Computer Networks 66 (2014) 102–111

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

From computer science to service science: Queues with human customers and servers

Hideaki Takagi

Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba Science City, Ibaraki 305-8573, Japan

ARTICLE INFO

Keywords:
Queueing theory
Little's law
Performance evaluation
Service science
"power"
patient flow in hospital

ABSTRACT

Professor Leonard Kleinrock has made key contributions during the initial stage of development in computer communication networks from several aspects. One of them is the ingenious application of queueing theory to the performance evaluation of communication networks. Queueing theory is still useful in contemporary service systems having human customers and servers as precious resources. However, some new theoretical development is needed to cope with human service systems. In this article, the potential of queueing theory is discussed in the scope of emerging *service science*. We begin with a snapshot of Professor Kleinrock's laboratory in the early 1980s. We then review the performance metric called *power* worked out by Kleinrock as it determines the optimal input rate in a service system. As an example of service science for healthcare, we show modeling of obstetric patient flow in a hospital by means of Little's law and a network of M/M/m and M/G/∞ queues.

1. Personal retrospective and view of queueing theory

My career as a researcher of computer and communication science started when I received the Ph.D. in Computer Science from the Computer Science Department of the University of California, Los Angeles (UCLA), in May 1983 [23]. My dissertation was supervised by Professor Leonard Kleinrock, who was then the principal investigator (PI) of the Contract MDA 903-82-C-0064 from the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense. My doctoral committee consisted of Professors Mario Gerla, James R. Jackson, Steven A. Lippman, Richard R. Muntz, and Leonard Kleinrock (Committee Chair). Two volumes of *Queueing Systems* [13,14] were textbooks in Professor Kleinrock's classes "CS212A Queueing Systems: Theory and Applications" and "CS212C Computer Communications Network". Later the solutions manuals were published coauthored by Kleinrock and Gail [18–20].

I spent a great time with Professor Kleinrock's students during that period including Mart Molle, Randy Nelson, Richard Gail, Ken Kung, Hanoch Levy, Yehuda Afek, Fathi Belgith, and Joe Green. I also talked a lot with other professors' students like Mike Molloy, Bruce Walker, Paul Hurley, Edmundo de Souza e Silva, Mary Vernon, Kin Leung, Frank Schaffa, and Luis Filipe de Moraes. All of them later became outstanding figures in academia and industry across the world. Several ex-students such as Ed Coffman, Simon Lam, Fouad Tobagi, Farouk Kamoun, Parviz Kermani, John Silvester, and Yechiam Yemini occasionally dropped by at the weekly students meeting in Professor Kleinrock's office and talked to younger *academic brothers* in a very friendly fashion. I sat in classes of Professors Richard R. Muntz, Mario Gerla, Walter J. Karplus, Wesley W. Chu, and Izhak Rubin (Systems Science) and met many distinguished visitors such as Vinton Cerf, Wim Cohen, Bob Kahn, Hisashi Kobayashi, Alan Konheim, and Steve Lavenberg.

Cheerful support staff members Lillian Larijani, George Ann Hornor, Brenda Ramsey, Verra Morgan, Ruth Pordy, and Terry Peters helped students in administration as well

E-mail address: takagi@sk.tsukuba.ac.jp

http://dx.doi.org/10.1016/j.comnet.2014.03.014
1389-1286/

as in preparation for research papers, because typewriting and figure drawing were done manually at that time.

On June 9 and 10, 1994, a two-day symposium entitled *Experts on Networks* was held at the Sequoia room of the UCLA Faculty Center in order to honor Professor Leonard Kleinrock and his contributions to computer science on his 60th birthday and over 30 years at UCLA. On that occasion, a genealogical tree of his students, students of his students, and so on, was compiled by Bob Felderman. According to the tree, I am the 21st student amongst those 35 students (at that time) who received doctoral degrees under the supervision of Professor Kleinrock, the first of whom being Ed Coffman. On June 11, participants in the workshop were invited by Mrs. Stella Schuler Kleinrock to the party at Professor Kleinrock's house in the Brentwood area of western Los Angeles.

