Systemic Risks/Benefits of Selfish Network Operations & Management in Dynamic Environment

V. Marbukh
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920
E-mail: marbukh@nist.gov

Abstract—**Allowing selfish agents to acquire and exploit system information has both positive and negative effects on the overall performance of resource allocation systems. The positive effect results from reduction in the uncertainty inherently present in large-scale systems. The negative effect, which can be mitigated through congestion pricing, is due to agent selfishness. However, current research, concentrated around the notion of "Price of Anarchy", is mostly concerned with the negative effect. This paper evaluates systemic risks/benefits of selfish agent ability to acquire and exploit dynamic system information in a specific case of selfish routing in a large-scale, random, loss network. Our analysis indicates that the beneficial effect of this ability dominates in a case of high system uncertainty - low load, while the negative effect dominates in a case of low system uncertainty - high load. In the intermediate cases while the beneficial effect still dominates in the "normal" operating mode, the negative effect manifests itself in a risk of cascading overload driving the system to an emergent metastable, i.e., persistent, congested mode. Future research should consider resource allocation models with elastic selfish users and evaluate effect of the congestion pricing.**

Keywords-selfish agents, information availability, systemic risk.

I. Introduction

Assessment of centralized vs. decentralized control schemes in complex distributed systems is of utmost importance for understanding and controlling these systems performance. Computer science and networking communities have developed and investigated several measures, including "Price of Anarchy", quantifying loss in performance due to lack of coordination among selfish agents [1]. These measures, however, do not take into account arguments in favor of decentralization in large-scale system comprised of a large number of components. These arguments, initially articulated in the context of economics [2], are due to inability of a "central planner" to acquire and process all the necessary system information in real time.

This paper evaluates the systemic benefits/risks of decentralization on an example of selfish routing in a loss network. Following [3]-[4] we assume that routing information can be separated into its structural and dynamic components, the former being stable on "much longer" time scale than the latter. In [3]-[4] structural information was identified with network topology while dynamic information was identified with less stable quality of paths across the network. We take a more general view that network parameters may be uncertain and model uncertainty by a probability distribution. This probability distribution is "stable" and can be viewed as structural information available to the central planner who implements the centralized routing. The specific values of the network parameters are more "volatile", and thus constitute the dynamic information, which is unavailable to the central planner but can be discovered by selfish agents in a decentralized manner through probing. In broader context of complex systems, structural information characterizes the macroscopic parameters while dynamic information can be identified with the microscopic system state.

Our analysis under mean-field approximation indicates that risk/benefit trade-off of allowing selfish requests to access and exploit dynamic system information is determined by the system load and uncertainty. The beneficial effect dominates in a case of high system uncertainty – low load, while the opposite is true in a case of low system uncertainty - high load. In the intermediate cases while the beneficial effect still dominates in the "normal" operating mode, the negative effect manifests itself in a risk of cascading overload driving the system to an emergent metastable, i.e., persistent, congested mode. Since metastability is a result of non-linear stochastic system dynamics, conventional resource allocation models may underestimate systemic risks. Note that metastability in large-scale loss network with dynamic routing has been confirmed by simulations and some analytical results [5]-[7].

The paper is organized as follows. Section II describes a model of random loss network. Section III develops mean-field approximation for the network performance. Using this approximation Section IV analyses and interprets this model. Finally Section V briefly summarizes and outlines directions of future research.

II. Loss Random Network

Consider a network with N nodes, where each pair of nodes (n,k), $n,k=1,..,N$, $n \neq k$ is connected by an undirected link l_{nk} of capacity C with probability $1-\pi$ and is not connected with probability π. We assume that existence of links l_{nk} is jointly statistically independent for different origin-destination pairs (n,k), $n,k=1,..,N$, $n\neq k$, and describe network topology by vector of binary variables $\delta=(\delta_{nk})$, where $\delta_{nk}=1$ if link l_{nk} exists and $\delta_{nk}=0$ otherwise.

