QUESTA manuscript No.
(will be inserted by the editor)

Stability and Performance of Greedy Server Systems:

A Review and Open Problems

Leonardo Rojas-Nandayapa · **Sergey Foss** ·
Dirk P. Kroese

Received: date / Accepted: date

Abstract Consider a queueing system in which arriving customers are placed on a circle and wait for service. A traveling server moves at constant speed on the circle, stopping at the location of the customers until service completion. The server is *greedy*: always moving in the direction of the nearest customer. Coffman and Gilbert conjectured that this system is stable if the traffic intensity is smaller than 1; however, a proof or counterexample remains unknown. In this review we present a picture of the current state of this conjecture and suggest new related open problems.

Keywords Greedy Server · Stability · Continuous Polling System · Random Measures

Mathematics Subject Classification (2000) 60K25 · 68M20 · 90B22

1 Introduction

Polling systems form a subclass of queueing models where multiple queues are served in a fixed order. Applications are found in telecommunications and computer networks, reliability, manufacturing, and transportation. Basic polling models are typically formulated as systems with a finite number of queues attended in a cyclic order by a single server. Usually, the "switchover" time to move from one queue to another

L. Rojas-Nandayapa
Department of Mathematics, The University of Queensland, Brisbane, Australia
Tel.: +61-733652309
Fax: +61-733651477
E-mail: l.rojasnandayapa@uq.edu.au

S. Foss
School of Mathematics and Computer Sciences, Heriot-Watt University, Edinburgh,United Kingdom

D.P. Kroese
Department of Mathematics, The University of Queensland, Brisbane, Australia

is non-zero. Although much is known about the behavior of such discrete polling systems (see, for example, Tagaki [24–26] for an overview), it has been recognized that the analysis of *continuous polling systems* — obtained by letting the number of queues go to infinity — is often much more transparent than that of their discrete analogues [8,12]. A classical example is the continuous polling system proposed by Fuhrmann and Cooper [12]. Here, the server travels at constant speed and direction (say clockwise) around a circle of length 1 and stops at the location of the customers for service. After completion of service the customer is removed from the circle and the server resumes his journey. The arrival and service times are assumed to be of M/G/· type, while the locations of arriving customers are chosen according to a uniform distribution over the circle, independent of everything else. The model was further analyzed by Coffman and Gilbert [5,6] and Kroese and Schmidt [15–17], and its workings are now well understood.

This fundamental continuous polling system has been generalized in many ways. A natural modification is to consider a *random* server, whose movement is governed by a stochastic process that is independent of the positions of the customers [1,17, 18]. Another simple modification of the standard continuous polling system is to allow the server to reverse his direction and to let him travel (at constant speed) always in the direction of the nearest customer [6,19]. This system is state dependent, in the sense that the server needs to know the positions of the customers in order to determine his movement. Such a model is called a *greedy server system*. Figure 1 gives an illustration.

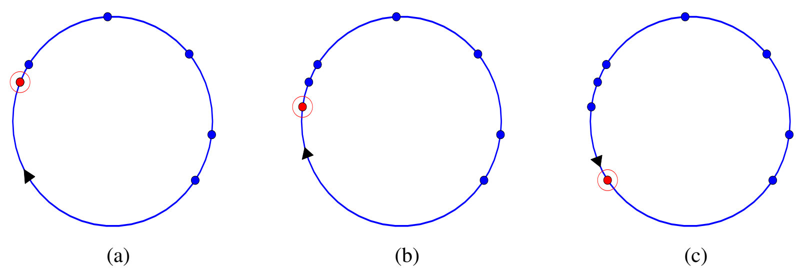

Fig. 1 A greedy server system. The server (triangle) continuously monitors the position of the nearest customer (encircled point) and travels towards it (a) and (b). When a new customer arrives that is closer to the server, the server aims for this new customer, possibly changing his direction as in (c).

A natural question is under which conditions these systems are *stable*, in the sense that certain associated random processes, such as the workload process W_t, converge in distribution as $t \to \infty$. Typically, these processes are *regenerative*, in which case the system is stable if the expected length of the regenerative cycle (for example, the busy period) is finite.

Stability for a discrete polling server has been extensively studied and there is a vast literature covering it (see [4,25,26] and references therein). For its continuous counterpart we can reason as follows: By relating the system to an M/G/1 queue with the same arrival rate and service time distribution (basically, ignoring the travel time), it is clear that a *necessary* condition for stability is that the *traffic intensity is smaller than 1*; that is,

$$\rho = \lambda b < 1, \tag{1}$$

where λ is the arrival rate and b is the expected service time.

