web Service "enable" Narada. We are currently gathering performance numbers for the JXTA based integration, which would contrast the latencies in pure JXTA environments with those in Narada's integrated environment. The final version of this paper would incorporate these comprehensive results.

References

1. D.J. Watts and S.H. Strogatz. Collective Dynamics of Small-World Networks. *Nature*. 393:440. 1998.
2. R. Albert, H. Jeong and A. Barabasi. Diameter of the World Wide Web. *Nature* 401:130. 1999.
3. The Narada Event Brokering System http://grids.ucs.indiana.edu/ptliupages/projects/narada/
4. Geoffrey Fox and Shrideep Pallickara, An Event Service to Support Grid Computational Environments, to be published in *Concurrency and Computation: Practice and Experience*, Special Issue on Grid Computing Environments.
5. Geoffrey C. Fox and Shrideep Pallickara, An Approach to High Performance Distributed Web Brokering. *ACM Ubiquity* Volume2 Issue 38. November 2001.
6. Pallickara, S., "A Grid Event Service." PhD Syracuse University, 2001.
7. Geoffrey C. Fox and Shrideep Pallickara .The Narada Event Brokering System: Overview and Extensions Submitted to the *2002 International Conference on Parallel and Distributed Processing Techniques and Applications* (PDPTA'02).

8. Geoffrey Fox, Ozgur Balsoy, Shrideep Pallickara, Ahmet Uyar, Dennis Gannon, Aleksander Slominski. Community Grids. To appear in the proceedings of the *International Conference on Computational Science* (ICCS 2002). Amsterdam, Netherlands April 2002.
9. Geoffrey Fox, "Peer-to-Peer Networks," *Computing in Science & Engineering*, vol. 3, no. 3, May2001.
10. openp2p P2P Web Site from O'Reilly http://www.openp2p.com.
11. "Peer-To-Peer: Harnessing the Benefits of a Disruptive Technology", edited by Andy Oram, O'Reilly Press March 2001.
12. Sun Microsystems. The JXTA Project and Peer-to-Peer Technology http://www.jxta.org
13. The JXTA Protocol Specifications. http://spec.jxta.org/v1.0/docbook/JXTAProtocols.html
14. SETI@home Project http://setiathome.ssl.berkeley.edu.
15. Folding@home Project http://www.stanford.edu/group/pandegroup/Cosm
16. Jabber http://www.jabber.org
17. Gnutella. http://gnutella.wego.com
18. Groove Network http://www.groove.net
19. Antony Rowstron and Peter Druschel. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. Proceedings of Middleware 2001.
20. Ken Arnold, Bryan O'Sullivan, Robert Scheifler, Jim Waldo and Ann Wollrath. The Jini Specification. Addison-Wesley. June 1999.
21. Mark Happner, Rich Burridge and Rahul Sharma. Sun Microsystems. Java Message Service Specification. 2000. http://java.sun.com/products/jms
22. SonicMQ JMS Server http://www.sonicsoftware.com/
23. The OpenJMS Project http://openjms.exolab.org/
24. Geoffrey C. Fox and Shrideep Pallickara JMS Compliance in the Narada Event Brokering System. Submitted to the *2002 International Conference on Internet Computing* (IC-02).
25. The Anabas Conferencing System. http://www.anabas.com
26. The Online Knowledge Center (OKC) Web Portal http://judi.ucs.indiana.edu/okcportal/index.jsp
27. Web Services Description Language (WSDL) 1.1 http://www.w3.org/TR/wsdl.
28. Presentation on Web Services by Francesco Curbera of IBM at DoE Components Workshop July 23-25, 2001. Livermore, California. http://www.llnl.gov/CASC/workshops/components2001/viewgraphs/ FranciscoCurbera.ppt
29. Frank Laymann (IBM), Web services Flow Language WSFL, http://www.ibm.com/software/solutions/webservices/pdf/WSFL.pdf
30. Harumi Kuno, Mike Lemon, Alan Karp and Dorothea Beringer, (WSCL – Web services Conversational Language), "Conversations + Interfaces = Business logic", http://www.hpl.hp.com/techreports/2001/HPL-2001-127.html
31. Semantic Web from W3C to describe self organizing Intelligence from enhanced web resources. http://www.w3c.org/2001/sw/
32. Berners-Lee, T., Hendler, J., and Lassila, O., "The Semantic Web," Scientific American, May2001.
33. Paul J. Leach and Rich Salz. Network Working Group. UUIDs and GUIDs. February, 1998.

Appendix A: (Garbage collection of duplicate interactions)

JXTA interactions, when they are propagated into the Narada system, are encapsulated inside events outlined in earlier sections. Unlike advertisements, JXTA messages do not have an UUID **[33]** (IETF specification guaranteeing uniqueness until 3040 A.D.) associated with them. These JXTA messages do have the peer id (the source) and the timestamp in milliseconds, based on the system time at the node hosting the peer, associated with them. These are used for computing the unique id associated with the constructed event. In the case of advertisements the UUID is used as the identifier within the constructed event. Thus every Narada-JXTA Proxy, where interactions are received, computes the same unique ID for the generated event to be used only within the Narada broker network. Destinations computed for duplicate events are identical; when these events are being routed within the Narada network the route they take is more or less the same, i.e. they propagate through the same cluster controllers. Paths traversed by later events (duplicates) en route to final destinations are more or less the same as the event that traversed prior to it. These paths only vary during sudden network changing conditions such as failures of brokers/links and spikes in network usage over sustained durations in which case the paths computed based on network costs would vary. Even in such cases subsequent duplicates continue to traverse along identical computed paths. Having brokers keep track of the events (event id's specifically) they received and discarding duplicates allows us to solve the duplicate detection problem and prevent the network from expending network cycles for routing duplicates. This scheme allows for faster disseminations, with a survival of the fittest scheme, where duplicate events are discarded if they were not the first to arrive at a broker.