We also assume that requests with origin-destination (n,k), $n,k=1,..,N$, $n\neq k$ arrive at node n according to a Poisson process of rate λ. Upon arrival a request with origin-destination (n,k) immediately occupies a unit bandwidth on the direct link l_{nk} if this link exists, i.e., $\delta_{nk}=1$, and it has sufficient available bandwidth. Otherwise the request is either lost or routed on a feasible two-link path r_{nmk} where both links (n,m) and (m,k) exist, i.e., $\delta_{nm}=\delta_{mk}=1$, and have sufficient available bandwidth. A routed request simultaneously occupies and then releases a unit of bandwidth on each link along the chosen route. The request holding time is distributed exponentially with unit average.

Our assumption that all requests have the same bandwidth requirements allows us to measure link capacity C in terms of the maximal number of flows the link can accommodate. Thus, $C\in\{1,2,..\}$, and the total number of requests, including direct and transit, carried on each link (n,k), x_{nk} cannot exceed the link capacity: $x_{nk}\leq C$. We formalize the admission and routing strategies by a set of conditional probabilities $q_{nk}(r|\delta,x)$ that a request arriving when the connectivity vector is $\delta=(\delta_{nk})$ and vector of carried requests on all links is $x=(x_{nk})$ is immediately sent to a feasible route $r\in R$ and is lost with probability $q_{nk}(\varnothing|\delta,x)=1-\sum_r q_{nk}(r|\delta,x)$. Probabilities $q_{nk}(r,x)$ satisfy self-consistency conditions, e.g., a request is always accepted on the available direct link, and a request is rejected in neither the direct link or any two-link bypass route is available.

We quantify the long-term system performance by the aggregate loss rate

$$L=\frac{1}{N(N-1)}E_\delta\left[E_X\left(\sum_{(n,k)}q_{nk}(\varnothing|\delta,x)\middle|\delta\right)\right]. \qquad (1)$$

where $E_\delta[.]$ is the mathematical expectation with respect to network topology δ, and $E_X(.|\delta)$ is the steady-state mathematical expectation with respect to process $x(t)$ conditioned on the network topology δ. One may attempt to minimize loss rate (1) over set of conditional probabilities $Q(x)=\{q_{nk}(r|\delta,x)\}$ representing the admission and routing strategies. This solution, however, may not be implementable since it assumes that up to date vectors δ and x are available throughout the network. To model that only local information is available at the point of making the admission and routing decisions, we assume that an arriving at node n request is immediately aware only the components $\delta_{nk},x_{nk},k=1,..,N;k\neq n$ characterizing the current state of all links adjacent to node n. Thus, an arriving request has no difficulty of finding the existing available direct route. However, finding a two-link bypass route requires non-local information and can be done only through probing.

We assume that once an arriving at node n request determines that the direct route is unavailable, this request immediately attempts to find a feasible two-link route r_{nmk}. Since in networking context a route discovery process is typically done through probing, we assume that an arriving request is allowed at most I attempts. Once an available two-link route is found, the probing stops and the request occupies this route. If after I attempts a feasible route is not found, the request is lost. Due to our assumption that an arriving at node n request immediately becomes aware of the status of all adjacent to node n links, to maximize the probability of finding an available two-link route this request shall probe only transit routes (n,m,k) for which the "first-leg" link (n,m) is available, i.e., $\delta_{nm}=1$ and $x_{nm}\leq C-1$.

III. MEAN-FIELD PERFORMANCE MODEL

Evolution of our network with bypass routes is described by Markov process $y(t)=(y_r(t))$ where components $y_r(t)$ characterize the numbers of flows in progress on all direct and two-link routes r. Astronomically high dimension of the corresponding Kolmogorov system for the probabilities of this Markov process makes solving this system computationally infeasible even for moderate size networks. In this paper we consider a sequence of networks with increasing the number of nodes $N\to\infty$ under assumption that asymptotically the numbers of carried requests on different links becomes jointly statistically independent:

$$\lim_{N\to\infty}P^{(N)}(x_{n_jk_j}=i_j,j=1,..,J)=\prod_{j=0}^{J}p_{i_j} \qquad (2)$$

for any fixed $J=2,3,..$; $i_j=0,1,..,C$, and any set of different existing links: $\{(n_j,k_j):\delta_j=1\}$. Assumption (2) originated in statistical physics and was verified by numerous simulations and extensively used for loss networks in [5]-[9].

