Stability analysis of regenerative queues: recent results

Evsey Morozov

Inst. Appl. Math. Research

Karelian Research Centre, Russian Academy of Sciences;

Petrozavodsk University

- Stability analysis of modern queueing systems: a challenging problem;
- main known approaches: fluid approach, Lyapunov functions (for Markovian models);
- we present regenerative stability analysis (see [1]):
  - describe the main steps;
  - give stability conditions for some queueing models;
  - an advantage: can be applied to non-Markovian models.

Definitions

- Consider a stochastic process

$X = \{X(t),\, t \in T\}$, $T = [0,\, \infty)$ or $T = \{0, 1, \ldots\}$, state space $(R^d,\, \mathcal{B})$

- a path is divided by *regeneration points* (r.p.'s) β_n onto iid *regeneration cycles*:

$$G_n \;:=\; \{X(t):\; \beta_n \le t < \beta_{n+1};\, \beta_{n+1} - \beta_n,\},\; n \ge 0,\; \beta_0 := 0,$$

with the iid *cycle periods* $\beta_{n+1} - \beta_n$ (in general, $G_0 \neq_{st} G_n$).

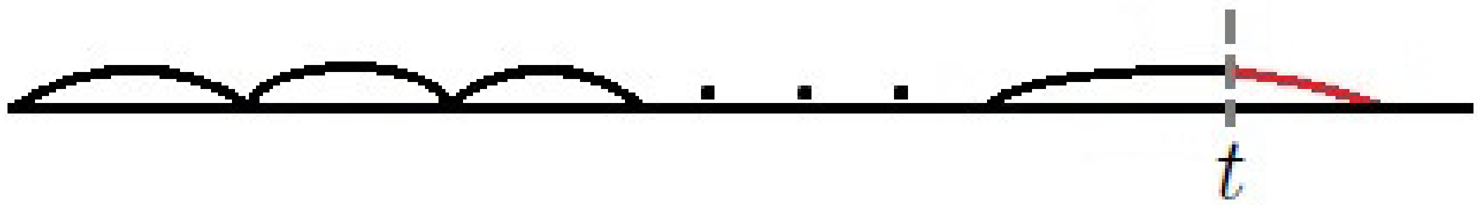
- *Regeneration measure* μ: distribution of $X(\beta_n)$; (for queues often $\mu\{0\} = 1$)

- The renewal process $\{\beta_n\}$ is *positive recurrent* if (β is generic cycle period)

$$\mathsf{E}_\mu\beta := \mathsf{E}\beta < \infty \quad (\text{and } \beta_1 < \infty \text{ with probability (w.p.) } 1). \tag{1}$$

- If $\mathsf{E}\beta = \infty$ then for each $\beta(0) = x$, the forward renewal time at instant t [9]:

$$\beta(t) := \min\{\beta_k - t : \beta_k - t > 0\} \to \infty \text{ in } P_x\text{-probability}$$



• (1) is crucial for stability because: for a measurable f: $\mathsf{E}\{\int_0^\beta \mid f(X(t)) \mid dt\} < \infty$:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(X(t))dt = \frac{\mathsf{E}\int_0^\beta f(X(t))dt}{\mathsf{E}\beta}(= \mathsf{E}f(X) \text{ if weak limit } X(t) \Rightarrow X \text{ exists}).$$

Main steps of the proof: using predefined conditions we show:

i) *negative drift condition*: X has *negative drift* outside bounded set B;

ii) *regeneration condition*:

$$\inf_{x\in B} \mathsf{P}_x(X \text{ regenerates within a finite interval}) > 0;$$

• i), ii) imply: $\beta(t) \not\to \infty$ and $\mathsf{E}\beta < \infty$;

An example: Kiefer-Wolfowitz workload (Markov) process in a $GI/G/m$ queue:

$$W_n = (W_n^{(1)}, \ldots, W_n^{(m)}),\ n \geq 0,$$

• $W_n^{(i)}$ is the ith smallest workload at arrival instant of customer n

• the renewal input with rate λ, service time S; m parallel servers;

$i)$ negative drift condition: $\rho = \lambda \mathsf{E} S < m$;

$ii)$ regeneration condition: $\mathsf{P}(\tau > S) := \varepsilon > 0$.

