The Case for an Adaptive Integration Framework for Data Aggregation/Dissemination in Service-Oriented Architectures

Dennis M. Moen and Lynn Meredith-Lockheed Martin
dennis.moen@lmco.com lynn.meredith@lmco.com

Abstract

The migration to Service Oriented Architectures (SOA) implies many real-time applications distributed across large geographic areas with highly mobile users and sensors that require exchange of critical data among local as well as distant users across resource constrained networks. These emerging applications can be characterized as distributed collaborative adaptive systems. They are likely to rely on ad hoc wireless networks particularly in military and emergency response applications for transport of critical information and in many cases in multimedia form. Users of these systems are likely to have different needs or views of sensor data either because of organizational role or geographic location. In this distributed architecture, available resources must dynamically reconfigure themselves to respond to external factors such as changes in the environment, changes in short-term objectives, reallocation of responsibilities, and changes in information flow patterns.

This paper describes a framework for dynamic resource management (DRM) and Quality of Service (QoS) in support of network aware applications and resiliency in ad hoc delay tolerant networking (DTN). The proposed framework is based on managing per-flow, end-to-end provisioning of heterogeneous network resources in support of mission-driven resource management.

1. Introduction

Many real-time applications distributed across large geographic areas and highly mobile users and sensors are emerging that require exchange of critical data among local as well as distant users across resource constrained networks [1]. These emerging applications can be characterized as distributed collaborative adaptive systems. They are likely to rely on ad hoc wireless networks particularly in military and emergency response applications for transport of critical information and in many cases in multimedia form. Users of these systems are likely to have different needs or views of sensor data either because of organizational role or geographic location.

A broad range of standards, technologies, and products are used to accommodate this digital communication. For example, typical communication services include email, chat, web portals, voice, video, collaboration, command and control, situational awareness messaging, fire control messaging, and file transfers to name a few.

Historically, communications among systems were all preplanned and communications links were statically allocated. The applications to which military systems are being put today are more dynamic than they were in the past. System configurations and missions change more rapidly than the applications built to support them. The applications thus have to be more flexible and accommodating about the combinations of capabilities they will be asked to provide complicated by the nature of disruptive networking environments. It is no longer possible to simply dedicate resources to each application on the assumption that no other application or network constraint will interfere with its use. In the distributed architectures of the future, available resources must dynamically reconfigure themselves to respond to external factors such as changes in the environment, changes in network services, changes in organizational objectives and responsibilities, and changes in information flow patterns.

The concept of adaptive integration framework for data dissemination/aggregation supported with dynamic resource management is emerging as customer requirements are becoming more clearly defined. For example, there are several different types of resources that can be allocated; processors, memory, disk space, network bandwidth, communication channels, and middleware services (e.g., publish-subscribe services, CORBA event channels, topics in the Java Message Service) [2][3][4]. The technologies, standards, and products for resource allocation are specific to a particular type of resource. For example:

- Middleware. Message oriented middleware (e.g., Data Distribution System (DDS), Java Messaging Service (JMS)), Tuxedo, CORBA, J2EE, and Service Oriented Architectures.
- Operating Systems. Windows, Solaris, MacOS, Real-Time Operating Systems, Trusted Operating Systems.
- Networking. Differentiated Services (DSCP), Integrated Services (RSVP and its companion Aggregate RSVP), Traffic Engineering, MPLS-TE, OSPFv3, router capabilities (e.g., Cisco policies).

In this paper, we propose a new strategy to facilitate the exchange of information in real time distributed applications. Our approach will support the acquisition of data and information from sensors and information networks, provide storage and distribution of acquired information, and enable information fusion and decision support capabilities. The approach allows for establishment of core ontologies in the information model, application of business rules for information accumulation and distribution, and new communication mechanisms (overlay networking and non-IP protocols).

In the next section of this paper, we provide background information and our motivation for the research work. In section 3, we present our ideas on a new way of thinking about the protocol stack and interworking vertically across the protocol layers.

In section 4, we describe the need for resilient networking and proposed strategies for implementation. Section 5 provides our approach to an adaptive information framework and we conclude with a summary of our proposed future work.

2. Background

Next generation networking and applications that are Service Oriented Architecture based combine advanced technologies in several layers. The implication is that new ideas are required on how transparency can be achieved with Quality of Service for putting relevant data and information into immediate relationships for highly mobile and demanding user environments. There are several efforts underway looking at these issues. The Navy Open Architecture [5] provides expectations for dynamic resource management. Here, network capacity constraints imply the need for allocation in a mission-driven fashion.

