SYSTEMS WITH DYNAMIC ROUTING

N.D. Vvedenskaya

Institute for Information Transmission Problems RAS, Moscow

We consider a network with k servers and ℓ Poisson input flows, each flow is assigned to some subgroup S_j of m_j servers. A dynamic routing protocol is used: upon its arrival a message is directed to the least loaded server.

The multi-server systems with dynamic routing protocol were investigated in a number of papers, see for example [1]-[9], in particular different asymptotic properties of such systems were investigated.

Consider a system $\mathcal{S}$ with k servers $\{s_i\}$, $i = 1, ..., k$, and ℓ input Poisson flows $\{f_j\}$, $j = 1, ..., \ell$, of intensity λ_j. Each flow is assigned to some subgroup S_j of m_j, servers $s_{j_1}, ..., s_{j_{m_j}}$, $2 \le m_j \le k$. Upon arrival a message selects a least busy server among the assigned ones and is directed to it (or is put into the server's buffer if the server is busy). The service is FCFS. Here "least busy" may mean the least loaded one (where a new message will have minimal waiting time) or the one with the shortest queue.

The flows are presented by the sequences

$$(\xi_n^{(j)}, \tau_n^{(j)}), \quad n = ..., -1, 0, 1, ..., \quad j = 1, ..., \ell, \tag{1}$$

where $\tau_n^{(j)}$ are the intervals between arrivals of two messages and $\xi_n^{(j)}$ - the message lengths. All variables are iid, $\tau_n^{(j)}$ are exponentially distributed, $\Pr(\tau_n^{(j)} > t) = e^{-\lambda_j t}$. The distributions of message lengths $\xi_n^{(j)}$ are identical. The service rate of a server is equal to μ.

Examples: I) k servers form a circle where each of $\ell = k$ flows is served by two neighboring servers. II) The system contains k servers, each of $\ell = \binom{k}{m}$ flows is served by m servers.

Denote by $w^{(j)}(t)$ the workload (the amount of unserved work) at server s_j at time moment t; $\boldsymbol{w}(t) = (w^{(1)}(t), ..., w^{(k)}(t))$, define by $v^{(j)}(t)$ the queue length (the number of messages) at the buffer of server s_j at time moment t (the served massage is also counted).

The system is stable if $\sum_{r=1}^{n} \lambda_{j_r} \mathbb{E}\xi < \mu \bar{l}$ for any set $\overline{S} = \bigcup_{r=1}^{n} S_{j_r}$, $1 \le n \le \ell$, that contains $\bar{l}$ servers.

We want to present a property of systems with dynamic routing that does not depend on stability.

We say that the system $\mathcal{S}$ is balanced if there exist such

$$\alpha(j, i_r),\ 0 \le \alpha(j, i_r) \le 1,\ j = 1, ..., \ell,\ i_r = i_1, ..., i_{m_j}$$

that

$$\sum_{r=1}^{m_j} \alpha(j, i_r) = 1,\ \sum_j \lambda_j \alpha(j, i) = \overline{\lambda} = const.$$

(Here j is number of flow and i – the number of server).

The system is connected if for any pair of servers s_p, s_q there exist such servers $s_p = s_{i_0}, s_{i_1}, \ldots, s_{i_n} = s_q$ and flows $f_{j_1}, ..., f_{j_n}$ that $\alpha(j_i, r_{i-1}), \alpha(j_i, r_i) > \hat{\alpha} > 0, \quad 1 \le i \le n.$

Let us first consider a case where the workload of servers guides the routing. Suppose that $\xi_n^{(j)}$ have finite second moments. The trajectories $\boldsymbol{w}(t) = (w^{(1)}(t), ..., w^{(k)}(t))$ define a Markov process U, $U \in \mathbb{R}^k$.

Let $(\mathbf{0}, \boldsymbol{w}_P(t))$ be the projection of vector $(\mathbf{0}, \boldsymbol{w}(t)) \in \mathbb{R}^k$ onto the hyperplane P that is orthogonal to the bisectrix $(w(1) = ... = w(k))$, $\boldsymbol{w}_P(t) \in \mathbb{R}^{k-1}$.

