Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Asymptotically Tight Bounds for Inefficiency in Risk-Averse Selfish Routing

Thanasis Lianeas and **Evdokia Nikolova**
University of Texas at Austin
{tlianeas|nikolova}@austin.utexas.edu

Nicolas E. Stier-Moses
Facebook Core Data Science*
nstier@fb.com

Abstract

We consider a nonatomic selfish routing model with independent stochastic travel times for each edge, represented by mean and variance latency functions that depend on arc flows. This model can apply to traffic in the Internet or in a road network. Variability negatively impacts packets or drivers, by introducing jitter in transmission delays which lowers quality of streaming audio or video, or by making it more difficult to predict the arrival time at destination. The *price of risk aversion* (PRA) has been defined as the worst-case ratio of the cost of an equilibrium with risk-averse players who seek risk-minimizing paths, and that of an equilibrium with risk-neutral users who minimize the mean travel time of a path [Nikolova and Stier-Moses, 2015]. This inefficiency metric captures the degradation of system performance caused by variability and risk aversion. In this paper, we provide the first lower bounds on the PRA. First, we show a family of *structural* lower bounds, which grow linearly with the size of the graph and players' risk-aversion. They are tight for graph sizes that are powers of two. We also provide asymptotically tight *functional* bounds that depend on the allowed latency functions but not on the topology. The functional bounds match the price-of-anarchy bounds for congestion games multiplied by an extra factor that accounts for risk aversion. Finally, we provide a closed-form formula of the PRA for a family of graphs that generalize series-parallel graphs and the Braess graph. This formula also applies to the mean-standard deviation user objective—a much more complex model of risk-aversion due to the cost of a path being non-additive over edge costs.

1 Introduction

A primary objective in many networked systems subject to congestion such as the Internet and transportation networks is minimizing latencies. A secondary objective in such systems, which has received considerably less attention than the first, is reducing variability. In telecommunication networks, variability amounts to jitter, which causes severe degradation in the quality of audio and video streams. In transportation networks, variability reduces the likelihood of correctly estimating the time of arrival. Consequently, this miscalculation can make the travelers late, which may have very bad consequences, or early which is less dramatic but sometimes inconvenient. This is of paramount importance for many routing services in online platforms such as Waze or Uber.

In both settings, a reasonable goal from the perspective of the agent (packet or traveler) is to consider a multi-objective shortest path problem with respect to both expected latency and its variability. A path with slightly larger expected latency and no variability may be preferred to one with a highly-variable lower-expectation latency, even if the latter stochastically dominates the former. As commonly done in settings with multiple objectives, we consider their linear combination, in this case given by the mean-variance latency of a path choice. The coefficient that is used to sum the two terms into one objective function typically represents agents' risk aversion, and is readily interpreted as how much agents can allow the expected latency to increase to reduce the variance. Since the networks we study are subject to congestion effects, we need to embed these agents' subproblems into a game setting in which each agent has to minimize that objective. This game has the structure as a routing game and its equilibrium concept is an extension of Wardrop equilibria [Wardrop, 1952; Beckmann *et al.*, 1956]. For surveys that discuss the role of these games in the context of telecommunication and transportation networks, see, respectively, Altman *et al.* [2006] and Correa and Stier-Moses [2011].

Related Work There has been an increased effort in recent years to model risk-averse preferences in routing games and understand the effect of such player preferences on network equilibria [Ordóñez and Stier-Moses, 2010; Nikolova and Stier-Moses, 2011; Nie, 2011; Angelidakis *et al.*, 2013; Piliouras *et al.*, 2013; Nikolova and Stier-Moses, 2014; 2015; Cominetti and Torrico, 2013; Fotakis *et al.*, 2015]. We refer the reader to a recent survey [Cominetti, 2015] for a more extensive review of equilibrium routing under uncertainty.

In an effort to decouple the effect of risk attitudes from the effect of selfish behavior on the degradation of system perfor-

* The majority of this work was done before the author joined Facebook.

mance, the *price of risk aversion (PRA)* has been defined as the worst-case ratio of the cost of a risk-averse equilibrium to that of a risk-neutral equilibrium (namely, the equilibria when agents are risk-averse and risk-neutral, respectively) and has been shown to be upper-bounded by $1 + \kappa\gamma n/2$, where κ is the worst-case variance-to-mean ratio of an edge at equilibrium, γ is the coefficient of risk-aversion and n is the number of vertices in the graph [Nikolova and Stier-Moses, 2015].