Another example, the Global Information Grid End-to-End Quality of Service (GIG E2E QoS) effort is also working on strategies for improved QoS. Their current strategy focuses on use of network layer protocols such as Multiprotocol Label Switching (MPLS) and Differentiated Services Code Point (DSCP), but the long term goal is a policy-based approach using the Net Centric Services for Enterprise Management [6].

There are a number of other efforts within the Department of Defense focused on improved QoS. These include WIN-T, Joint Tactical Radio System (JTRS), and Transformational Communications Satellite (TSAT).

There is a clear focus across all these programs on "end-to-end quality of service (E2E QoS)." Early research sponsored by the Office of Naval Research at the University of Michigan [7] focused on QoS-optimization algorithms and communication subsystem architectures that satisfy E2E QoS. This effort proposed a strategy where each client establishes a contracted QoS, while adapting gracefully to transient over-load and resource shortage. The research introduced a new concept of flexible QoS contract, specifying multiple acceptable levels of service and their corresponding rewards for each client.

Lockheed Martin in conjunction with the University of California, Irvine, developed early concepts for an architectural pattern called Quality Connector (QC) [8]. This concept is a meta-programming technique that enables applications to specify the QoS they require from their infrastructure, and then manages the operations that optimize the middleware to implement those QoS requirements.

Floyd [9] introduced the idea of End-to-End congestion control in the Internet. For the problem at hand, to provide a clear definition of this requires that the endpoints be clearly defined. Each customer defines endpoints differently, which causes the resources between (or among) the endpoints to be different. For example, if the endpoints are "point-of-presence" routers on the GIG, then networking technology is what is needed. In other cases, as we develop in this paper, endpoints are user devices (clients, servers, or more generally peer-to-peer devices).

3. New Model for Protocol Interworking

Traditional communications and networking models assign boundaries based on application of the networking protocol stack. This approach has served the community well in defining and developing the application of Internet protocols and communications systems in legacy wired networks. We are now challenged with introducing new services to support environments that are characterized by mobility with less dependency on wired communications.

Increased efficiencies are likely to be gained only through considerations of the interworking vertically and realizing the impact of choices made at one layer on another. An example is the formation of ad hoc wireless networking for randomly deployed sensor fields. Each node in such a network will need to optimize power consumption, spectrum use, network protocol, routing algorithm, data format, and capacity throughput with the overall needs of the sensor network to efficiently report events, yet maintain viability/life as long as possible before complete consumption of resources such as available electrical power.

We propose a strategy based on consideration for optimization across three layers of activity: access or network layer, data format, and application as indicated in figure 1. The reason for this approach is that careful management of resources is critical to achieving better performance in ad hoc networks and even in high performance wired networks in the presence of congestion. However, managing resources is challenging because today end-hosts (applications) cannot directly observe or even control network resources. In addition, data formats are typically chosen to meet the needs of the application, not necessarily considering the optimum choices for networking. An example of this is presented by Morse [10]. In the example, Morse indicates a nearly 30 times increase in traffic load when using a pure Web service protocol versus streaming data across an overlay multicast channel without the use of the Web service protocol overhead.

Interaction between layers is essential and requires a flexible architecture with a very heterogeneous approach accounting for full integration of wired and wireless networking at the lowest layer with data formats and application needs at the higher layers. The growing complexity of information exchange using a variety of protocols, interworking network elements and robust distributed applications envisioned by Service Oriented Architectures (SOA) is very challenging. The interaction of these components significantly influences quality of service (QoS), application performance, robustness, and reliability.

4. Resilient Networking

Nodes in ad hoc wireless network can communicate directly with another node located within its radio transmission range. To communicate with the node outside of its communication range, a sequence of intermediate nodes in ad hoc networks is required to

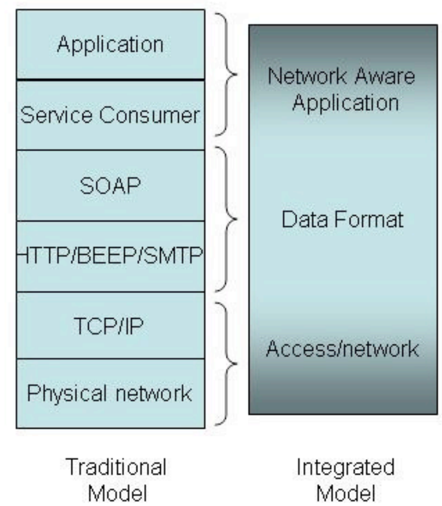

Figure 1. Optimized activity layers

relay messages on behalf of this node, resulting in multi-hop wireless network. The mobility of nodes in the ad hoc network causes the nodes to be in and out of range from one another; therefore, the connectivity varies dynamically with time, power, interference, and other factors.

