Stability of Polling Systems with State-Independent Routing

Serguei Foss[*], Natalia Chernova[†] and Artyom Kovalevskii[‡]

This paper deals with the stability study of polling systems with either finite or infinite (countable) number of stations (queues) and with a finite number of servers that poll (visit) the stations along some random (state-independent) routes. First, we formulate "global" and "local" stability theorems for systems with a single server and with a general stationary ergodic input. Their proofs are based on certain monotone properties of underlying stochastic processes (see [15-17]). Second, we give a stability criterion for systems with several servers, with a finite number of stations and with i.i.d. driving sequences. The proof of the latter criterion (see [18]) is based on the fluid approximation approach.

Stability conditions for polling systems have undergone study rather recently (see, e.g., [1-12]) and all available papers deal with systems with finitely many stations and (except [2]) with either Poisson or renewal input.

Keywords: polling system, stability, stationarity, monotonicity, saturation rule, fluid approximation.

§1. Systems with a Single Server: "Global" Stability

Our approach is based on ideas of the so-called saturation rule [13].

Introduce a polling system with $K \le \infty$ stations. Let $\langle \Omega, \mathcal{F}, P \rangle$ be a probability space. All random variables below are considered on this space.

The input. By the *input* we mean a marked point process T with points T_n ($T_0 = 0$) and marks ξ_n. The sequence $\xi_n = (\tau_n, \mu_n, \sigma_n), n = 0, \pm 1, \pm 2, \ldots$, is assumed stationary and ergodic. Here $\tau_n = T_n - T_{n-1}$ is the interarrival time between customer $(n-1)$ and customer n, μ_n is the number of the station to which customer n is directed, and σ_n is his service time.

Let $\mathbf{E}\tau_1 = \lambda^{-1}$ be finite and positive; $\mathbf{E}\sigma_1 = \sigma < \infty$; and $\mathbf{P}(\mu_1 = k) = p_k > 0$ for every $k = 1, 2 \ldots, \sum_{k=1}^{\infty} p_k = 1$.

[*] ColoState University and Novosibirsk State University, foss@stat.colostate.edu and foss@math.nsc.ru
[†] Novosibirsk State University, root@cher.nsu.ru
[‡] Novosibirsk State University, artyom@math.nsc.ru

The server route. Assume to be given a sequence $\{\nu_j, w_j\}_{j=-\infty}^{\infty}$ of pairs of random variables, where the random variable ν_j takes values $1, \ldots, \infty$ and equals the number of the queue visited by the server jth in succession and the random variable $w_j \geq 0$ is the walking time from queue ν_j to queue ν_{j+1}. Suppose that the sequence $\{\nu_j, w_j\}$ can be partitioned into independent identically distributed (i.i.d.) segments of random length (cycles); i.e., there exists an increasing sequence of integer-valued random variables $\{j_i\}_{i=-\infty}^{\infty}$ such that the random vectors ("cycles of the route")

$$\eta_i = (l_i; \nu_{j_i+1}, \ldots, \nu_{j_{i+1}}; w_{j_i+1}, \ldots, w_{j_{i+1}}), \quad i \in \mathbf{Z},$$

are i.i.d. Here $l_i = j_{i+1} - j_i$ is the number of queues (stations) visited by the server in cycle i. Assume that the cycles start with visiting the same queue. For definiteness, let $\nu_{j_i+1} = 1$ for all i. Denote by $c_k^i = I(\nu_{j_i+1} = k) + \ldots + I(\nu_{j_{i+1}} = k)$ the number of visits to queue k in cycle i, where $\mathbf{P}(c_k^1 > 0) > 0$ for all k, and denote by $\psi_i = w_{j_i+1} + \ldots + w_{j_{i+1}}$ the total walking time during the cycle.

Let $L = \mathbf{E} l_1 < \infty$, $W = \mathbf{E}\psi_1 < \infty$, and $C_k = \mathbf{E} c_k^1 < \infty$ for all k. Assume the sequences $\{\eta_i\}$ and $\{\xi_n\}$ to be independent.

By a *route of the server in the empty system* we mean a marked point random process whose points are the starting times of the cycles and the distance between points equals the total walking time during the corresponding cycle.

Denote by $\Psi = \{\Psi_i, \eta_i\}$ the point process with points Ψ_i and marks η_i, $i \in \mathbf{Z}$, in which $\Psi_0 = 0$ and $\Psi_i = \Psi_{i-1} + \psi_i$ is the finish time of cycle i if the server moves in the empty system.