An arriving request is rejected if the direct link is unavailable and I attempts to find a two-link bypass route with available "second leg" link fail. Due to assumed independence (2), the request loss probability is $L = L_0^{I+1}$, where $L_0 = \pi + (1-\pi)p_C$. Thus, the loss rate (1) is:

$$L = [\pi + (1-\pi)p_C]^{I+1} \quad (3)$$

Slightly modifying results [9] one can show that under assumption (2) the probability that an existing link has its entire bandwidth in use is described by the corresponding Erlang probability:

$$p_C = \frac{\mu^C}{C!} \Bigg/ \sum_{j=0}^{C} \frac{\mu^j}{j!} \quad (4)$$

where the "effective" arrival rate of requests for link bandwidth is

$$\mu = 2\lambda \frac{1+L_0}{1-L_0}(1-L) \quad (5)$$

Combining (3)-(5) we obtain the following fixed-point equation for probability p_C:

$$p_C = \varphi(p_C) \quad (6)$$

In a case when equation (9) has unique globally stable equilibrium point p_C^*, it is natural to identify this equilibrium and the corresponding loss rate with the stable, i.e., steady-state network state. In a case when equation (6) has multiple locally stable equilibrium points, we interpret these points as representing metastable, i.e., persistent network states. It the "fast" time scale the network converges to the "closest" metastable state, and in the "slow" time scale the network state alternates between the metastable states. This interpretation has been substantiated by numerous simulations and some analytical results [5]-[9]. The next Section analyzes solution of equation (6) as function of system parameters: I, π, and network utilization

$$\rho = \frac{2\lambda}{C}\frac{1+\pi}{1-\pi} \quad (7)$$

IV. DISCUSSION

Figure 1 which sketches solution of equation (6).

Figure 1. Solution of the fixed-point equation (6)

In cases of low and high load $\rho < \rho_1^*$ and $\rho > \rho_2^*$ equation (6) has unique globally stable equilibrium p_C^*. In the intermediate case $\rho_1^* < \rho < \rho_2^*$ equation (6) has two locally stable solutions p_{1C}^* and p_{2C}^*, where $p_{1C}^* < p_{2C}^*$, separated by an unstable solution. Due to limited space we only consider a case of large capacity links $C \to \infty$, when $p_C = \max(0, 1 - C/\mu)$.

Figure 2 compares the persistent loss probabilities for $I \to \infty$ with loss probability $L = \max(0, 1 - 1/\rho)$ for the deterministic system, which can sustain load $\rho \le 1$ without losses and is fully utilized if $\rho > 1$.

Figure 2. Stochastic vs. deterministic model for $C \to \infty$

This comparison makes sense since for a given ρ, assumption $C \to \infty$ eliminates load randomness. Indeed, rate of Poisson request arrivals for given origin-destination $\lambda \to \infty$ and thus the corresponding relative standard deviation of the amount of arriving work $\lambda^{-1/2} \to 0$.

For sufficiently light load $\rho < (1+\pi)/2$ fixed point equation (6) has unique equilibrium $p_C^* = 0$ which describes lossless deterministic system: $L^* = 0$. For intermediate load $(1+\pi)/2 < \rho < 1$ this equilibrium $p_{1C}^* = 0$ remains as locally stable, and thus describes lossless metastable state $L_1^* = 0$. Another locally stable equilibrium $p_{2C}^* = 1$ describes a congested metastable state with finite loss rate $L_1^* = 1 - (1+\pi)/(2\rho)$. For sufficiently heavy load $\rho > 1$ fixed point equation (9) has unique equilibrium $p_C^* = 1$ which describes stable congested mode with finite loss rate $L^* = 1 - (1+\pi)/(2\rho)$.

"Slow" load ρ increase from $\rho = 0$ to $\rho = \infty$ results in loss probability change along path 0ABCD, which includes discontinues, i.e., first kind phase transition at the point $\rho = 1$. "Slow" load ρ decrease from $\rho = \infty$ to $\rho = 0$ results in loss probability change along path DCS0, which includes continues, i.e., phase transition of the second kind at $\rho = (1+\pi)/2$. The hysteresis loop ABCA is indicative of the metastability. The deterministic model, which does not reveal a possibility of metastable states, predicts loss rate change along path 0ABE.

Figure 3 shows system phase diagram in parameter space (ρ, π) as $C \to \infty$.