Then for any $W_0 = x$ $\exists$ bounded B, $\varepsilon_1 > 0$ and deterministic $n_i \to \infty$:

$$\inf_i \mathsf{P}_x(W_{n_i} \in B) > 0; \tag{2}$$

• $\exists$ constants $T_B = T$:

$$\inf_{x \in B} \mathsf{P}_x(\beta_1 \le T) > 0; \tag{3}$$

• Then (discrete-time) forward regeneration(=renewal) time $\beta(n)$ satisfies

$$\inf_i \mathsf{P}_x(\beta(n_i) \le T) \ge \inf_i \mathsf{P}_x(W_{n_i} \in B) \inf_{z \in B} \mathsf{P}_z(\beta_1 \le T) > 0 \quad \Rightarrow \quad \mathsf{E}\beta < \infty.$$

- Using ii) one can show that for $\forall$ bounded B:

$$\nu_x(B) = \mathsf{E}_x(\sum_{k=0}^{\beta-1} I_{W_k \in B}) \leq T/\varepsilon,$$

and set B has finite *stationary measure* (mean number of visits B during cycle)

$$\nu(B) \ = \ \int \nu_x(B)\mu(dx) < \infty. \tag{4}$$

- *Null recurrence*: infinite mean cycle period, $\nu(\mathsf{R}_+^m) = \mathsf{E}_\mu \beta = \infty$;

$$\mathsf{P}_x(W_n \in B) \to 0, \quad \text{as } n \to \infty, \quad \forall\ B,\ x. \tag{5}$$

Similarly for null recurrent Harris Markov chains [4].

• An advantage: dimension reduction (non-Markovity): the same r.p's

replacement $X_n \in R^d$ by $Z_n = \sum_{1\le i\le d} X_n^{(i)}$.

A novel approach to prove $\beta_1 < \infty$ under arbitrary $X(0)$; we show:

• the total time $X(t)$ spends in a set B during 1st cycle is finite w.p.1;

• the total time $X(t)$ spends in B is infinite w.p.1 $\Rightarrow$

• the number of regeneration cycles $\ge 2 \;\; \Rightarrow \beta_1 < \infty$.

Recent queueing applications

- Main technique: coupling $\Rightarrow$ monotonicity/dominance

- **Classical m-server finite retrial system** $GI/G/m/n$ with infinite orbit:

renewal input rate λ, interarrival time τ; service time S; i.i.d. exponential retrial times;

Theorem 1 [2]. *Under conditions*

$$\lambda \mathsf{E} S < m, \quad \mathsf{P}(\tau > S) > 0$$

the basic processes (e.g. orbit size $N(t)$) are positive recurrent.

Underlying idea: as $N(t) \to \infty$, the idle time of server after each service goes to 0:

Discipline: *asymptotically work-conserving*.

- **Retrial system Σ with constant retrial rate μ_0 and exponential retrial time** (*joint work with INRIA, Sophia-Antipolis, France*)

  renewal input rate λ; service rate μ; finite buffer, c servers;

- Motivation: modeling telephone exchange systems [8]; short TCP transfers [6];

  ALOHA multiple access protocols [7]: n retrial customers, each has rate μ_0/n.

- To find stability condition, consider an auxiliary loss system $\hat{\Sigma}$:

- *an independent Poisson input with rate* μ_0;

- lost customers joins a virtual orbit $\hat{N}(t)$ with the "output rate" μ_0 (not affecting $\hat{\Sigma}$)

- Σ is **less loaded** than $\hat{\Sigma}$ $\Rightarrow N(t) \leq_{st} \hat{N}(t) \Rightarrow$

- stability of $\hat{N}(t)$ implies stability of $N(t)$ [5].

