2009 IEEE International Symposium on Parallel and Distributed Processing with Applications

Analysing Internet Information Propagation using Quality of Service Methodology based on Algebra of System

Wang Tianju
Department of Industrial Engineering, Tsinghua University
Beijing, China
wtshootj@gmail.com

Qu Qu
Department of Industrial Engineering, Tsinghua University
Beijing, China
ququjoy@gmail.com

Benjamin Koo
Department of Industrial Engineering, Tsinghua University
Beijing, China
koo0905@gmail.com

***Abstract*: In this paper, we analyze the basic characteristics of the propagation network. Using the modeling framework Algebra of System (AoS), we convert the complexity of the network structure into the quantity of data. By using OPN(Object-Process Network) which is an executable meta-language based on the mathematical specification of AoS, we build a framework to describe and evaluate the performance of internet networking during Beijing Olympic Games, which is collected by GPERF, a globally distributed performance measurement system of China Education and Research Network (CERNET). In this framework, we deal with the information of QoS for simulation, visualization, and analysis of user's needs in a uniform algebraic way.**

Keyword:* *Algebra of System, Quality of Service, Object-Process-Network

I. Introduction

A. Beijing Olympic Games and Propagation of Information

The Beijing Summer Olympic Games has taken attentions all over the world. By the convenience of internet and telecommunication networks to facilitate the transmission, making our events happen to Beijing the first time information, we can deliver the information of Olympic Games to all parts of the world in seconds. Therefore, the behavior and the measurement of high concern, high rate of information arising events in the global communication network is an important research topic. The main characteristics are shown in the following areas:

- Heterogeneity: The processes of transmission of information tend to use a variety of means, through a variety of media, so the physical environment and the software system of the propagation of information should be taken into consideration. Therefore, it belongs to the issue about heterogeneous network.
- Effectiveness: In many aspects, the measurement of the effectiveness of information is not only about guarantee the information at arrival is the same as the original one. Taking the semantic of information into account, a more accurate maybe take the understanding of the recipient into consideration.
- Suddenness: From the study of major events and sports, the model of the attention over times is useful to our general information propagation modeling and design of emergency response system.
- Regionality: Different users in different geographical regions have different stimulus intensity to different information, so the propagation of information has a geographical orientation, which may make sense in the design of network deployment and decision-making on the deployment.

B. Algebra and System Modeling

A mathematical structure of an operand set and operator set, we call it an algebra. As operators for an algebra, they must obey the closure property of an algebra, thus they must be closed under the operator set of the same algebra, that is to say, operators to an arbitrary element in the operand set always results in some element in the same set. Therefore, we can written an algebra A as following tuple:

$$\mathrm{A} = < \{Operands\}, \{Operators\} >$$

With operators and operands, algebras can provide a way to describe and reason about information of system level representationly and efficiently. Operators offer us ways to rewrite algebraic expressions, for example, by creating operators which operate in the system model domain, we can simplify, transform, and reveal hidden qualities of the system model, which can improve our understanding of the system. Operands offer us ways to encode the information content of algebraic expressions, for example, by using abbreviated variable names or mathematically-defined constant values, we can decompose the system into subsets of qualitative and quantitative properties, which make it manageable.

C. Methodology of Quality of Service

As the name suggests, in real life, QoS often reflects the customers' satisfaction of services, and QoS is a measurement and evaluation of service level. In computer systems, especially computer network systems, such as computing and information services, the same advantages and disadvantages exist with the quality of service issues.

The goal of QoS control is to provide distinguished services and performance guarantee: Distinguished Service means to provide different services between different applications in accordance with the demand; performance guarantee means to guarantee these issues such as bandwidth, loss, delay,

978-0-7695-3747-4/09 $25.00 © 2009 IEEE
DOI 10.1109/ISPA.2009.87

delay jitter and other performance indicators. The QoS description applied to the dissemination of information network should include the following aspects:

- Characteristics of information flow: the characteristics of information flow are described for the user application to allocate resources, including the peak cell rate(PCR) and average cell rate(ACR) of information flow, the sudden flow of information as well as the average cycle length.
- Performance requirements of information flow: it describes the performance requirements of network, including network throughput, information delays, jitter, loss rate and so on.
- Information Consistency: it describes the invariance in transmission of information, including the contrast between the characteristics of information flow to the recipient and the characteristics of recipient feedback flow of information, and feedback information about the target.
- Service levels: it is used to describe the extent of end-to-end QoS guarantees, for example controlled-load service, guaranteed services, and best-effort services. Although performance requirements of information flow allow users to quantify the expression of the required properties of scale, the Services levels make a qualitative scale refining.