With such background, I worked on the application of queueing theory to computers and communication networks during my most active research life. However, the use of analytic studies has soon become less important in these fields. I think there are two reasons for this decline. First, the rapid technological development, referred to as *Moore's law*, has brought powerful CPU and abundant memory for computers as well as virtually unlimited bandwidth of optical fiber for telecommunication networks so cheaply that we do not have to concern ourselves with the wise (and stingy) use of these resources any longer. Second, the protocol and control of the operation in these systems have become so complicated and interdependent that theoretical treatment, often based on simplifying fictitious assumptions such as independence of events and exponential distribution, cannot cope with the operation of real systems. Thus it is no wonder that analytic methods based individually on the human brain power have been replaced by algorithmic and event-driven simulation methods which can capitalize on the ever-growing computer power.

My recent interest is the mathematical and statistical treatment of service systems with human customers and servers, for example, queues in airports, call centers, hospitals, etc., in the context of an emerging discipline called *service science*. Notice that human beings still remain to be a precious resource that may not be abused in service systems. In most countries, whether developed or under development, the service sector of industry now takes a predominant portion of the national economy in terms of gross domestic product (GDP) as well as the population share of labor force. Unlike manufacturing, however, not much scientific approach has been exploited so far for the increase of productivity and promotion of innovation in the service industry. The *services science* was advocated as a new academic discipline in the so-called *Palmisano Report* from the Council on Competitiveness in the United States published in December 2004 [2].

According to my view, the study of service systems with human customers and servers is one of a few fields in which the queueing theory can still make prominent impacts on the practical side (another potential field may be the radio communication technology where the channel bandwidth continues to be a physically limited resource [26]). Application of queueing theory to human service systems is not new at all. It was only that the exclusively driving application was computers and communication networks from the 1960s to the 1980s. Science of service is the field of study in the 21st century on top of operations research, data science and computer science. I hope that the basis of our knowledge which was cultivated under the leadership of Professor Kleinrock enables us to pave the way for this new affluent area of research. Some books address the queueing theory for service systems [3,11,24].

The rest of this article is organized as follows. In Section 2, we characterize service systems involving human customers and servers by highlighting the difference from those systems in which the service is provided by tireless machines. In Section 3, we review the performance metric named *power* which I think Professor Kleinrock favored, and we confirm one of his conjectures numerically. In Section 4, as an example of queueing theory application to service science, we show a preliminary piece of work on the modeling of obstetric patient flow in a hospital by means of Little's law and a network of $M/M/m$ and $M/G/\infty$ queues. We conclude in Section 5 by discussing several features of queueing models for human service systems that challenge the queueing theorists.

2. Service systems with human servers and facilities

Service can be defined as the activity to bring value (satisfaction) to not only the recipients (customers) but also the providers (employees) by optimal management of a set of available resources. The service industry is so diverse. Thus the challenge of *service science* is to create the principles and architecture of service operation and to implement it in the system just as the computer science has developed the architecture of computer operation and implemented it in the system since the 1950s. Some service systems do not involve human servers in real time such as reservation of a hotel room, purchase of goods and money transfer over the Internet and the automatic check-in processing in the airport. However, there are still many service systems in which human servers play a major and indispensable role.

Fig. 1 shows a generic service system involving human servers and facilities. In a service organization, employees provide service by using some physical facilities as enablers. One of the characteristic features of service that differentiate "service" from "goods" is said to be the *intangibility*, meaning that the service is action or performance that cannot be touched physically in the same manner as goods can be sensed as objects [4, p.20]. However, our experience tells us that a comfortable physical facility and environment are essential for the customer satisfaction. Employees and facilities are the operational resources that are not free but require sizable investment to keep them available for customers. Therefore, the manager of service organization holds a finite number of employees and facilities in the pool to supply them when needed to administer the service.

Customers come to the service organization to get service. A certain set of resources (employees and facilities) must be allocated to each service for a certain duration of