- Let $\hat{\mathsf{P}}_{loss}$ be the stationary loss probability in $\hat{\Sigma}$ (always exists).

• **Theorem** ([5]). *Under negative drift assumption*

$$\textit{input to } \hat{N}(t) \ = \lambda \hat{\mathsf{P}}_{loss} < \mu_0(1 - \hat{\mathsf{P}}_{loss}) = \mu_0 \hat{\mathsf{P}}_{accept} = \textit{output from } \ \hat{N}(t), \qquad (6)$$

the basic regenerative process $\{N(t)\}$ *is positive recurrent.*

((6) is stability criteria for M/G/n-type retrial system).

• *Example:* for M/G/1/1-type (Erlang) system condition (6) reads

$$\rho := \frac{\lambda}{\mu} < \frac{\mu_0}{\lambda + \mu_0} \quad \Rightarrow \quad \text{becomes classic: } \rho < 1, \text{ as } \ \mu_0 \to \infty. \qquad (7)$$

The proof is based on monotonicity property of loss system [11, 12, 13].

- N-orbit retrial system: input rates λ_i; retrial rates $\mu_0^{(i)}$; service rates μ_i;

- necessary stability condition for $N = 2$:

$$\lambda_i \mathsf{P}_{loss} < (1 - \mathsf{P}_{loss})\mu_0^{(i)}, \qquad i = 1, 2. \tag{8}$$

P_{loss} - loss probability in auxiliary loss system with total input rate

$$\lambda_1 + \lambda_2 + \mu_0^{(1)} + \mu_0^{(2)}.$$

- P_{loss} is known for some systems; ((8) is stability criteria for Poisson inputs ?)