Figure 3. Phase diagram for $C \to \infty$

In the region of low load – high uncertainty: $(\rho, \pi) \in U_1$ as well as in the region of high load: $(\rho, \pi) \in U_2$ equation (6) has unique globally stable equilibrium. This interpretation of different regions in Figure 3 accounts for uncertainty in both network topology and demand. In the region $(\rho, \pi) \in U_1$ loss rate L is a decreasing function of I, and thus the optimal values $I^{opt} = \infty$, $L^{opt} = 0$ and this equilibrium describes "normal" stable system mode. In the region $(\rho, \pi) \in U_2$ where loss rate L is an increasing function of I, and thus the optimal values $I^{opt} = 0$, $L^{opt} = \pi$ and this equilibrium describes "congested" stable system mode. In the intermediate region $(\rho, \pi) \in U_{12}$ these both equilibria and the corresponding system modes coexist as metastable.

Finally, Figure 4 demonstrates risks/benefits of allowing selfish network management (black curve) vs. static centralized management (blue curve).

Figure 4. Metastable loss vs. information access for $C \to \infty$.

Allowing selfish agents to acquire and use system information while improving system performance in metastable normal regime worsens system performance in metastable congested regime as compared to static centralized management.

V. Conclusion and Future Research

This paper indicates an intriguing trade-off between the positive and negative effects of selfish agent ability to acquire and exploit system information due to a possibility of matastable, i.e., persistent, regimes. While availability of system information to selfish agents improves performance of the "normal" system operating mode, it may also increase systemic risk of cascading overload driving system to the undesirable metastable system mode. Since metastability is a result of non-linear stochastic system dynamics, conventional deterministic resource allocation models typically underestimate or completely miss systemic risks of abrupt system transition to an undesirable metastable state. In particular, our analysis indicates that metastability may make "Braess paradox" more severe than predicted by deterministic models [1].

This paper has not considered a possibility of pricing. It is well known that in a deterministic case congestion pricing under certain convexity conditions incentivizes selfish elastic users to maximize the aggregate system utility rather than the individual utility. Performance implications of selfish agent ability to instantaneously acquire and exploit near real-time pricing information in stochastic resource allocation systems is an important area of future research. We consider a specific case of a large-scale loss network with random topology under mean-field approximation. The main advantage of this particular case study selection is that mean-field approximation for these networks has been verified by numerous simulations. Extending our analysis to other classes of distributed systems, e.g., banking ecosystems [10] and bandwidth allocation in the Internet, are interesting possibilities for future research.

References

[1] T. Roughgarden. "The Price of Anarchy." MIT Press, Cambridge, MA, 2005.

[2] F.A. Hayek, The Road to Serfdom, University of Chicago Press, Chicago 1994.

[3] D. Zhu, M. Gritter, and D.R. Cheriton. Feedback based routing. Computer Communication Review, 33(1):71–76, 2003.

[4] F.P. Kelly and T. Voice, Stability of end-to-end algorithms for joint routing and rate control, Computer Communication Review 35:2 (2005).

[5] R. J. Gibbens, P. J. Hunt, and F. P. Kelly, Bistability in communication networks, Disorder in physical systems, Oxford Sci. Publ., Oxford Univ. Press, New York, 1990, pp. 113–127.

[6] V. Anantharam, A Mean Field Limit for a Lattice Caricature of Dynamic Routing in Circuit Switched Networks, Annals of Applied Probability 1 (1991) 481-503.

[7] Nelson Antunes, Christine Fricker, Philippe Robert, and Danielle Tibi. Stochastic networks with multiple stable points. *Annals of Probability*, 36(1):255-278, 2008.

[8] V. Marbukh "Asymptotic Investigation of a Complete Communication Network with a Large Number of Nodes and Bypass Routes", *Problems of Information Transmission*, Vol. 16, No. 3, (1981), pp. 212-216.

[9] V. Marbukh "Loss Circuit-Switched Communication Networks: Performance Analysis and Dynamic Routing" - invited paper for a special issue of *The Queuing Systems: Theory and Applications*, Vol 13, (1993), pp. 111-141.

[10] A.G. Haldane and R.M. May, "Systemic risk in banking ecosystem," Nature, 469, 351-355 (20 January 2011).