Meir and Parkes [2015] defined biased smoothness and provide a technique that can be used to compare an equilibrium under modified cost functions to the social optimum of the original game. As an example of this technique, they derive an upper bound on the price of risk aversion of $(1 + \kappa\gamma)(1 - \mu)^{-1}$ when cost functions are $(1, \mu)$-smooth. As we establish in this paper, this upper bound and that of Nikolova and Stier-Moses [2015] are of a different type, i.e., functional (based on the latency function classes) vs. structural (based on the network topology), which is why they cannot be compared directly. In our paper, we present a new lower bound of this functional form, as well as a simpler proof of the functional upper bound that relies on a generalization of the earlier price of anarchy proof based on variational inequalities [Correa *et al.*, 2008]. Consequently, this method makes possible a more direct comparison with the existing literature including the traditional price of anarchy proofs.

Our contributions Related to AI research on autonomous agents, our results shed light in how risk aversion experienced by agents impact system-wide metrics. The main technical contributions are the first lower bounds to the price of risk-aversion, which are tight or asymptotically tight. We also provide a simplified analysis of the upper bound of Meir and Parkes [2015] for the price of risk-aversion, as well as new tight bounds for a class of graphs that apply to both mean-variance and mean-stdev risk-averse objectives. Our main conceptual contribution is a new *structural* vs. *functional* perspective on the price or risk-aversion, demonstrating that bounds can be given either in terms of the network structure or in terms of the class of edge latency functions that are allowed. As such, we also provide an understanding of the connection of two prior upper-bound results on the price of risk-aversion [Nikolova and Stier-Moses, 2015; Meir and Parkes, 2015], as well as a connection to the extensive literature on price of anarchy, which has so far provided only functional bounds.

Our first result is a tight lower bound to the price of risk aversion of $1 + \kappa\gamma n/2$, with number of vertices n that are powers of two (Theorem 3), where κ is the maximum variance-to-mean ratio, and γ is the degree of risk-aversion of users. This bound essentially closes the gap to the previously known upper bound and characterizes the exact price of risk aversion for an infinite number of graph sizes. We call the bound *structural*, since it matches the upper bound that only depends on the network structure and not on the latency functions used. Our construction of the worst-case graph family involves finding an instance in which an alternating path (a path that consists of appropriately defined forward and backward edges) goes through every vertex in the graph and alternates between forward and backward edges at every internal vertex. We achieve this by inductively defining a graph family with appropriate mean and variance functions for each edge.

Our second lower bound is an asymptotically tight functional bound (Theorem 5). It follows from the same inductively defined family of graphs but with different appropriately chosen mean and variance functions. We also give a new proof of the *functional* upper bound $(1 + \kappa\gamma)(1 - \mu)^{-1}$ for $(1, \mu)$-smooth latency functions via a variational inequality characterization of equilibria (Theorem 4). This bound implies, for example, that the price of risk aversion is at most $(1+\kappa\gamma)4/3$ for linear latency functions and it generalizes the classic result that the price of anarchy is $(1 - \mu)^{-1}$ [Roughgarden, 2003] because, when there is no variability, $\kappa = 0$. We remark that for unrestricted functions, the functional upper bounds become vacuous since $\mu = 1$, which provides further support for the structural analysis of Section 3.

Finally, we provide a closed-form formula for the price of risk aversion under the alternative mean-stdev model of risk-aversion where players minimize a linear combination of the mean and standard deviation of the route travel time. This model is significantly more difficult to analyze than the mean-var model due to its non-additive nature, namely the mean plus standard deviation of a path cannot be decomposed as a sum of costs over the edges in the path. The formula is valid for a family of graphs that generalizes series-parallel graphs (specifically, the family of graphs where the domino-with-ears graph, shown in Fig. 3(b), is a forbidden minor). This formula extends to the mean-var model as well, and it refines our understanding on the topology of graphs for which the price of risk aversion is $1 + 2\kappa\gamma$, as opposed to the cruder bound in terms of number of vertices only.

Both our structural and functional lower bounds also readily provide corresponding lower bounds in the mean-stdev model. Showing a structural and functional upper bound for general graphs in that model remains an open problem.