This dynamic connectivity imposes major challenges for the network layer to determine the multi-hop route between a given pair of source and destination nodes. The traditional routing techniques such as distance vector and link state (proactive protocols) that are used in fixed networks cannot be directly applied to ad hoc networks. First, although they do adapt dynamically to the changes of network topology, they are not designed for this kind of dynamics; second, the periodic updates of routing tables waste a large portion of the scarce capacity in the ad hoc network [11].

At the same time that more and more information exchange moves to wireless networking, organizational information environments are moving to SOA and shifting to "network aware" applications. Because users cannot all be served at predictable and satisfactory levels of quality [12], applications have to be able to adapt to changes in networks. These new applications that are designed to adapt to network conditions are called "network aware" [13]. As indicated in figure 2, these network aware applications provide the opportunity to enable additional host level network services that are not available at lower layers of network protocols (TCP/IP).

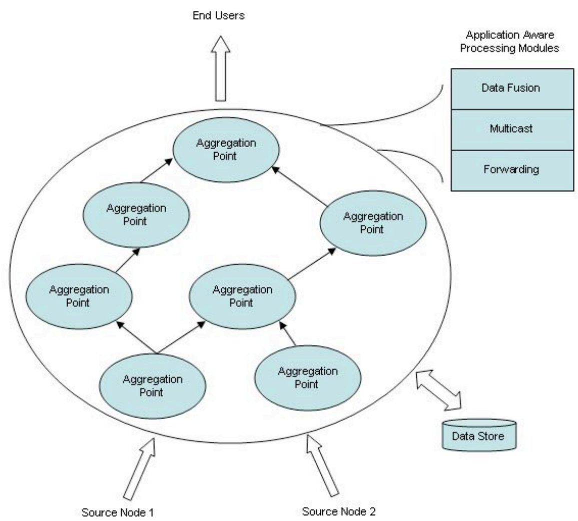

Figure2. Application aware networking

A number of approaches such as peer-to-peer (P2P) systems provide load balancing, content routing, or dynamic selection among multiple paths. These approaches move the route selection functionality to the application or transport layer, through the use of overlay networks of cooperating end systems. Application layer approaches are also useful in the case of multicasting when multicast is not otherwise available [14]. This strategy enables consideration for end-to-end performance management and supports, as Savage suggests [15], use of informed transport protocols.

Constraint based routing comprises both policy and QoS routing. Policy routing is important for providing better and more flexible services. Internet routing involves two concepts: routing protocols and routing algorithms. Routing protocols capture information about network resources and routing algorithms use the information to compute paths based on relatively static measures [16].

QoS routing includes considerations for application requirements as well as the availability of network resources. However, this implies additional needs for managing routing such as dissemination of dynamic information and more complex computations for route path determination. Results of research at George Mason University [17] proposed the idea of using only local information for node performance measurements for use in calculating paths in overlay networks as a strategy for simplifying the computations.

The concept is to consider application layer implementation of a constraint based algorithm based on network performance information locally available without extensive information exchange between the local node and distant nodes in the network. While this approach may use less accurate information from an overall network perspective, it significantly reduces network overhead associated with routing protocols and enables consideration for new approaches for routing algorithms.

In our current research we are investigating the use of bio-inspired routing algorithms (swarm intelligence) enabled by the use of local node performance information. We propose to use these algorithms to aid the provisioning of resilient networks for support of network aware applications over disruptive network environments such as those presented by ad hoc mobile networks found in military and emergency response operations.

5. QC Adaptive Integration Framework

A flexible, efficient approach for the deployment of QoS-sensitive applications in networks should facilitate the monitoring of the QoS received by an application and allow easy deployment of application aware processing at intermediate nodes of the network. This is especially important at any instant during an operation where military commanders need to know if their forces can communicate with one another and have the information needed to accomplish their mission objectives. If only a portion of the force can communicate because the resources are overloaded then the commanders need to decide how to reallocate resources among their forces.

The problem is that there is no straightforward method for operational personnel or system architects to describe how resources should be allocated in an environment that is complex, dynamic, information-rich, and time-sensitive. The problem is complicated by changing operational conditions as well as the wide variety of technologies, standards, and products from which systems are assembled.

Our approach is an infrastructure for dynamic resource management (DRM) and QoS within an adaptive integration framework we call the Quality Connector. The Quality Connector (QC) is a framework for managing per-flow, end-to-end provisioning of heterogeneous network resources in support of mission-driven resource management. Our strategy is to apply the QC as policy enforcement points integrated into the networking components and controlled via policies supplied by the control plane.

