# CanQueue 2011: 13th Annual Conference for Canadian Queueing Theorists and Practitioners

**Winfried Grassmann (University of Saskatchewan, Javad Tavakoli (University of British Columbia Okanagan)**

**August 26-28, 2011**

## 1 Overview of the Field

Queueing theory is concerned with developing and investigating mathematical models of systems where customers wait for service. The terms customers and servers are generic. Customers could, for example, be humans waiting in a physical line or waiting on hold on the telephone, jobs waiting to be processed in a factory, or tasks waiting for processing in a computer or communication system. Examples of service include a medical procedure, a phone call, or a commercial transaction. Queueing theory started with the work of Danish mathematician A. K. Erlang in 1905, which was motivated by the problem of designing telephone exchanges. The field has grown to include the application of a variety of mathematical methods to the study of waiting lines in many different contexts. The mathematical methods include Markov processes, linear algebra, transform theory, and asymptotic methods, to name a few. The areas of application include computer and communication systems, manufacturing systems, and health care systems. Introductory treatments of queueing theory can be found, for example, in [1] and [2].

## 2 Recent Developments and Open Problems

Many recent developments in queueing theory have been driven in large part by a greater interest in applications that involve human customers, for example in the rapidly growing call centre sector (see [3]). Humans behave in less predictable ways than, say, jobs in a factory or tasks in a computer system. For example, they may renege (abandon the queue), and retry later. The needs of human customers are likely to be heterogeneous (motivating the use of skills-based routing to connect different customers to different servers) and to vary with time (sometimes requiring transient rather than steady-state solutions). All of these complications lead to interesting mathematical challenges. The interest in modeling reneging has led to a substantial literature by now, for example see [5]. Asymptotic analysis, which in the past typically considered

situations where the arrival rate approached the capacity of a system with fixed number of servers, has been rejuvenated by a focus on situations where the arrival rate and the number of servers approach infinity simultaneously (see [4] for the first such analysis). Such many-server asymptotic analysis has resulted in a collection of simple-to-use formulas for recommended staffing, consisting of a linear term (minimum staffing for stability) and a square root term (safety staffing, to protect against random fluctuations). In addition to the focus on the call centre sector, applications in health care are becoming increasingly important, for example see [6]. Successful health care applications are likely to require further extensions to the queueing theory toolkit to accommodate customers that are given different priorities and have different needs and servers that may group together to work on one service and then move on to other tasks. Typical queue performance measures, such as average wait or average cost, will also need to be re-examined, and a greater focus on measuring equity as well as quantification of the medical consequences of waiting may be necessary.

# 3 Objectives

The objective of the CanQueue workshops, held annually since 1999, is to promote research and application of queueing theory. The workshops emphasize discussion of research projects in their early stages, to facilitate feedback from colleagues and collaboration while the direction of the project can still be influenced. It provides a venue for graduate students to present their work (in process or completed) in a supportive environment and it introduces interesting application areas. In this way, it strengthens the community of researchers in queueing theory in Canada.

This year, the meeting was held at the BIRS center in Banff, and as usual, it was a great success. The sessions were well attended, and there was often a lively discussion. We also appreciate the help we got from the center for organizing this meeting. Their effort greatly contributed to the success of CanQueue 2011.

# 4 Presentation Highlights

The meeting started on Friday with three interesting talks on queueing in health care. Sherry Weaver, from the University of Calgary discussed how a model for a finite queue that can be used to accommodate urgent patients for knee replacement. Of course, when giving urgent patients priority, the wait of non-urgent patients may become unduly long. Problems of an unduly long wait for non-urgent patients is quite general in health care. To counteract this problem, David Stanford, University of Western Ontario, suggested increasing the priority as the waiting-time increases The results of his study were provided in a number of graphs. A similar problem arises in transplant queues where the transplants can become more urgent as time passes. This problem was discussed in detail by

Steve Drekic, University of Waterloo in a paper co-authored by David Stanford and Dough Woolford, Wilfrid Laurier University.

A number of talks dealt with the application of queueing theory to radio networks. In particular, the following problem was addressed. The wireless spectrum has traditionally been assigned statically to licensed users. However, it is known that less than 5% of the assigned spectrum is typically used, and as the availability of the spectrum is getting more scarce and demand for radio access is increasing, it has become clear that unlicensed users have to gain access to the unutilized portions of the assigned spectrum. This situation gave rise to a number of research projects that were investigated by Attahiru Alfa, University of Manitoba, and his students, including Samitha Umwiththige, Sofia Alvarenga, Chamara Devanarayana, Charith Gunasekara, Samitha Wijedasa, and Nkouatchah Ngatched. They analysed cognitive radio, a technology to assist with dynamic spectrum sharing between licensed users, also known as primary users, and unlicensed users, also known as secondary use. Coognitive radio can be adopted based on underlay or overlay approaches. Based on the overlay approach, secondary users seek idle channels for use when they are free of licensed users. However, in order to do this, secondary users have to detect the availability of idle channels. The durations of busy and idle periods are correlated random variables followong a general distribution. Some of the projects presented developped stochastic models for describing the states a radio channel. Based on these models, it is possible for an secondary user to develop a sensing strategy. A sensing strategy involves determining how frequently to sense a channel. If the sensing is too frequent an secondary user spends too much resource (e.g. battery power); on the other hand infrequent sensing leads to possible missed opportunities of channel idle times.