It turns out that condition (1) is also *sufficient* for the stability of the continuous polling server that always moves in one direction, irrespective of the server speed [16]. Also the random server governed by Brownian motion [17,18] and the general random server [1] are stable under (1). It is widely believed that (1) is also sufficient for stability of the greedy server, but a proof remains to be found. This conjecture was first formulated by Coffman and Gilbert in [6] and is further analyzed in [3]. Partial stability results concerning a light-traffic regime are reported in [19]; other partial results may be found in [2].

The main difference between greedy and non-greedy server systems is that the motion of a greedy server is state dependent, i.e., it depends on the position of the customers. This makes the analysis of such systems much more difficult. The purpose of this paper is to provide an overview of the general class of *continuous cyclic server systems*, of which the continuous polling and greedy server system are examples, and to highlight a number of intriguing open problems for such systems, in particular with respect to greedy servers.

2 Spatial and Greedy Server Systems

Continuous cyclic server systems are idealized queueing system in which customers arrive in a continuous waiting space and are served by one or more servers that roam this space. The description of such models comprises the following ingredients.

- *Waiting space*. This describes the medium where the customers and servers are located. In many studies the waiting space is a line segment or a closed curve (circle) [5,6,12,15–19,21,22]. However, graphs [1,7] and bounded convex regions [2,3] have also been considered.
- *Arrival and service times.* This specifies when customers arrive and how long their service takes. The standard assumption is that the interarrival times are independent and identically distributed (iid), and are independent of the iid service times. The arrival and service characteristic can thus be summarized in Kendall's notation as $G/G/k$.
- *Customer location.* This describes where customers arrive in the waiting space. Typically, the position is chosen at random and independent of everything else according to a continuous distribution on the waiting space. Hence, the space-time distribution of the arrival process is then modeled as a marked point process (Poisson random measure in the $M/G/\cdot$ case). Continuous space-time arrival processes (*snowfall*) have also been considered [8,9,14].
- *Server discipline*. The order in which the customers are served is largely determined by the motion regime of the server:

- A *continuous polling* or *scanning* server moves in a deterministic pattern over the waiting medium [5–7, 15–17].
- The movement of a *random* server is governed by a stochastic process, such as a Brownian motion [1, 17, 18], independent of the positions of the customers.
- A *greedy* server moves in the direction of the closest customer.
  - A *non-dynamic* greedy server chooses the closest customer and travels to it ignoring all arriving customers.
  - A *semi-dynamic* greedy server moves according to a predefined pattern. After completing each service this pattern is redefined according to the current positions of the server and customers.
  - A *dynamic* greedy server always aims at the closest customer so it can change its destination and even reverse its direction with the arrival of a new customer [6, 19].

There is also a significant amount of literature on *discrete* greedy server systems, consisting of a finite number of service stations. Arriving customers choose one service station at random and wait for service. The system is attended by a single server who travels among stations. The server discipline is partly determined by the way the server handles waiting customers in a single queue:

- *Exhaustive:* The server resumes his journey only after emptying the current station [10, 13].
- *Gated:* The server resumes his journey after serving all customers initially found in the queue. Those arriving during service periods are set aside to be served in the next round [2, 3].
- *Fixed/Random:* The server resumes his journey after serving a fixed/random number of customers [11].

A related symmetric discrete greedy system with exhaustive server discipline has been studied in [23] and its stability was proved. In [10] stability is proved for a discrete system with general topology, exhaustive server discipline, and for a class of policies that includes the greedy policy. Similar results are obtained in [11] for other server policies.

3 Open Problems

Guided by the classification of models in Section 2, we formulate a number of open problems for continuous cyclic server systems, starting with Coffman and Gilbert's 24-year old conjecture. Two recurring themes are: stability of the system and characterization of the random measure describing the steady-state customer positions.

Open Problem 1 (Stability for the Greedy Server System) Consider the following greedy server system on a circle of circumference 1. Customers arrive and are served in an M/G/1 way with arrival rate λ and expected service time b. The location of customers is chosen uniformly over the circle. A single server travels at constant speed α^{-1} toward the nearest customer and remains at the location of the customer during service. After completion the customer is removed from the system and the server resumes his journey in the direction of the nearest customer. The discipline of this greedy server is dynamic in the sense that the server can aim for a different customer or reverse his direction with the arrival of a new customer. Let $\rho = \lambda b$ be the traffic intensity of the system. *Prove that* $\rho < 1$ *is sufficient for stability or give a counterexample. Analyse this conjecture also for the cases of semi- and non-dynamic greedy service mechanisms.*

The key here is: assuming that the server starts at the bottom point of the circle and a single customer is located at the very top of the circle, prove that the server has a finite expected traveling time from the bottom to the top. A partial result is given in [19] where stability of the greedy server is proven if the following *light traffic condition* holds:

$$\lambda\left(b+\frac{\alpha}{2}\right) < 1. \tag{2}$$

Notice that α is the time that the server needs to walk around the whole circle, so $\alpha/2$ serves as an upper bound of the traveling time between any two points on the circle. Adding this to the service time of each customer a stable M/G/1 queue which works slower than the greedy server can be obtained. Hence, for λ sufficiently small the greedy server empties infinitely often.