Consider also a stationary version (in continuous time) of the process Ψ which we denote by $\Psi^{(1)} = (\Psi_i^{(1)}, \eta_i^{(1)})_{i=-\infty}^{\infty}$. Assign the number 0 to the first positive point of this process, so that $\Psi_0^{(1)} > 0 \geq \Psi_{-1}^{(1)}$ a.s.

Denote by $\Psi^{(-n)}$, $n \geq 0$, the stationary ergodic point process that is obtained from the process $\Psi^{(1)}$ by shifting each point to the left by the random variable $\sum_{j=-n}^{0} \sigma_j$ and is renumbered so that $\Psi_0^{(-n)}$ is its first positive point.

Since $\Psi^{(1)}$ and $\{\xi_k\}_{k=-\infty}^{\infty}$ are independent, for every $n \geq -1$ the process $\Psi^{(-n)}$ is independent of the sequence $\{\xi_k\}_{k=-\infty}^{\infty}$ and its distribution coincides with the distribution of $\Psi^{(1)}$.

The service policies. If the server, on visiting station k at time j, finds x customers in a queue, then it serves, without interruption, $f_k^j(x) \equiv f_k(x, D_k^j) \leq x$ customers in the *FIFO* order, and then moves to the next station of the route. Upon service completion, customers leave the system. Here (for every k) the random variables D_k^j, $j = 0, \pm 1, \pm 2, \ldots$, are i.i.d. Suppose that the service policies satisfy the conditions $\mathbf{P}(f_k^j(1) = 1) = \delta_k > 0$ and $f_k^j(x) \leq x$ a.s. for all $x \in \mathbf{Z}^+$, $k = 1, \ldots, K$, and belong to the class $M = \{f : f(x, y) \leq f(x+1, y) \leq f(x, y) + 1$ for all $x \in \mathbf{Z}^+$, $y \in \mathbf{R}\}$. We call the class M the *class of monotone service policies*. For the service policies in M, there always exists a (finite or infinite) limit

$$F_k = \mathbf{E} \lim_{x \to \infty} f_k(x, D_k^j) = \mathbf{E} F_k(D_k^j) \leq \infty.$$

Henceforth the number $[m, n]$ and the arguments Ψ and $T = \{T_k\}$ in some characteristic of the system (queue length, exhaustion time, etc.) signify that the characteristic is considered in the system that is governed by the route Ψ of the server and to which only the customers with numbers $m \le \ldots \le n$ are submitted at respective times $T_m \le \ldots \le T_n$. By a *nonempty* cycle we mean a cycle during which there are customers in the system.

Denote by $X_{[m,n]} = X_{[m,n]}(T, \Psi)$ and $\widetilde{X}_{[-k,l]} = X_{[-k,l]}(T, \Psi^{(-k)})$ the finish times of the last nonempty cycles in the corresponding systems.

Given two systems of the above-described type with (possibly) different arrival times of customers and the service policies, we shall write $T \le T'$ a.s., provided that $T_n \le T'_n$ a.s. for all n, and write $f \ge f'$ a.s., provided that $f_k^j(x) \ge f'^j_k(x)$ a.s. for all $k = 1, 2 \ldots$, $j = 1, 2, \ldots$; $x \in \mathbf{Z}^+$.

The above objects enjoy the *monotonicity property*:
If $T \le T'$ and $f \ge f'$ a.s. then

$$X_{[m,n]}(T) \le X_{[m,n]}(T'), \quad \widetilde{X}_{[-k,l]}(T) \le \widetilde{X}_{[-k,l]}(T').$$

Introduce the following notations:

$$Z_{[m,n]}(T) = X_{[m,n]}(T) - T_n, \quad \widetilde{Z}_{[-n,m]}(T) = \widetilde{X}_{[-n,m]}(T) - T_m,$$

$$X_1(T) = X_{[1,1]}(T), \quad Z_1(T) = Z_{[1,1]}(T).$$

Assume the following condition to be satisfied:
(A_1) $\mathbf{E}X_1(T) < \infty$.
The condition (A_1) is always valid for systems with finite number of stations.
Under condition (A_1), we have

Lemma 1 (Law of Large Numbers). There exists a finite constant $\gamma \ge 0$ such that

$$\frac{Z_{[1,n]}}{n} \xrightarrow{\text{p}} \gamma, \quad \lim_{n\to\infty} \frac{\mathbf{E}Z_{[1,n]}}{n} = \gamma;$$

$$\frac{Z_{[-n,-1]}}{n} \xrightarrow{\text{p}} \gamma, \quad \lim_{n\to\infty} \frac{\mathbf{E}Z_{[-n,-1]}}{n} = \gamma.$$

Given an arbitrary $0 \le c < \infty$, denote by cT the process that consists of the points $\{cT_i\}, i \in \mathbf{Z}$, and the marks $(c\tau_i, \mu_i, \sigma_i)$. The monotonicity property and (A_1) imply

Lemma 2 . *For every $c \ge 0$, there exists a nonnegative constant $\gamma(c)$ such that*

$$\frac{Z_{[1,n]}(cT)}{n} \xrightarrow{\text{p}} \gamma(c);$$

moreover, $\gamma(c)$ decreases in c, whereas $\gamma(c) + c\lambda^{-1}$ increases in c.