Figure 3 presents our QC adaptive integration framework. Ovals represent run-time components of the QC and the rectangles represent externally supplied

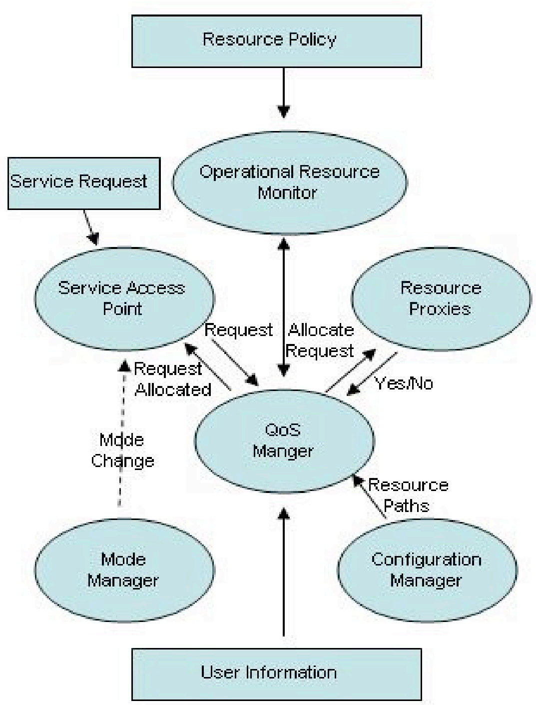

Figure 3. QC adaptive integration framework

information. The Service Access Points can be thought of as a set of policy enforcement points that sit between applications and the underlying infrastructure resources. The infrastructure resources can be message oriented middleware like a CORBA event service, a network hub, a router, or even a CPU.

The assumptions for the QC adaptive integration framework are:

- Systems will increasingly be composed from commercial and government off-the-shelf (COTS/GOTS) products.
- End-to-end resource allocation and QoS can be described in technology and product neutral terms.
- End-to-end QoS descriptions can be sensitive to system modes.
- It is possible to map these specifications onto the middleware technologies and product(s) chosen by different system architects.

To implement any form of resource allocation the technologies and products chosen by system architects must provide an interface that allows them to be controlled. Our early efforts focused on allocation of network resources and in some cases supported various aspects of QoS. For example, QC prototypes have used the following technologies: Differentiated Services (DiffServ), Real-Time Event Channels (RTEC), and Fault Tolerant Real-Time Event Channels (FTRTEC). Most of these prototypes were written in C++ and Java, and utilized real-time CORBA technologies (e.g., TAO v1.4).

More recently, we demonstrated prioritized resource allocation for the Java Messaging Service (JMS). This QC prototype managed resources for over 800 clients simultaneously. The prototype, implemented in Java, controlled resource allocation for several JMS topics running on multiple hosts. An administrator created sets of resource allocation policies, specified using technology-neutral Information Exchange Requirements (IERs), and associated the sets with individual user roles. Included are detailed performance measurement statistics.

Our approach to operational resource allocation utilizes the following pieces of information:

- *Users*. The primary focus is on operational personnel but could include machine-to-machine.
- *Roles*. User roles defined in a registry service.
- *Infrastructure Services*. Email, chat, data distribution service (DDS), Java Messaging Service, CORBA Event Service, etc.
- *Information Exchange Requirements (IERs)*. We assume that IERs will be defined during the system architecture process and contain enough information to allocate resources.
- *Modes*. We assume that modes are also defined during the system architecture process.

6. Conclusions and Future Work

We are developing capabilities that focus on new approaches to quality information exchange that recognize the need for optimization of services across three layers of activity: access or network layer, data format, and application. The driving force is the shift to SOA strategies for information processing and distribution in networking dominated by highly mobile users using ad hoc networking infrastructure. In this new environment, management of resources with QoS agreements is critical.

Our early prototyping has shown that our proposed QC framework represents a promising approach for increasing the efficient use of network resources and delivering managed QoS services for users. The QC prototype managed resources for over 800 clients simultaneously. The prototype, implemented in Java, controlled resource allocation for several JMS topics running on multiple hosts. Our plans for future work include:

- Investigate how to apply a QC implementation with other resource management approaches.

- Implement and study different strategies for resource allocation and network routing.
- Extend the QC beyond resource allocation to support other QoS aspects such as latency, jitter, fault tolerance, and security.
- Develop a simulation to drive the resource proxies and mode changes for use in composing complex systems and test them under different operating modes or with different resource configurations.