Consider a Markov processes $U_P\ :\ \boldsymbol{w}_P(t)$ of dimension $k-1$.

Theorem 1 *If the system is balanced and connected then the Markov process U_P is ergodic.*

That means that the trajectories $\boldsymbol{w}(t)$ are mainly concentrated in the neighborhood of bisectrix $(w(1) = ... = w(k))$ even in case where the process is not stable and vector $(\mathbf{0}, \boldsymbol{w}(t))$ grows with t.

The proof is based on the following property:

If at moment t of a message arrival $\max_{i,j} |w^{(i)}(t) - w^{(j)(t)}|$ is sufficiently large then $\mathbb{E}(|\boldsymbol{w}_P(t+0)| - |\boldsymbol{w}_P(t-0)|) < c < 0$. Here $|\boldsymbol{w}|$ is the norm of vector $(\mathbf{0}, \boldsymbol{w})$, see [10], [11].

Consider a case where the queue lengths guide the routing. Suppose that $\xi_n^{(j)}$ are distributed exponentially with mean equal to 1. The trajectories $\boldsymbol{v}(t) = (v^{(1)}(t), ..., v^{(k)}(t))$ define a Markov process V, $V \in \mathbb{Z}^k$. Let $(0, \boldsymbol{v}_P(t))$ be the projection of $(0, \boldsymbol{v}(t))$ onto the hyperplane P that is orthogonal to the bisectrix $(v(1) = ... = v(k))$.

Theorem 2 *If the system is balanced and connected then the Markov process V_P is ergodic.*

The proof is similar to the proof of Theorem 1.

The case where the messages are of unit length and $\mu = 0$ is presented in [12].

References

[1] *S.Foss, N.Chernova* On a stability of a partially accessible multi-station queue with state-dependent routing // Queueing Systems, 1997, V. 29, P. 55-73.

[2] *M. Alanyali, B. Hajek* On Large Deviations in Load Sharing Networks // Ann. Appl. Probab. 1998, V. 8, ϵ 1, P. 67-97 .

[3] *J.B. Martin and Yu.M. Suhov*, Fast Jackson networks, Ann. Appl. Probability., 1999. V.9, No 4, P.854-840

[4] *D.R. McDonald and S.R.E. Turner*, Resource Pooling in Distributed Queueing Networks // Fields Inst. Commun. V. 28, Providence: Amer. Math. Sci., 2000, P. 107-131.

[5] *R.D. Foley , D.R. McDonald,* Join the Shortest Queue: Stability and Exact Asymptotics // Ann. Appl. Probab. 2001, V. 11, ϵ 3, P. 569-607.

[6] K. Duffy, E.A. Pechersky, Y.M. Suhov, N.D.Vvedenskaya, Using Estimated Entropy in a Queueing System with Dynamic Routing // Markov Process. Related Fields. 2007, V.13, ϵ 1, P. 57-84.

[7] *Y.M. Suhov, N.D. Vvedenskaya*, Fast Networks with Dynamic Probl. Inform. Transmission, V 38, ϵ 2, P.44-63, 2002.

[8] *A.A. Puhalskii, A.A. Vladimirov,* A Large Deviation Principal for Join the Shortest Queue // Math. Oper. Res.2007, V. 32, ϵ 3, P. 700-710.

[9] *N.D. Vvedenskaya, E.A. Pechersky*, Circl of Interacting Servers: Spontanious Collective Fluctuations in Case of Large Fluctuations. Probl. Inform. Transmission, 2008, V. 44, ϵ 4, P. 370-384.

[10] *G. Fayolle, V.A. Malyshev, M.V. Menshikov*Njhics in the Constructive Theory of Countable Markov Chains // Cambridge University Press, 1995.

[11] *S.P. Meyn, R.L.Tweede* Markov Chains and Stochastic Stability // Springer-Verlag, 1993.

[12] *V.V. Menshikov, V.V. Sisko, M. Vashkovskaia* On shape stability for a storage model //arXiv:0709.0121v2[math.PR] 17 Oct 2009.