Denote

$$\gamma(0) = \lim_{c \searrow 0} \gamma(c) = \lim_{n} \frac{Z_{[1,n]}(0 \cdot T)}{n}.$$

Theorem 1 . *1.* $\gamma(0) = \sigma + \sup_k \frac{p_k}{F_k C_k} W$.

2. There exists $\lim_n \widetilde{X}_{[-n,0]}$ *in the sense of convergence a.s.*

3. The event $\{\lim_n \widetilde{X}_{[-n,0]} < \infty\}$ *has probability 0 or 1.*

4. Let condition (A_1) *hold. If* $\lim \widetilde{X}_{[-n,0]}(T) = \infty$ *a.s. then* $\rho \equiv \lambda\gamma(0) \geq 1$. *If* $\rho > 1$ *then* $\lim \widetilde{X}_{[-n,0]}(T) = \infty$ *a.s.*

For the system governed by the process $\Psi^{(-n)}$, let $Q^k_{[-n,m]}(t)$ stand for the queue length at station k at time t; $\chi^k_{[-n,m]}(t)$, the residual service time at station k at time t; $\chi^0_{[-n,m]}(t)$, the residual interarrival time; $\eta_{[-n,m]}(t)$, the cycle of the route in which the server is at time t (the random vector composed of the numbers of stations and the walking times between them); and $\varphi_{[-n,m]}(t)$, the residual (total) walking time of the server in the cycle $\eta_{[-n,m]}(t)$.

Set the corresponding quantities equal to zero if their values at time t are not defined. All above characteristics are assumed right continuous. Put

$$Y_{[-n,m]} = \{\{Q^k_{[-n,m]}(t)\}^\infty_{k=1}, \{\chi^k_{[-n,m]}(t)\}^\infty_{k=0}, \eta_{[-n,m]}(t), \varphi_{[-n,m]}(t), 0 \leq t \leq T_m\}.$$

Given random variables X and Y on the probability space $\langle\Omega, \mathcal{F}, P\rangle$, we call X a *copy* of Y if there exits a one-to-one measure-preserving $\mathcal{F}$-measurable shift transformation θ on Ω such that $X(\omega) = Y(\theta\omega)$ for all ω.

We say that the process $\breve{X}(t)$ is *Palm-stationary* (with respect to the nested times $\{T_n\}$, where $T_0 = 0$) if for every n the process $\{\breve{X}^n(t) = \breve{X}(t + T_n),\ t \geq 0\}$ is a copy of the process $\{\breve{X}(t), t \geq 0\}$.

For the process $\{\breve{X}(t),\ \infty \leq t \leq \infty\}$ and for any $m = 1, 2, \ldots$, put $\breve{Y}^m = \{\breve{X}(t), 0 \leq t \leq T_m\}$.

Denote by $Q^k_n = Q^k_{[1,n]}(T_n)$ the queue length at station k at time T_n in the polling system governed by the process Ψ, set $\vec{Q}_n = \{Q^k_n; k = 1, 2, \ldots\}$. Let $Q_n = \sum_{k=1}^\infty Q^k_n$ stand for the total queue length.

Theorem 2 . *Assume that condition* (A_1) *holds.*

1. If $\rho < 1$ *then, on the probability space* $\langle\Omega, \mathcal{F}, P\rangle$, *there exists a Palm-stationary process* $\{\breve{X}(t),\ \infty \leq t \leq \infty\}$ *such that for every* $m = 1, 2, \ldots$ *there is a sequence* $\{\breve{Y}^{n,m}\}^\infty_{n=1}$ *of copies of the process* $\breve{Y}^m$ *for which*

$$\mathbf{P}(Y_{[-n,m]} \neq \breve{Y}^{n,m}) \to 0 \quad \text{as} \quad \text{n} \to \infty.$$

In particular, there exists s stationary sequence $\{\vec{Q}^{(n)}; \infty < n < \infty\}$ *such that*

$$\mathbf{P}(\vec{Q}_n = \vec{Q}^{(n)}) \to 1$$

as $n \to \infty$.