The remainder of this paper is organized as follows. We review the related work in Section 2, the basic definition of Algebra of System and using an executable meta-language OPN to describe QoS indicators in the propagation of information is shown in Section 3. In section 4, we will take the Beijing Olympics as a case to simulate and analysis the information propagation process.

II. RELATED WORKS

The development of the theory of mathematical mechanization provides an important theoretical basis and practical ideas to the algebraic meta-language to describe and model complex network system. C.A.R. Hoare[1] put forward the system modeling methodology based on algebra, and he thought that complex systems can be reasoned using concise and uniform algebraic language. Through the study of Process Algebra and Algebra of Computer Programs [2], Hoare pointed out that the basic role of algebra in modeling of entire system. Patrick Cousot and Rahdia Cousot's theory of Abstract Interpretation [3], reach a method that deal with different language to describe the system uniformly through the establishment of a concise algebraic language.

Ben Koo's theory the Algebra of System (AoS) which describe, model and abstract complex systems in a general algebraic method, has been used in a number of analysis of complex applications[20], and you can find more details in Simmons' and Cameron's papers [4, 5]. Based on the methodology of AoS, Ben Koo has developed a meta-language called "Object-Process Network" (OPN), which abstract and model complex network systems in its general form [6].

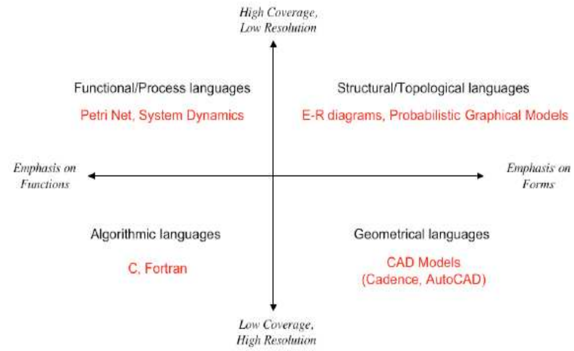

Figure 1: The Design Space of Meta-language

SimSync which is created by Ke Huang and Zhihong Li, using a form-based constraint language to describe the synchronous control logic of complex system, is designed as a parallel algorithm to solve the Synchronization constraint satisfiability problem.

The basic theory of the propagation of information are information theory (Shannon, 1948), it treat the propagation of information as a statistical phenomenon to consider, set up a basic model of information propagation, as shown in Figure 2, from the perspective of math, to measure the capacity of the communication channel, the basic unit of information is entropy [7].

$$H = -\sum p_i \log p_i$$

W. Weaver summarized Shannon's work is, pointing out the three basic levels of the information propagation: the technical problem, which is Shannon's research; semantic problem; effictiveness problem. He also pointed out that the "communication problem with the semantic has nothing to do with aspects of engineering and technology", at the same time he put forward a generalized information model [8]. Although he did not make any specific mathematical formulas, but his basic idea has lead the way for future generations.