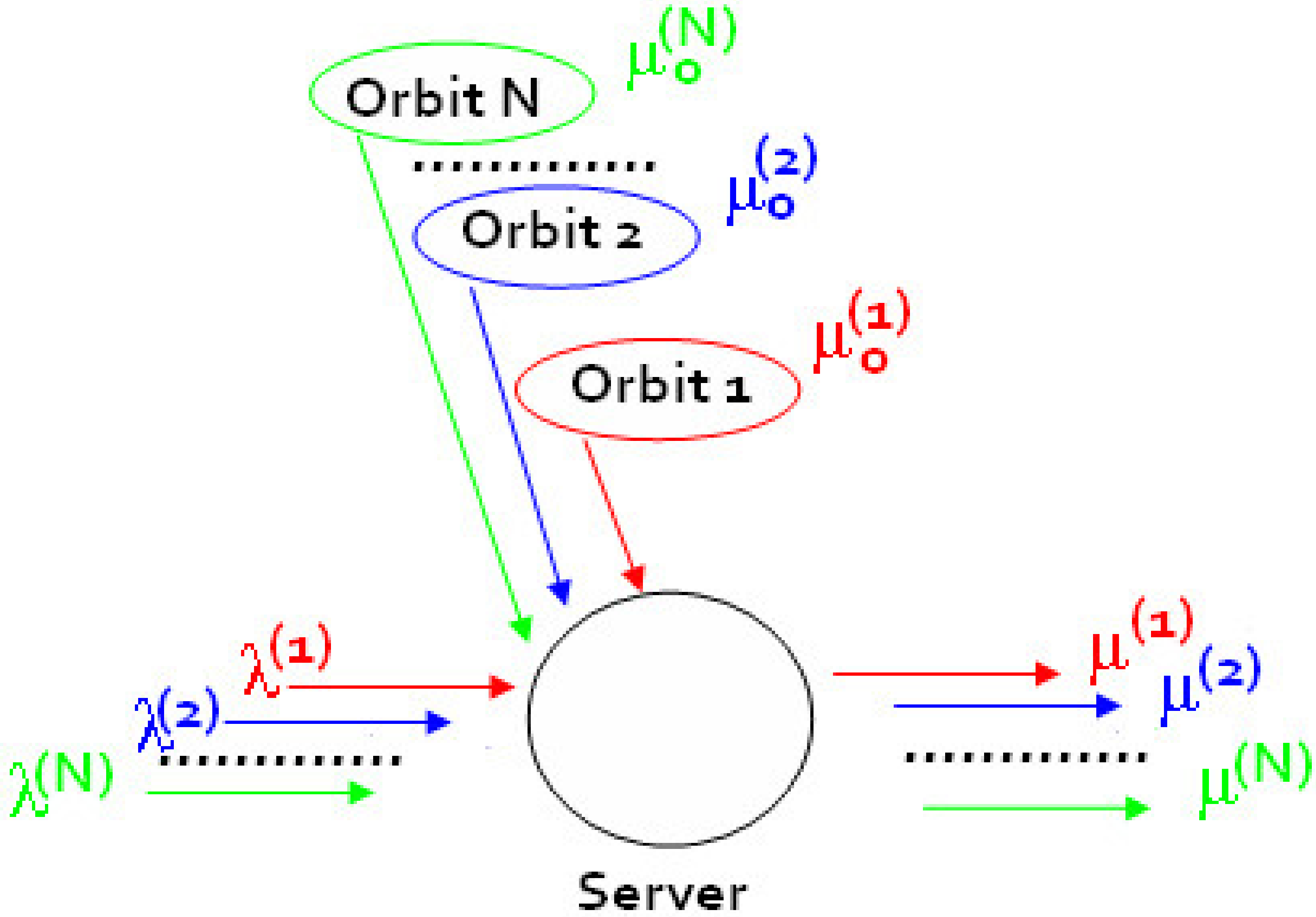


Figure 1: N-orbit retrial system with constant retrial rates $\mu_0^{(i)}$

- **Discrete queue with m non-reliable servers:** input rate λ; service time S, server i: interruption rate $\lambda_0^{(i)}$ (non-Poisson), repair time $X^{(i)}$;

- preemptive repeat service interruption: stability condition

$$\lambda \mathsf{E}S + \mathsf{E}S \sum_{i=1}^{m} \lambda_0^{(i)} + \sum_{i=1}^{m} \lambda_0^{(i)} \mathsf{E}X^{(i)} < m\,, \tag{9}$$

lost capacity: interruption$+$ repeat service$=\sum_{i=1}^{m} \lambda_0^{(i)} \mathsf{E}X^{(i)} + \mathsf{E}S \sum_{i=1}^{m} \lambda_0^{(i)}$;

- preemptive resume service interruption: stability condition

$$\lambda \mathsf{E}S + \sum_{i=1}^{m} \lambda_0^{(i)} \mathsf{E}X^{(i)} < m. \tag{10}$$

- main idea: construction of common renewal points for the superposition of m independent discrete work/repair processes [3].

- m**-wavelength system with optical fiber lines** (*joint with UGENT*):

lengths $a_0 < a_1 < \cdots, \ \lim a_n = \infty$:

renewal input rate λ, interarrival time T; transmission time S;

$$g_n \;=\; a_{n+1} - a_n = \text{difference of adjacent lengths};$$

$$\Delta^* = \max g_n < \infty;$$

$$\Delta_0 \;=\; \limsup_{n\to\infty} g_n = \text{ difference between "longest" lines/orbits} \tag{11}$$

- buffer control uses a FIFO rule to choose the shortest line with sufficient length $\Rightarrow$

- a modified Lindley's recursion for waiting time W_k of the kth signal:

$$W_{k+1} = \lceil W_k + S_k - T_k \rceil_A \tag{12}$$

where $A = \{a_i\},\ \lceil x \rceil_A = \inf\{y \in A : y \geq x\}$ (to keep FIFO order).