2. If $\rho > 1$ *then there exists* $k < \infty$ *such that* $\sum_{j=1}^k Q^j_n \to \infty$ *a.s. as* $n \to \infty$.

Remark 1. It is noteworthy that the existence of the Palm version of a stationary process implies the existence of a stationary process in continuous time (and vice versa) and some formulas are known that connect the distributions of these marked point processes (see, for instance, [14]).

Remark 2. Extend the class of the service policies under consideration as follows: Let $\widehat{B} = \{f :$ for every y there is $\lim_{x\to\infty} f(x, y) = F(y) \leq \infty\}$ and $B = \{f :$ for every y there is $\lim_{x\to\infty} f(x, y) = F(y) \leq \infty$, $f(x, y) \leq F(y)$ for all $y \in \mathbf{R}$, $x \in \mathbf{Z}^+\}$.

It is easy to see that $M \subset B \subset \widehat{B}$. For the service policies in the class $\widehat{B}$, there exists

$$F_k = \mathbf{E} \lim_{x\to\infty} f_k(x, D_k^j) = \mathbf{E}F_k(D_k^j) \leq \infty.$$

Theorem 3 *Consider a system with service policies in the class $\widehat{B}$. If $\rho < 1$ then Q_n is bounded in probability; i.e.,*

$$\lim_{x\to\infty} \sup_n \mathbf{P}(Q_n > x) = 0.$$

Theorem 4 *Consider a system with policies in the class B. If $\rho > 1$ then there exists $k < \infty$ such that $\sum_{j=1}^k Q_n^j \xrightarrow{\mathrm{P}} \infty$ as $n \to \infty$.*

The claim of Theorem 4 fails (in general) for systems with policies in the class $\widehat{B}$.

§2. Systems with a Single Server: "Local" Stability

Consider a model with $K < \infty$ stations. Assume that $\rho > 1$, i.e. the "global" system is unstable. The problem is: do some stable station still exist? We give a positive answer on this question, but under slightly more restrictive assumptions on distributions of driving sequences.

Assume that service time at each station k form a stationary ergodic sequence $\{\sigma_n(k)\}$ with finite mean $\sigma(k)$, all these sequences are mutually independent and do not depend on the sequence $\{(\tau_n, \mu_n)\}$. Put $a_k = \frac{p_k}{F_k C_k} \geq 0$ and permit stations in such an order that

$$a_1 \leq a_2 \leq \ldots \leq a_K.$$

For $k = 1, \ldots, K$, set

$$\rho_k = \lambda(\sum_{j=1}^{k} \sigma(j)p_j + a_k(W + \sum_{j=k+1}^{K} \sigma_j p_j F_j C_j))$$

(here $0 \times \infty = 0$). Note that $\rho_k \leq \rho_{k+1}$ for all k, and $\rho_K = \rho$.

Theorem 5 *If $\rho_k < 1$ for some $k = 1, \ldots, K$, then there exists a stationary k-dimensional sequence $\{\vec{Q}^{(n)}(k)\}$ such that*

$$\mathbf{P}((Q_n^1, \ldots, Q_n^k) = \vec{Q}^{(n)}(k)) \to 1$$

as $n \to \infty$.

Under assumption (A_1), a similar result takes place for systems with infinite number of stations. One can formulate natural analogs of Theorem 1-4 also.

§3. Systems with Finite Number of Stations and with Several Servers

Consider a model with $K < \infty$ stations. Assume that interarrival times $\{\tau_n\}$ form an i.i.d. sequence with mean λ^{-1}, each customer (independently of everything else) is sent to station $k = 1, \ldots, K$ with probability p_k. Each server has its own regenerative routing mechanism and service policies. All service policies are assumed to be limited: $F_k^{(m)} < \infty$ for any server m and for any station k. Within one cycle, a server m visits any station k a random number of times with finite mean $C_k^{(m)}$, and its mean cycle walking time is $W^{(m)}$.

Theorem 6 *Assume interarrival times to have an unbounded distribution. Then the model is stable if and only if* $\tilde{\rho} < 1$.

The "unboundedness" assumption may be weakened variously. Here $\tilde{\rho} = \lambda \sum_{r=1}^{R} \beta^{(r)}$, and R and $\beta^{(r)}$, $r = 1, \ldots, R$ are defined by the following recursive procedure.