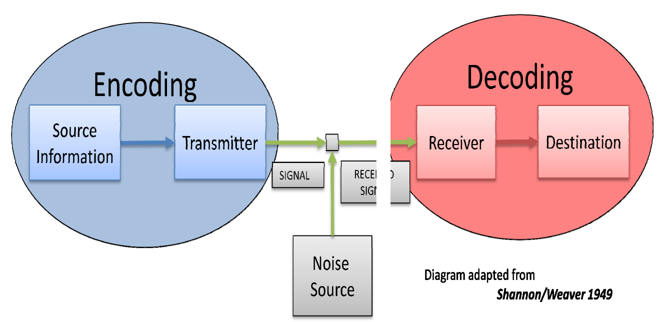

Figure 2: Communication Model Created by Shannon[7]

In the field of Qos(Quality of Service, QoS) which is mainly used in information network domain, the design of systematic language point out the direction of development of the theory, especially in the aspects of cross-domain modeling and decision making support. QoS refers to the agreement of quality between users of sending and receiving information as well as users and the general services network [9]. Since the birth of the computer system, there has been the issues to improve service performance and service quality. In early 80s of twentyth century, Seitz and Wortendyke have thoughts on user-based performance evaluation problem in the research of the X. 25 communications of ARPANET[10]. In late 80s, with the appears of B-ISDN technology and ATM networks as well as the dramatically increase of distributed multimedia applications, people began to study the management and control of QoS. Some experimental system began to appear. With the great success of Internet, the QoS problem of Internet went further. In September 1997, IETF formulated a range of RFC standards of the definition and service of the QoS, and they put forward two different Internet QoS Architecture: IntServ (Integrated Service, IntServ) [11] and DiffServ (Differentiated Service, DiffServ) [12]. Nowadays the analysis of the QoS of heterogeneous network has become a hot research topic: Maurizio D'Arienzo, Antonio Pescape and Giorgio Ventre proposed the initial definition of Service Condition [13]. Alessio Botta, Donato Emma, Antonio Pescapé, and Giorgio Ventre proposed the modeling method and characteristics of heterogeneous network based on QoS[14].

III. Analysis of QoS based on Algebra of Systems

A. *Information Propagation Model under Heterogeneous Network*

To solve problem of QoS analysis for the information propagation network, the key is on the definition of service condition. In fact, because in real life information often propagates through various forms of media, we need to use a unified approach to describe the propagation process of demand being satisfied in different service condition. We call it the heterogeneity of service condition, mainly in the following aspects:

- Hybrid network structure: user terminals, network media, the type of service and so on
- Dynamic user behavior: the study of information, propagation, practice, feedback and so on

Therefore, taking the real-life information propagation networks into account, the propagation of information is often in different channels and ways simultaneously. While when we are in the actual measurement process, we tend to consider between the two endpoints of the process of information propagation. Figure 3 indicate the real-world information propagation network, any arbitrary one of the two endpoints can complete the propagation process.

Figure 3: Information Propagating Network in Reality

To build the service condition of end-to-end propagation process of information, we can use the n-dimensional space to describe it, as the following 3-dimensional space:

（1）$\mathit{Propagation\ Network} \in PN$

$$\mathrm{PN} = \{\mathrm{Ethernet\ (10/100/1000), WLAN\ 802.11\ b, CDMA, Sound\ Wave}, \dots\}$$

（2）$\mathit{User\ Terminal} \in UT$

$$\mathrm{UT} = \{\mathrm{Television, Phone, Laptop, PC\ Desptop}, \dots\}$$

（3）$\mathit{Service\ Type}$

$$\mathrm{ST} = \{\mathrm{Transactional, Real\ Time, Isochronous}, \dots\}$$

Therefore, we get the following *Service Condition Space*

$$SC = (PN) \times (UT) \times (ST)$$

Of course, if we face a service condition of more heterogeneous elements, then we will get a higher dimension of Service Condition Space. At the same time for an end-to-end information propagation process, we can treat both ends of the service environment as a whole. If we consider a more general server-client network model, we can also use the structure shown in Figure 4 to describe:

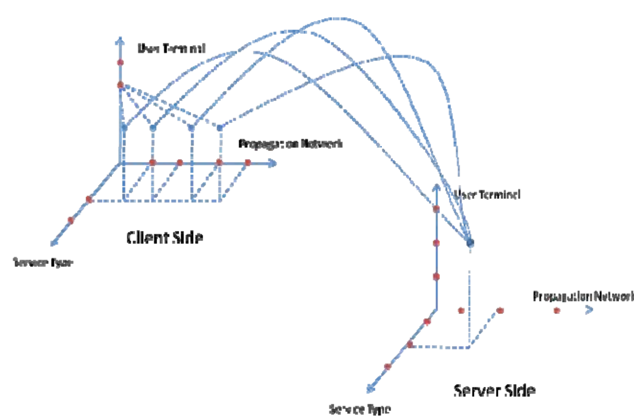

Figure 4: Server-Clients Services Condition Space Model

B. *Management of QoS of Information Propagation Based on Algebra of System*

Based on the model in information theory of Shannon, we consider the abstract model of end-to-end process of information propagation, as shown in Figure 5, the entire process can be regarded as the process that information from the sender to the destination. The sender as the source of information (for people what is brain) transmit information into the signal through transmitter (such as vocal sounds), through the network propagation medium to receiver(such as sound waves through the air reach the ears of receiver), then the receiver get the information through understanding.

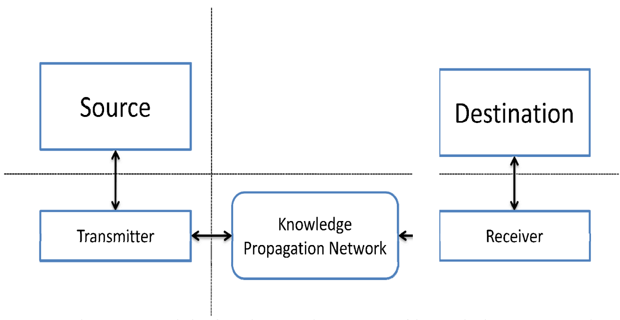

Figure 5: Abstract model of end-to-end process of knowledge propagation

We introduce the QoS mechanisms into this situation and use meta-language to model this. First of all, we will use "Object-Process Network" (OPN) which is a Petri-Net-like model to describe the information propagation network, which will convert the flow of information into the flow of token. Each token will bring the information of both sender's and receiver's Service Condition as the basic parameters to decision-making. We build the abstract model of OPN language to model end-to-end information propagation, as shown in Figure 6, the definition of OPN is in *Appendix A.*:

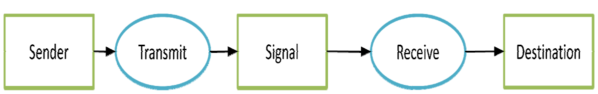

Figure 6：OPN description for end-to-end information propagating

Therefore, in the whole process of QoS measurement of Token flow, for example we will study bandwidth (Bit Rate,B), Delay(D), Delay jitter(J), Package Loss(PL), constitute a QoS space of description of characteristics of information flow, as follow:

$$\mathrm{QoS} = [(\mathrm{B}), (\mathrm{D}), (\mathrm{J}), (\mathrm{PL})] \in \mathrm{R}^4$$

IV. Design of Experiment

In this part, we focus on the change of users' needs when a major event or emergency happen. We use GPERf, a globally distributed measurement system located in 156 countries, 271 AS, nearly 1000 measurement points of the global monitoring network to simulate user request. It covers more than 156 counties to collect the world-wide internet QoS information. We construct the OPN diagram in figure 7 as an executable kernel which will explain details later:

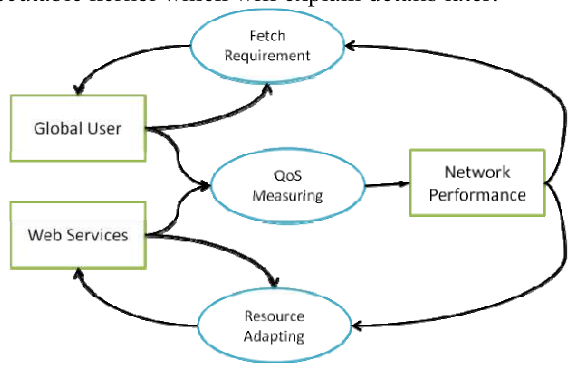

Figure 7: OPN diagram for global performance measurement architecture

A. *User Requirement Fetching Process*

Therefore, the structure of the token which represent global users is shown as follows:

$$\mathrm{User} \in \mathrm{User} = \{(\mathrm{AS}), (\mathrm{TURL}), (\mathrm{Time})\}$$

(1) AS: the AS (Autonomous System, AS) which users (measurement server) is in. Through the unique id of AS (ASN), we can get basic geographical and networking information of related user.
(2)TURL: the URL information of users' visiting target website. We have considered the visits information of 24 Olympic-related websites (*Appendix B*).
(3)Time: the current point of time. Taking the ability of collecting and store data, we measure the whole network by the frequency of once per hour. The format of time is "YY-MM-DD-HH".

