Poster: Providing Quality of Service in Wide-Area Publish/Subscribe Systems

Kyriakos Karenos, Minkyong Kim, Hui Lei, Hao Yang, Fan Ye
IBM T. J. Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532
Email: {kkarenos,minkyong,hlei,haoyang,fanye}@us.ibm.com

ABSTRACT

In recent years, the industry has witnessed an increasing trend of exploiting the publish/subscribe paradigm for integrating and expanding enterprise application domains over large geographic areas. The quality of the publish/subscribe service plays a critical role in the overall system performance as perceived by the end users. In this work, we present an overlay-based publish/subscribe system, called Harmony, that can manage the end-to-end QoS in latency, throughput and availability, based on the application requirements. This is achieved through a holistic set of overlay routing mechanisms that actively exploit the diversity in the network paths and direct the traffic over links with high quality, based on the feedback from a distributed network monitoring component. Harmony also provides highly available services with minimal traffic overhead, through carefully establishing multiple paths and backup brokers. We have validated these capabilities in the context of a national airspace surveillance system, where a large volume of real-time aircraft tracking data from different sources is transported among various federal agencies located throughout the US continent.

1. INTRODUCTION

In recent years, the industry has witnessed an increasing trend of exploiting the publish/subscribe (pub/sub) paradigm for integrating and expanding the enterprise application domains. For example, most large enterprises are revolutionizing the way they interconnect different branches and offices to accommodate business practices like mergers, acquisitions, off-shoring and outsourcing. Multiple organizations, such as government agencies, may form an alliance by opening up their respective domains for service reusing and data sharing among major stakeholders. There is also a proliferation of cyber-physical systems that significantly extend traditional enterprise service domains with massive sensing and actuation capabilities deployed in the physical world. Examples of such systems include national airspace surveillance, emerging Smart Grid for energy distribution, smart city management and intelligent transportation.

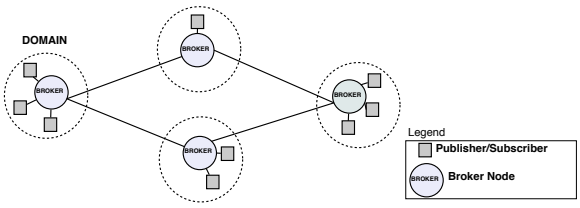

Figure 1: A broker overlay network for wide-area publish/subscribe system

Providing predictable Quality-of-Service (QoS) in the underlying pub/sub communication is an essential requirement for these mission-critical applications. The pub/sub system must be able to ensure timely and reliable delivery of critical messages, such as emergency alerts or real-time control commands. Formally stated, our goal is to provide QoS-aware publish/subscribe service in terms of *message latency, delivery rate and service availability between all matching pairs of publishers and subscribers*. This is challenging because these application typically span a large geographic area, and the wide-area networks are inherently dynamic and unreliable and cannot provide any assurance in end-to-end performance.

We present the design and validation of Harmony, an industry-strength pub/sub system developed at IBM Research. One salient feature of Harmony is the holistic provisioning of dependable and predictable QoS by effectively addressing system and network dynamics, heterogeneity and failure conditions. It allows the specification of required QoS properties (i.e., latency, message rate and service availability) for each message topic or connection session, and transports messages across autonomously administered domains respecting the above requirements end-to-end. More details of the Harmony system and our technical approach can be found in [3].

2. DESIGN

Harmony provides the QoS-aware publish/subscribe service through a wide-area broker overlay network. As illustrated in Figure 1, each local domain sets up one or multiple brokers, and these broker collectively form an overlay network that can route messages among publishers and subscribers from different domains. Such an overlay-based approach has several benefits in large-scale federated systems. It facilitates the autonomous management of constituent domains

and improves the system scalability by avoiding the maintenance of pair-wise connections. It also accommodates heterogeneous systems by allowing each domain to employ a different set of protocols and implementations to connect its own endpoints internally.

The end-to-end QoS is achieved in Harmony through a holistic set of overlay network techniques for handling network dynamics and various types of failures. Harmony actively exploits diversity in the network paths and directs messages over those links with high quality, e.g., low latency and high bandwidth. In order to cope with network dynamics, Harmony continuously monitors the link quality and adapts the routes whenever their quality deteriorates below the application requirements. It also provides highly available services with minimal traffic overhead, through carefully establishing multiple paths and backup brokers. Furthermore, it manages the network resources through laxity-based message scheduling over each overlay link, by breaking the end-to-end latency requirement into per-hop latency budgets.