An interesting application of queueing theory is to predict the time between yellow alerts for ambulances, that is, the time when the number of available ambulances falls below a certain level $k$. Mathematically, this corresponds to a $k$-partial busy period and can be solved as such. This was explained in a talk by Armann Ingolfsson, University of Alberta, and his two students Amir Rastpour and Bora Kolfal. Another contribution of Armann Ingolfsson, with co-authors Mohammend Delassay, Bora Kolfal, and Z. George Zhang (Professor with Simon Fraser University) involved a two-dimensional Markov chain to solve queueing systems where servers adjust their service rate in response to system load.

There are close relationships between queueing and inventory as shown by Zhe George Zhang, Simon Fraser University. His queueing model dealt with switching from MTO (Work-To-Order) to MTS (Make-To-Stock). Another area having a close relation to queueing is the ruin problem in insurance as highlighted in a talk by David Landiault, University of Waterloo. A third area related to queueing theory is scheduling, as shown by two talks, both given by Tony Tran, Dara Terekhov (University of Toronto) and Douglas Down (McMaster University) and Christopher Beck.

The workshop included a number of more theoretical talks, reflecting the mix

of mathematical developments and practical applications that has been a hall mark of the CanQueue workshops. A talk by Attahiru Alfa and V. Ramaswami showed that the number in the system in a Geo/G/$\infty$ queue does not lead to simple expressions, quite in contrast to the M/G/$\infty$ queue, where the number in the system follows a Poisson distribution. Percy Brill (University of Windsor) and Mei Ling Huang (Brock University) applied level crossing analysis to a bulk queue. A problem of great practical interest is how the tails of queue length distributions behave. In most, but not all cases, the distributions in question show a exponential decay. Two talks dealt with the tail behaviour of distributions observed in queueing theory. One was given by Hui Li (Mount Saint Vincent University), Javad Tavakoli (University of British Columbia Okanagan) and Yiqiang Zhao (Carleton University), and it dealt with singular cases, that is, cases where the decay is not exponential. The other one, given by Zafar Zafari, a student of Javad Tavakoli, University of British Columbia Okanagan, analysed as to when the number in a shorter queue system is singular. Another student of Javad Tavakoli analysed when to interrupt and restart a finite queueing system with server breakdowns.

For practical applications, time-varying arrival rates to queues are very important. Such arrivals where explored by numerical methods by Barbara Margolius, Cleveland State University. Numerical experimentation was also discussed in a talk by Winfried Grassmann, University of Saskatchewan (coauthored by Javad Tavakoli). who showed that the speed of modern computers, even laptops, allows for experimentations impractical in the past. However, not all problems benefits equally form higher speeds. In particular, in queueing networks, non-simulation methods increase exponentially with the number of queues in the network, and even if the speed of computers increases 100 fold, only networks with a limited number of queues can reasonably be solved numerically.

# 5 Scientific Progress Made

As at past CanQueue meetings, the unique feature that facilitates progress on research is that these workshops attract both queuing theorists, who focus on developing new methodology, and researchers who apply queueing theory in various settings. Thus, it provides opportunities for theorists to learn about new application areas and the types of models that are needed for these areas and opportunities for researchers with a more applied bent to get suggestions from queueing theory specialists on potentially useful methodologies or approaches. As well, the 2011 workshop brought together theorists with different foci, for example those that focus on asymptotic analysis versus those that focus on matrix analytic methods, and the interchanges between these groups brought valuable insights on what each of these fields can add to the other.

CanQueue 2011 ended with a business meeting, and it was decided to create a special interest group within CORS. The name for this groups is still to be

determined. It was also decided to hold the next CanQueue in London, Ontario. This ended the very successful meeting at BIRS in Banff.

# 6 A Note of Appreciation

The organizers would like to thank BIRS staff for their very competent assistance in organizing this workshop. Many participants commented favourably on the superb BIRS facility. The location and the amenities made it considerably easier for us to attract the group of distinguished researchers that attended the workshop. We thank again for the excellent help of the staff of BIRS which greatly contributed to the success of CanQueue 2011.

## References

[1] L. Kleinrock, Queueing Systems, Volume 1: Theory, Wiley, New York, 1975.

[2] D. Gross and C. M. Harris, Queueing Theory, third edition, Wiley, New York, 1998.

[3] Telephone call centers: Tutorial, review, and research prospects, Manufacturing & Service Operations Management 5 (2003), 79141.

[4] S. Halfin and W. Whitt, Heavy-traffic limits for queues with many exponential servers, Operations Research 29 (1981), 567588.

[5] S. Zeltyn and A. Mandelbaum, Call centers with impatient customers: many-server asymptotics of the M/M/n + G queue, Queueing Systems 51 (2005), 361402.

[6] L. Green, Capacity planning and management in hospitals. In Operations Research and Health Care: A Handbook of Methods and Applications (M. L. Brandeau, F. Sainfort, W. P. Pierskalla, eds.), Kluwer, Boston, 1541.