The whole *User Requirement Fetching Process*: at a particular point in time, we store a collection of tokens which represent the user requirement. After using the process of fetch requirement, we get the collection of the next point of time. The following figure demonstrates a token change with time in a user requirement fetching process.

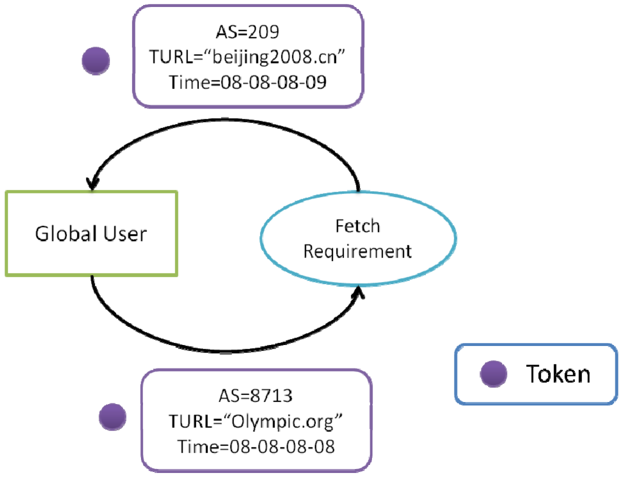

Figure 8: Part of OPN diagram for User Requirement Fetching Process

B. *Network Performance Evaluating Process*

Network performance evaluation of network is based on the information of bsic QoS measurement, from which we can get the load of network on the overall situation. To consider whether the deliver ability of the network satisfies the needs of users, we consider the entire network as the object which provides services, and we take time and space as the basic units of measurement, mainly referring to the end-to-end communication time and information space. Thus our tokens which bring the QoS information is in the following format:

$$\text{Network Performace} \in \text{NP} = \{(\text{RTT}), (\text{Loss}), (\text{UAS}), (\text{TURL})\}$$

- RTT: round-trip time, referring from the time when the sending end (measurement server) start to send information to the time when the sending end receive from the receiving end (target website) confirmation (the receiving end send confirmation immediately after receiving data). The basic units of RTT is milliseconds (ms). Therefore it can a measurement to the service which user received from target website
- Loss: Package Loss, referring to the completeness of information received by users, can be used to measure the stability of end-to-end information propagation between ends. We use a real number in 0 ~ 1 to express how many percentage it is that the loss of information packet over the total information packages. 1 is regarded that we can't touch the target through network, therefore we can't get the RTT data.
- UAS: the unique id of AS which user (a measure server) is in. To some extent, it can reflect the situation on the whole region.
- TURL: the URL of target website, as a record of the target end.s
- Time: the current point of time

Now we analyze the flow of tokens in the process of QoS measurement. As shown in the following figure, the user (measure server) visit (measure) the official site of the Beijing 2008 website at 9:00 on August 8, 2008.

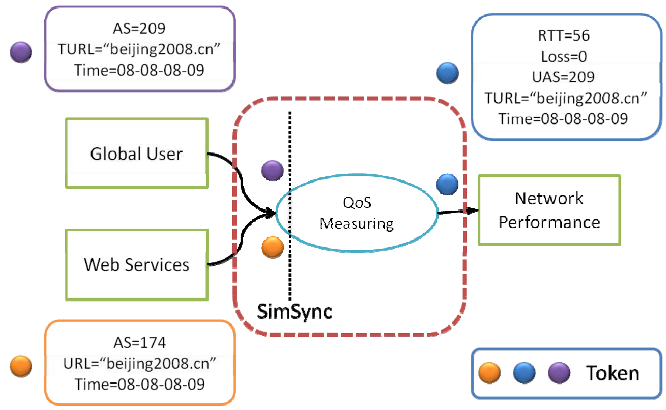

Figure 9: Part of OPN diagram for process of QoS measurement

We use SimSync in OPN, which is a table-based constraint language to describe the system synchronization control logic for controlling match between these tokens, which carry the user information, and those tokens which bring the information of web server load. The matching constraints are as follows:

$$\big((\text{GU.TURL} == \text{WS.URL})\&\&(\text{GU.Time} == \text{WS.Time})\big)$$

This express that the target website of user and the website, which provide the service, is the same one. Moreover, the QoS of the user and server matches. Then these two tokens will enter QoS measuring process to measure the information of QoS.

In the actual process, there are a lot of tokens, which bring the needs of users, are stored in the Global User; there are a lot of tokens, which bring the information of websites, are stored in the Web Services. With the mechanisms of SimSync, we ensure that tokens in object can travel in proper process in highly dynamic situation. From statistics of tokens, we can get the basic analysis for the QoS of Olympic-related websites. The following figure shows that the average RTT of official website of Olympic Games (www.olympic.org) in August 8,2008 .

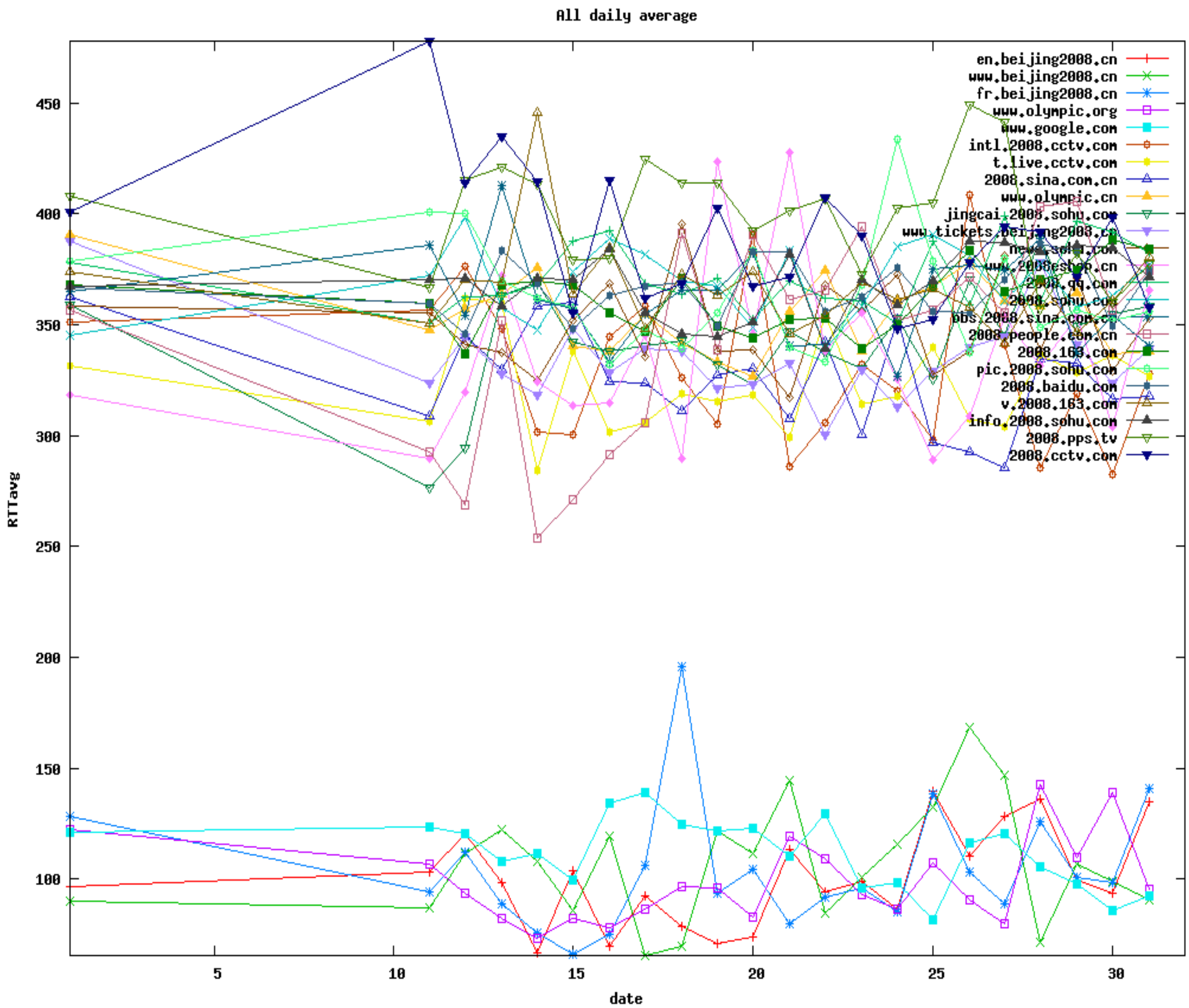

Figure 11: average RTT of related website of Olympic Games in August, 2008[17]

C. *QoS Decision Making with Time and Space*

Based on basic QoS information provided by our framework, we can make decision of service dynamically. On the one hand, QoS itself is a measurement of time and space, on the other hand, the process which produce the information of QoS also contain specific information of time and space. Therefore, we can integrate information of QoS, the information of time and space of server, to support decision making dynamically. The following figure shows the average RTT of Google in provinces of China. We can get that except some province, users in most of provinces can arrive target website in a acceptable delay time. Thus we can say that Google has taken this into consideration. Figure 13 shows that the RTT of the official Olympic web-site in provinces of China at August 8, 2008. We can find that after 20:00, RTT in some of the provinces increase dramatically because of the open ceremony.

Figure 12: Average RTT of Google website in provinces of China in August 8, 2008[18]

V. Conclusion

2008 Beijing Olympic Games as a world-wide event, Large-scale modeling can be effective here, because it abstracts away from modeling of individual behavior and considers population-based representations. In the present paper we have used a high-level modeling language (the algebra of systems) to generate frequency-based population dynamics models of Internet propagating process. If AoS model describes the behavior of the major system components well then the modeler is well-placed to interpret the results of the time series analysis which results from solving the initial value problem for the ODEs. In practice, we have found this to readily admit an intuitive interpretation.