Open Problem 2 (Stationary Measure for the Greedy Server System) Suppose the greedy server system in Problem 1 is stable. Let Q denote the stationary random counting measure representing the positions of the waiting customers relative to the position of the server, given that the server is not busy. *What is the mean measure of* Q*? Or, more generally, what is the distribution of* Q*?*

For the polling server with *constant* service times the Laplace functional of Q is known explicitly:

$$\mathbb{E}\,\mathrm{e}^{-Qf} = \mathrm{e}^{-c_f}\left(\frac{1-\rho}{1-\rho\int_0^1 \mathrm{e}^{-h(y)}\,dy}\right)^{\alpha/b}, \tag{3}$$

where $c_f = \alpha\lambda\int_0^1(1-x)(1-\mathrm{e}^{-f(x)})\,dx$ and $h(y) = \rho\int_0^y(1-\mathrm{e}^{-f(x)})\,dx$. In particular, the mean measure is

$$\mathbb{E}\,Q(dx) = \frac{\lambda\alpha}{1-\rho}(1-x)dx. \tag{4}$$

For polling servers with *general* service times the mean measure is still of the form (4), but the distribution of Q can only be specified implicitly in terms of a system of functional equations [16]. The mean measure of Q is also known in the case of the Brownian server [18]. Under *heavy traffic* conditions, that is $\rho \approx 1$, the random measure $(1-\rho)Q$ for the polling server is approximately of the form $|Q^*|2(1-x)dx$, where $|Q^*|$ has a gamma distribution [15].

No such results are known for the case of the greedy server. However, [19] provides a second-order expansion of the mean measure Q:

$$\mathbb{E}\,Q([0,x]) = \int_0^x m_Q(u) + O(\lambda^3) \qquad \text{as } \lambda \to 0,$$

where

$$m(u) = \lambda\alpha(1+\rho)\left(\frac{1}{2}-u\right) + \lambda^2\alpha^2\left(\frac{u^2}{2}-\frac{2u^3}{3}\right) + O(\lambda^3) \qquad \text{if } u \in [0,1/2],$$

and $m(u) = m(1-u)$ for $u \in [1/2,1]$.

Open Problem 3 (Performance Measures) Suppose again that the greedy server in Problem 1 is stable. *Find expressions for stationary performance measures of the system, such as the stationary workload, the stationary mean sojourn, and the stationary waiting times of a random customer. Based on these stationary performance measures, find conditions under which the greedy server is more efficient than the polling server and vice versa.*

Examples of such performance characteristics for the polling and other spatially distributed service systems may be found in [18,19], where the expansion (4) of the mean measure Q is used to provide approximations of several such performance measures under light traffic conditions ($\lambda \to 0$). These approximations are used to verify that the greedy server works more efficient than the polling server in light traffic conditions.

Open Problem 4 (Maximum Position of Greedy Server) Consider a greedy server on an infinite line or a plane $\mathscr{X}$. Initially, the space $\mathscr{X}$ is empty of customers, and the server is located at the origin. Customers arrive according to a homogeneous Poisson process on $\mathbb{R}_+ \times \mathscr{X}$. *What can be said about the maximum distance from the origin, say M_t, that the server reaches in some time interval $[0,t]$? Does it tend to infinity or is it bounded? If the former is true, what is the rate of increase?* A clue is given by Kurkova and Menshikov [20], who consider a discrete greedy Markov system and show that the server turns its direction finitely often almost surely and its position tends either to plus or minus infinity with equal probabilities. They also prove that the ratio $M_t/\log t$ converges almost surely to a random variable taking values C and $-C$ with equal probabilities, where C is some constant.

Open Problem 5 (Greedy Server on a General Space) Consider a greedy server system, where the arrival and service times are again of M/G/$\cdot$ form, with arrival rate λ and mean service time b. The customers are now placed on a general space (e.g., an n-dimensional surface or manifold) according to some continuous distribution, and a single dynamic greedy server moves at constant speed over the space in the direction of the nearest customer. *Under what conditions is the system stable? How are the customers (points) distributed in steady state, relative to the position of the server?*

Open Problem 6 (Multiple Greedy Servers) This time, k servers travel in some space, serving each customer they encounter. Clearly, a dynamic greedy discipline may by inefficient here. For example, two or more servers can aim at the same customer at the same time. In this case a non-dynamical greedy approach might be more appropriate. A natural question in additional to performance and stability is: *what is the most efficient service/movement policy for the servers, e.g., to minimize the steady-state expected number of customers in the system.*

A number of related open problems for discrete-systems with greedy server and with fixed/random service policies are discussed in [11]. Two recent papers [21,22] study related models with multiple servers, each of which chooses a customer that is closer to the server than the others.