- Classical regenerations for $\{W_k\}$:

$$\beta_0 = 0,\ \ \beta_{n+1} = \inf(k > \beta_n : W_k = 0),\ \ n \geq 0.$$

- Theorem ([14, 15, 16]). *The workload process is positive recurrent if*

$$\lambda \mathsf{E} S + \lambda \Delta_0 < m, \quad \textit{equivalently,} \tag{13}$$

arrived load + lost capacity (for busy high orbits) < system capacity

- *and (regeneration) condition holds:* $\mathsf{P}(T > \Delta^* + S) > 0.$

- **Remark.** For iid (or constant) distances $g_n =_{st} \Delta$: inspection paradox makes more exact stability zone by replacement in (13):

$$\Delta_0 \to \frac{\mathsf{E}\Delta^2}{2\mathsf{E}\Delta}.$$

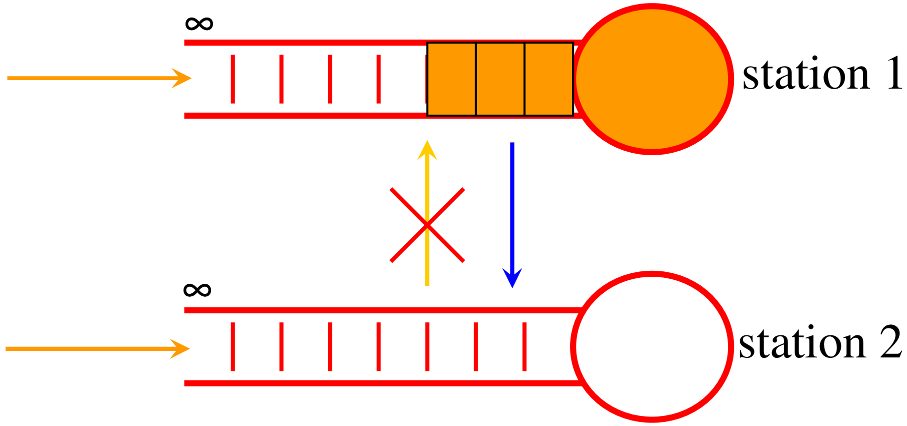


Figure 2: A cascade network with two stations

- **2 -station cascade system** [17] : stability condition (negative drift)

$$\mu_2 > \lambda_2 + (\lambda_1 - \mu_1)^+, \tag{14}$$

- capacity of 2nd station $>$ input to 2nd station$+$extra rate from 1st station;

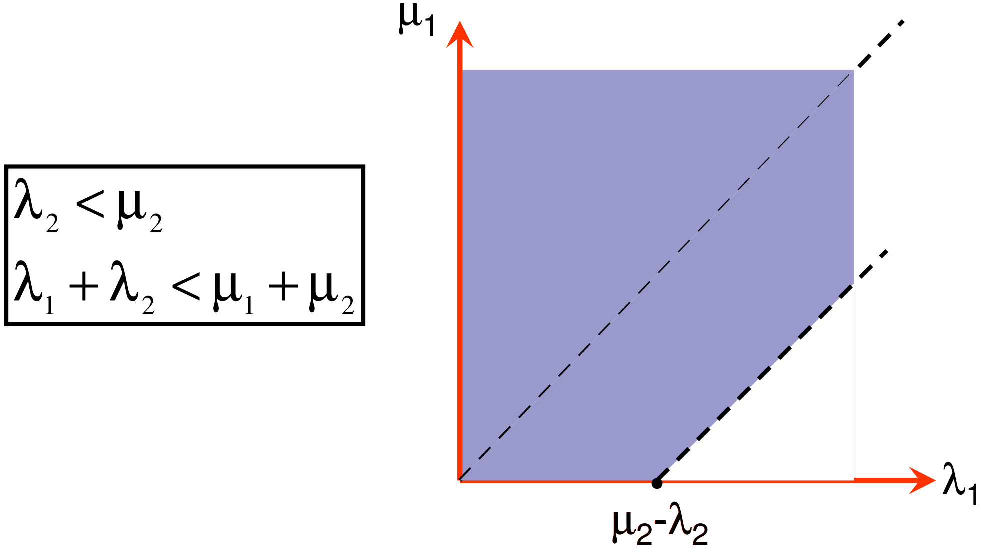


Figure 3: Stability region (coloured area) of a cascade network

- equivalent form of negative drift:

$$\mu_2 - \lambda_2 > 0, \quad \lambda_1 < \mu_2 - \lambda_2 + \mu_1$$

Conclusion

obtained conditions are transparent and close to being stability criteria;

method: applicability to non-Markovian processes;

a wide class *recurrent Harris Markov processes* posses regenerative property [4];

confidence estimation: regenerative simulation based on regenerative CLT [18]:

group data over cycles; apply classic CLT for the new iid enlarged data;

new (macroscopic/cycle) scale: simulation up to an integer number of cycles.

References

[1] E. Morozov and R. Delgado, Stability analysis of regenerative queueing systems, Automation and Remote control, v. 70, No. 12, 1977-1991 (2009).

[2] E. Morozov, A multiserver retrial queue: regenerative stability analysis, Queueing Systems, v. 56, 157-168 (2007).

[3] E. Morozov, D. Fiems, and H. Bruneel, "Stability analysis of multiserver discrete-time queueing systems with renewal-type server interruptions", *Performance Evaluation*, v. 68, 1261-1275 (2011).

[4] S. Asmussen (2003) Applied Probability and Queues, 2nd ed., Springer, NY.

[5] Avrachenkov K. , Morozov E. V. Stability analysis of GI/G/c/K Retrial Queue with Constant Retrial Rate, INRIA, Res.Rep. No 7335, July, (2010).

[6] K. Avrachenkov and U. Yechiali, “Retrial networks with finite buffers and their application to Internet data traffic”, *Probability in the Engineering and Informational Sciences*, v.22, pp.519-536 (2008).

[7] B.D. Choi, K.H. Rhee and K.K. Park, “The M/G/1 retrial queue with retrial rate control policy”, *Probability in the Engineering and Informational Sciences*, v.7, pp.2946, (1993).

[8] G. Fayolle, “A simple telephone exchange with delayed feedback”, In O.J. Boxma, J.W. Cohen and H.C. Tijms (eds.), *Teletraffic Analysis and Computer Performance Evaluation*, v.7, pp.245-253, Elsevier, Amsterdam (1986).

[9] W. Feller, An Introduction to Probability Theory and its Applications, Vol.2 (2nd edn), Wiley, NY (1971).

[10] E. Morozov, The tightness in the ergodic analysis of regenerative queueing processes, Queueing Systems, 27, 179-203 (1997).

[11] D. Sonderman, Comparing multi-server queues with finite waitng rooms, I: Same number of servers, AAP, v.11, pp.439-447 (1979).

[12] D. Sonderman, Comparing multi-server queues with finite waitng rooms, II: Different number of servers, AAP, v.11, pp.448-455 (1979).

[13] W. Whitt, Comparing counting processes and queues, AAP, v.13, pp.207-220 (1981).

[14] W. Rogiest, E. Morozov, D. Fiens, K. Laevens, H. Bruneel. Stability of Single-Wavelength Optical Buffers, European Transactions on Telecommunications (ETT) 21(3), 202-212, (2010).

[15] E. Morozov, W. Rogiest, K. De Turck, D. Fiems. Stability of multiwavelength optical buffers with delay-oriented scheduling Transactions on Emerging Telecommunications Technologies Vol 23, Issue 3, pp. 217 -226, April (2012).

[16] K. Avrachenkov, E. Morozov, R. Nekrasova, B. Steyaert, Stability analysis and simulation of a N-class retrial system with constant retrial rates and Poisson inputs, Asia-Pacific Journal of Operations Research (accepted).

[17] E. Morozov, B. Steyaert. Stability analysis of a two-station cascade queueing network, Annals of Operations Research (2013)

[18] G. Shedler, Stochastic Regenerative Simulation, Academic Press Inc. (1993)