Our plans for future work include the extension of AoS analysis to dynamic decision making and distributed computation. We seek to expand our coverage of the language to include various networks for information propagation.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (NSFC) under the contract 70771059.

Appendix A:

The Algebra of System(AoS) and Object-Process Network

AoS uses a composite data structure, a triple $< P, B, C >$ as its operand domain. P denotes the quantitative and qualitative properties, B denotes the Boolean value status, and C denotes the compositional structure of a system. The definitive reference for the language is [6].

- Domain Definition 1 (P: Properties domain): P is a domain of 2-tuples. It is a set of tuples with two elements, where the data content of the keys are non-repetitive. There are four closure operators in this domain which are merge, substitute, delete, and interp.P $= \{< key, value >\}$.
- Domain Definition 2 (B: Boolean domain): The constraints in AoS model is encoded in a Boolean expression where the values of the variables are supplied by the information content in the domain P. The operators in this domain are Boolean operators such as and, or, negate and interp.
- Domain Definition 3 (C: Composition domain): The composition domain, C, is a bi-partite graph data structure. An element in C encodes a system as a collection of smaller building blocks and a set of relationships between them. It uses two kinds of building blocks of a system that represent data and functional units of abstraction which are named Objects and Processes. Object is a pool for token which denotes the data as $< key, value >$ pair in P domain and records all the traces it passes though. Process can change the information that token carries. The relationships between Objects and Processes can be graphically represented as directed arcs between them, we call it pre-condition which from objects to processes and post-condition which from processes to objects. Each directed arc is associated with a Boolean expression, and its applicability is constrained by the result of the Boolean expression. We call these bi-partite graphs "Object-Process Network" (OPN), Examples of these bi-partite graphs are shown in Figure 6,7,8,9,14.