Authors

Dennis M. Moen is a research scientist with the Advanced Technology group of Lockheed Martin Maritime Systems and Sensors. He is a graduate of the United States Military Academy. He received his M.S. degree in electrical engineering from the University of Arizona and his Ph.D. in Information Technology from George Mason University. He is a licensed Professional Engineer and a Senior Member of the IEEE. His research interests are network performance modeling and survivable network architectures.

Lynn Meredith is a Lockheed Martin Fellow and has over 25 years of experience as a systems engineer and software systems architect in the context of both research and development, and full scale development. He has worked on systems and software architecture for many different operational systems. Mr. Meredith also worked several advanced research projects including DARPA programs in program composition for distributed systems, fault tolerant networking, coalition information management, information assurance, and intelligence integration. He has also worked on high speed secure networking technologies for the Department of Energy (DOE).

References

[1] Li, D., K. Wong, Y.H. Yu, and A. Sayeed, "Detection, Classification and Tracking of Targets in Distributed Sensor Networks," *IEEE Signal Processing Magazine*, Vol. 19 Issue 2. Mar 2002.

[2] Patterson, D., T. Anderson, N. Cardwell, R. Fromm, K. Keeton, C. Kozyrakis, R. Thomas, and K. Yelick. "A Case for Intelligent RAM," *IEEE Micro*, Vol. 17 Issue 2, Apr. 1997.

[3] Eggers, S., J. Emer, H. Levy, J. Lo, R. Stamm, and D. Tullsen, "Simultaneous Multithreading: A Platform for Next-generation Processors," *IEEE Micro*, Vol. 17, Issue 5, Sept 1997.

[4] Patterson, D., G. Gibson, and R. Katz. "A Case for Redundant Arrays of Inexpensive Disks (RAID)," *Proceedings of the International Conference on the Management of Data*, pages 109-116, June 1989.

[5] Navy PEO IWS, "Open Architecture Computing Environment Technologies and Standards," Version 1.0, Aug 2004.

[6] Defense Information Systems Agency, http://www.disa.mil/nces/core_enterprise_services/enterprise_service_management.html, February 2008.

[7] Tarek Abdelzaher and Kang G. Shin. "End-Host Architecture for QoS-Adaptive Communication," http://www.cc.gatech.edu/classes/AY2001/cs6235_spring/papers/00683195.pdf., February 2008.

[8] Joseph Cross and Douglas Schmidt. "Quality Connector: An Architectural Pattern to Enhance QoS and Alleviate Dependencies in Distributed Real-time and Embedded Middleware," http://www.cs.wustl.edu/~schmidt/PDF/QualConnPat.pdf, February 2008.

[9] Sally Floyd and Kevin Fall. "Promoting the Use of End-to-End Congestion Control in the Internet," May 3, 1999. http://www.icir.org/floyd/end2end-paper.html.

[10] Morse, Katherine L., Ryan Brunton, J. Mark Pullen, Priscilla McAndrews, Andreas Tolk, and James Muguira, "An Architecture for Web Services Based Interest Management in Real Time Distributed Simulation," *Proceedings of the Eighth IEEE Distributed Simulation--Real Time Applications Workshop*, 2004.

[11] J. Broach, D.A. Maltz, and D.B Johnson, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks," Internet Draft, December, 1998.

[12] Bolliger, J. "A Framework for Network-Aware Applications," ETH Zurich Institute of Computer Systems, 2000.

[13] Cao, Jinwei, Dongsong Zhang, Kevein M. McNeill, and Jay F. Nunamaker, Jr. "An Overview of Network-Aware Applications for Mobile Multimedia Delivery," Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.

[14] Moen, Dennis M., J. Mark Pullen and Fei Zhao, "Implementation of Host-based Overlay Multicast to Support Web Based Services for RT-DVS," *Proceedings of the Eighth IEEE Distributed Simulation--Real Time Applications Workshop*, 2004.

[15] Savage, Stefan, Neal Cardwell, and Tom Anderson, "The Case for Informed Transport Protocols," *Proceedings of the Seventh Workshop on Hot Topics in Operating Systems*, March 1999, pg 58-63.

[16] Kuipers, Fernando and Peit Van Mieghem, "An Overview of Constraint-Based Path Selection Algorithms for QoS Routing," *IEEE Communications Magazine,* Dec 2002, pg 50-55.

[17] Pullen, J Mark, and Dennis M. Moen, "Overlay Multicast for Real-Time Distributed Simulation," George Mason University Technical Report, DMSO Contract NBCH00-02-D-0037, May 2005.