For $1 \le i, j \le K$, $1 \le m \le M$, put

$$\varphi_{i,j} = \sum_{m=1}^{M} \frac{F_i^{(m)} C_i^{(m)}}{W^{(m)} + \sum_1^j \sigma_k^{(j)} F_k^{(j)} C_k^{(j)}}.$$

Set

$$p_i^{(1)} = p_i; \varphi_1^{(i)} = \varphi_{i,K}; \alpha_i^{(1)} = \frac{p_i^{(1)}}{\varphi_i^{(1)}}; \beta^{(1)} = \min_{1 \le i \le K} \alpha_i^{(1)}.$$

Take $K(1) = K$. Assume that, for some $k \ge 1$, we have previously defined $K(1) > K(2) > \ldots > K(r) \ge 1$ and, for $1 \le k \le r$, $1 \le i \le K(k)$, we know the values of $p_i^{(k)}, \varphi_i^{(k)}, \alpha_i^{(k)}$ and $\beta^{(k)} = \min_i \alpha_i^{(k)}$.

Permit a sequence $\{1, 2, \ldots, K(r)\}$ in such a way that (after permutation)

$$\alpha_1^{(r)} \ge \alpha_2^{(r)} \ge \ldots \ge \alpha_{K(r)}^{(r)}.$$

If $\alpha_1^{(r)} = \alpha_{K(r)}^{(r)}$, then stop with the procedure and put $R = r$. Otherwise, set

$$K(r+1) = \max\{l : \quad \alpha_l^{(r)} > \alpha_{l+1}^{(r)}\}$$

and, for any $1 \le i \le K(r+1)$, put

$$p_i^{(r+1)} = \varphi_i^{(r)}(\alpha_i^{(r)} - \alpha_{K(r)}^{(r)}); \varphi_i^{(r+1)} = \varphi_{i,K(r+1)};$$

$$\alpha_i^{(r+1)} = \frac{p_i^{(r+1)}}{\varphi_i^{(r+1)}}; \beta^{(r+1)} = \min_i \alpha_i^{(r+1)}.$$

References

[1] H. Lévy, M. Sidi, and O. J. Boxma, "Dominance relations in polling systems," Queueing Systems, **6**, 155–171 (1990).

[2] L. Massoulie, "Stability of nonmarkovian polling systems," in: Abstracts: Proc. Conf. Appl. Probab., Paris, 1993. (Submitted to Queueing Systems.)

[3] A. A. Borovkov and R. Schassberger, "Ergodicity of a polling network," Stochastic Process Appl., **50**, 253–262 (1994).

[4] S. Foss and G. Last, "On the stability of greedy polling systems with exhaustive service policies," in: Technical Report 18/94, Technical University Braunschweig, 1994. (To appear in Ann. Appl. Probab.)

[5] S. Foss and G. Last, "On the stability of greedy polling systems with general service policies," in: Technical Report 6/95, Technical University Braunschweig, 1995.

[6] C. Fricker and M. R. Jaibi, "Monotonicity and stability of periodic polling models," Queueing Systems, **15**, 211–238 (1994).

[7] L. Georgiadis and W. Szpankowski, "Stability of token passing rings," Queueing Systems, **11**, 7–34 (1992).

[8] D. P. Kroese and V. Schmidt, "A continuous polling system with general service times," Ann. Appl. Probab., **2**, 906–927 (1992).

[9] D. P. Kroese and V. Schmidt, "Single-server queues with spatially distributed arrivals," Queueing Systems, **17**, 317–345 (1994).

[10] R. Schassberger, "Stability of polling networks with state-dependent server routing," in: Technical Report 12/93, Technical University Braunschweig, 1993.

[11] V. Sharma, "Stability and continuity of polling systems," Queueing Systems, **16**, 115–137 (1994).

[12] H. Takagi, Queueing Analysis of Polling Systems in: Stochastic Analysis of Computer and Communications System, Hideaki Takagi (ed.), Elsevier Science Pub. Horth-Holland (1990).

[13] F. Baccelli and S. Foss, "On the saturation rule for the stability of queues," J. Appl. Probab., **32**, No. 2, 494–507 (1995).

[14] F. Baccelli and P. Bremaud, Elements of Queueing Theory, Springer, Berlin (1994).

[15] S. Foss and N. Chernova, " Comparison theorems and ergodic properties of polling systems," Preprint No. 6, Institute of Mathematics, Novosibirsk (1995) - to appear in Problems of Information Transmission (1996).

[16] S. Foss and N. Chernova, "On polling systems with infinitely many stations," Siberian Math. J., **37**, 940-956 (1996).

[17] S. Foss and N. Chernova, "On a local stability in polling models" (in preparation).

[18] A. Kovalevskii and S. Foss, "Stability of polling models with many servers" (in preparation).