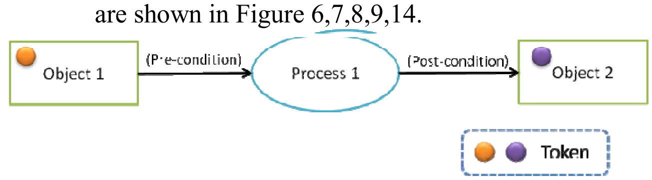

Figure 14:The Composition domain, C, depicted in OPN diagram

Appendix B.

TABLE I. The list for 24 Olympics-related websites[19]

The list for 24 Olympics-related websites		
www.google.com 2008.sina.com.cn	news.sohu.com 2008.sohu.com	bbs.2008.sina.com.cn www.olympic.cn
2008.people.com.cn	pic.2008.sohu.com	www.2008eshop.cn
in-fo.2008.sohu.com	2008.cctv.com	intl.2008.cctv.com
2008.163.com	2008.baidu.com	en.beijing2008.cn
t.live.cctv.com	www.beijing2008.cn	www.tickets.beijing2008.cn
2008.pps.tv	fr.beijing2008.cn	v.2008.163.com
www.olympic.org	jing-cai.2008.sohu.com	2008.qq.com

Reference

[1] Andy Dearden, Janet Finlay (2006). "Pattern Languages in HCI: A critical review." Human Computer Interaction, Vol.21, No.1

[2] Hoare, C.A.R., A theory of engineering design, EUROMICRO 94. System Architecture and Integration. Proceedings of the 20th EUROMICRO Conference.

[3] Patrick, Cousot, Radhia Cousot, BASIC CONCEPTS OF ABSTRACT INTERPRETATION, wcc2004

[4] W. L. Simmons, B. H. Y. Koo, E. F Crawley, A Computational Method for Mapping the Decision Space of the Lunar Exploration Program, Proceedings of 57th International Astronautical Congress 2006, Valencia, Spain, Oct. 2006, IAC-06-D1.3.7

[5] Bruce Cameron, Sandro N. Catanzaro, Edward F. Crawley, Value Based Architecture Selection, Proceedings of 57th International Astronautical Congress 2006, Valencia, Spain, Oct. 2006,IAC-06-D3.1.3

[6] B. H. Y. Koo, W. L. Simmons, and E. F. Crawley. (2007) Algebra of systems: an executable framework for model synthesis and evaluation. 2007 International Conference on Systems Engineering and Modeling

[7] Shannon C E. A mathematical theory of communication, Bell System Technical Journal,27 (1948), 379−429, 623−656

[8] C. E.Shannon and W.Weaver,The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949.

[9] Hutchison D, Coulson G, Gampbell A, et al. Quality of service management in distributed systems. In: Sloman M, ed. Network and Distributed Systems Management. Chapter 11. Addison Wesley, 1994

[10] Seitz N B, Wortendyke D R, Spies K P. User-Oriented performance measurements on the ARPANET. IEEE Communication Magazine, 1983, 1(5): 28~44

[11] IETF Working Group on Integrated Services

[12] http://www.ietf.org/html.charters/OLD/intserv-charter.html

[13] IETF Working Group on Differentiated Services

[14] http://www.ietf.org/html.charters/OLD/diffserv-charter.html

[15] M. D'Arienzo, A. Pescapé, G. Ventre, Dynamic service management in heterogeneous networks, Int. J. Network System Management (ISSN 1064-7570) 12 (3) (September, 2004) 349–370.

[16] A Botta, D Emma, A Pescapé, G Ventre,Systematic performance modeling and characterization of heterogeneous IP networks, Int. J. Computer and System Sciences 72 (2006) 1134–1143

[17] http://gperf.edu.cn/200808.php

[18] http://203.91.120.211/users_new/rtt_map_china.php?id=1&startdate= &enddate=

[19] http://gperf.edu.cn/index.php