Acknowledgements Dirk Kroese would like to acknowledge the support of the Australian Research Council through the Discovery Grant DP0985177.

References

1. E. Altman and S. Foss. Polling on a space with general arrival and service distribution. *Oper. Res. Lett.*, 20:187–194, 1997.
2. E. Altman and H. Levy. Queueing in space. *Adv. Appl. Prob.*, 26:1095–1116, 1994.
3. D. J. Bertsimas and G. van Ryzin. A stochastic and dynamic vehicle routing problem in the euclidian plane. *Operations Research*, 39:601–615, 1993.
4. S C Borst. *Pollin Systems*. CWI Tract, Amsterdam, 1996.
5. E. G. Coffman and E. N. Gilbert. A continuous polling system with constant service times. *IEEE Trans. Inform. Theory*, 32:584–591, 1986.
6. E. G. Coffman and E. N. Gilbert. Polling and greedy servers on line. *Queueing Systems*, 2:115–145, 1987.
7. E. G. Coffman and A. Stolyar. Continuous polling on graphs. *Probab. Engrg. Inform. Sci.*, 7:209–226, 1993.
8. I. Eliazar. The snowblower problem. *Queueing Systems*, 45:357–380, 2003.
9. I. Eliazar. From polling to snowplowing. *Queueing Systems*, 51:115–133, 2005.
10. S. Foss and G. Last. Stability of polling systems with exhaustive service policies and state-dependent routing. *Ann. Appl. Probab.*, 6:116–137, 1996.
11. S. Foss and G. Last. On the stability of greedy polling systems with general service policies. *Probab. Engrg. Inform. Sci.*, 12:49–68, 1998.
12. S. W. Fuhrmann and R. B. Cooper. Applications of the decomposition principle in M/G/1 vacation models to two continuum cyclic models. *AT&T Tech. J.*, 64:1091–1098, 1985.
13. A. Harel and A. Stulman. Polling, greedy and horizon servers on a circle. *Operations Research*, 43:177–186, 1995.
14. D. E. Knuth. *The Art of Computer Programming*, volume 3. Addison-Wesley, Reading, MA, 1973.
15. D. P. Kroese. Heavy-traffic analysis for continuous polling models. *J. Appl. Probab.*, 34:720–732, 1997.
16. D. P. Kroese and V. Schmidt. A continuous polling system with general service times. *Ann. Appl. Probab.*, 2:906–927, 1992.
17. D. P. Kroese and V. Schmidt. Queueing systems on a circle. *J. Math. Methods Oper. Res.*, 37:303–331, 1993.
18. D. P. Kroese and V. Schmidt. Single server queues with spatially distributed arrivals. *Queueing Systems*, 17:317–345, 1994.
19. D. P. Kroese and V. Schmidt. Light-traffic analysis for queues with spatially distributed arrivals. *Math. Oper. Res.*, 21:137–157, 1996.
20. I. A. Kurkova and M.V. Menshikov. Greedy algorithm, z^1 case. *Markov processes and related fields*, 3(2):243, 1997.
21. L. Leskelä and F. Unger. Stability of spatial polling system with gready miopic server. *Annals of Operation Research*, 2010. To appear, http://arXiv.org/abs/0908.4585.
22. Ph. Robert. The evolution of a spatial stochastic network. *Stoch. Proc. Appl.*, 2010. To appear, http://arXiv.org/abs/0908.3256.
23. R. Schassberger. Stability of polling networks with state-dependent server routing. *Probab. Eng. Inform. Sci.*, 9:539–550, 1995.
24. H. Takagi. *Analysis of Polling Systems*. MIT Press, Cambridge, MA, 1986.
25. H. Takagi. Queueing analysis of pollin systems: An update. In H Takagi, editor, *Stochastic Analysis of Computer and Communication Systems*, pages 267–318. Elsevier Science Inc., North Holland, Amsterdam, 1990.
26. H. Takagi. Queueing analysis of polling models: Progress in 1990-1994. In J. H. Dshalalow, editor, *Frontiers in Queueing: Models and Applications in Science and Engineering*, pages 119–146. CRC Press, Boca Raton, FL, 1997.