While the general approach of overlay routing and scheduling has been used in other system, there are several nontrivial challenges in the context of wide area pub/sub systems. First, how can we establish overlay routes from all publishers to all subscribes of a given topic in a distributed and scalable manner? Second, how do we provide high availability in face of overlay link or broker failures, which may even disconnect all the publishers and subscribers in a domain from the overlay network? Third, how should different brokers coordinate their scheduling decisions to ensure the end-to-end latency?

Harmony uses several novel techniques to address these challenges. First, we develop a model-informed multipath computation algorithm to find overlay routes with the desired latency and resiliency performance. The basic idea is to compute the K-shortest paths using measurements from the monitoring component, find multi-path combinations that satisfy the latency requirement, then pick the combination that maximizes the aggregated resiliency. The multi-path combination ensures the resiliency against failures of overlay links or forwarding brokers. Note that a brute-force search has exponential computational complexity due to its combinatorial nature. We use a heuristic search based on the branch-and-cut technique, by constructing a route selection tree and discard a branch once its resiliency drops below the required level. After the paths are computed, the routes are installed via a signaling protocol, and data tunnels are created between neighboring brokers to transport the streaming messages [2].

Second, we provide highly available services for an application endpoint by connecting it to multiple brokers, one as the primary (typically the local broker) and others as backup. This prevents the failure of a broker from disconnecting all the subscribers and publishers in its local domain. There is an intrinsic trade off between availability and overhead in selecting the backup brokers, because brokers may have local subscribers and publishers on overlapping topics. We have formulated this as an optimization problem that takes into account such overlapping, solved the problem using integer programming and applied the results to backup broker configuration.

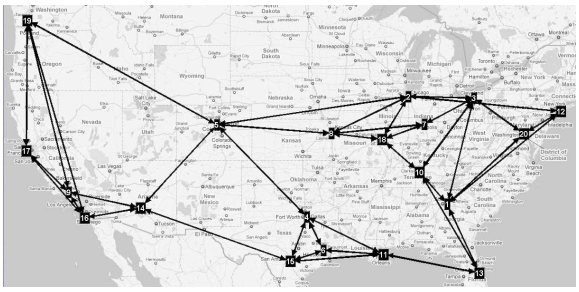

Figure 2: Network topology in an exemplary Harmony deployment

Third, we develop a distributed algorithm to divide the end-to-end latency requirement into latency budgets at each hop along a route, and leverage a laxity-based message scheduler (TransFab [1]) to decide the order of message transmission at each broker. The basic idea is to divide the latency margin, the difference between the delay requirement and the current end-to-end delay among all brokers. Thus each broker will have some "buffer" to absorb sudden latency increases, provided they are small enough compared to the margin. In addition, a message with a smaller latency margin will have higher priority when it competes with other messages in the transmission queue of an intermediate broker. Thus multiple brokers can coordinate their scheduling decisions through the budget allocation.

We have implemented the Harmony system within the IBM Websphere Message Broker (WMB), a leading industry messaging platform, and successfully deployed it in a pilot project in the context of national airspace surveillance. An example of the Harmony deployment scenario is illustrated in Figure 2. In this scenario, federal agencies located throughout the US subscribe to the real-time airspace surveillance data which is published from various sources, such as ground radars, airborne beacons and air control centers. These surveillance data provide real-time information (identity, longitude, latitude, height, flight plans, etc) of all aircrafts within the US airspace, and the message delivery is subject to very stringent latency and availability requirements posed by mission-critical applications such as threat analysis and response. The validation has confirmed Harmony's capability to provision holistic QoS in wide area pub/sub systems.

3. REFERENCES

[1] Astley, M., Bhola, S., Ward, M., Shagin, K., Paz, H., and Gershinsky, G. Pulsar: A resource-control architecture for time-critical service-oriented applications. *IBM Systems Journal 47*, 2 (April 2008).

[2] Chow, A., Yang, H., Xia, C., Kim, M., Liu, Z., and Lei, H. Ems: Encoded multipath streaming for real-time live streaming applications. In *IEEE ICNP (to appear)* (2009).

[3] Dube, P., Halim, N., Karenos, K., Kim, M., Liu, Z., Parthasarathy, S., Pendarakis, D., and Yang, H. Harmony: Holistic messaging middleware for event-driven systems. *IBM Systems Journal 47* (2008).