2 Model and Preliminaries

We consider a directed graph $G = (V, E)$ with a single source-sink pair (s, t) and an aggregate demand of d units of flow that need to be routed from s to t. We let $\mathcal{P}$ be the set of all feasible paths between s and t. We encode the players' decisions as a flow vector $f = (f_p)_{p\in\mathcal{P}} \in \mathbb{R}_+^{|\mathcal{P}|}$ over all paths. Such a flow is feasible when demand is satisfied, as given by the constraint $\sum_{p\in\mathcal{P}} f_p = d$. For notational simplicity, we denote the flow on an edge e by $f_e = \sum_{p\ni e} f_p$. When we need multiple flow variables, we use x, x_p, x_e and z, z_p, z_e.

The network is subject to congestion, modeled with stochastic delay functions $\ell_e(f_e) + \xi_e(f_e)$ for each edge $e \in E$. Here, the deterministic function $\ell_e(f_e)$ measures the expected delay when the edge has flow f_e, and $\xi_e(f_e)$ is a random variable that represents a noise term on the delay. Functions $\ell_e(\cdot)$, generally referred to as *latency functions*, are assumed continuous and non-decreasing. The expected latency along a path p is given by $\ell_p(f) := \sum_{e\in p} \ell_e(f_e)$. Random variables $\xi_e(f_e)$ are pairwise independent, have expectation zero and standard deviation $\sigma_e(f_e)$, for arbitrary continuous functions $\sigma_e(\cdot)$. For the variational inequality characterization used in Section 4, we further assume that mean-variance objective of users, defined below, is non-decreasing. The variance along a path equals $v_p(f) = \sum_{e\in p} \sigma_e^2(f_e)$, and the standard deviation (stdev) is $\sigma_p(f) = (v_p(f))^{1/2}$.

We consider the *nonatomic* version of the routing game where infinitely many players control an infinitesimal amount of flow each so that the path choice of a single player does not unilaterally affect the costs experienced by other players.

Players are risk-averse and strategically choose paths taking into account the variability of delays by considering a *mean-var* objective $Q_p^\gamma(f) = \ell_p(f) + \gamma v_p(f)$. We refer to this objective simply as the *path cost* (as opposed to latency). Here, $\gamma \geq 0$ is a constant that quantifies the risk-aversion of the players, which we assume homogeneous. The special case of $\gamma = 0$ corresponds to risk-neutrality. In the last section, we consider the *mean-stdev* objective where the variance in the objective is replaced with the standard deviation of the path.

The variability of delays is usually not too large with respect to the expected latency. Following previous work, we assume that $v_e(x_e)/\ell_e(x_e)$ is bounded from above by a fixed constant κ for all $e \in E$ at the equilibrium flow of interest $x_e \in \mathbb{R}_+$. This means that the variance cannot be larger than κ times the expected latency in any edge at equilibrium.

In summary, an instance of the problem is given by the tuple (G, d, ℓ, v, γ), which represents the topology, demand, latency functions, variability functions, and degree of player risk-aversion. The following definition captures that at equilibrium players route flow along paths with minimum cost.

Definition 1 (Equilibrium)**.** *A* γ-equilibrium *of a stochastic nonatomic routing game is a flow f such that for every path $p \in \mathcal{P}$ with positive flow, the path cost $Q_p^\gamma(f) \leq Q_q^\gamma(f)$ for any other path $q \in \mathcal{P}$. For a fixed risk-aversion parameter γ, we refer to a γ-equilibrium as a* risk-averse Wardrop equilibrium *(RAWE), denoted by x. For $\gamma = 0$, we call the equilibrium a risk-neutral Wardrop equilibrium (RNWE).*

Notice that since the variance decomposes as a sum over all the edges on the path, the previous definition represents a standard Wardrop equilibrium with respect to modified costs $\ell_e(f_e) + \gamma v_e(f_e)$. For the existence of the equilibrium, it is sufficient that the modified cost functions are increasing.

Our goal is to investigate the effect that risk-averse players have on the quality of equilibria. The quality of a solution that represents collective decisions can be quantified by the cost of equilibria with respect to expected delays since, over time, different realizations of delays average out to the mean by the law of large numbers. For this reason, a social planner, who is concerned about the long term, is typically risk neutral. Furthermore, the social planner may aim to reduce long-term emissions, which would be better captured by the total expected delay of all users.

Definition 2. *The* social cost *of a flow f is defined as the sum of the expected latencies of all players: $C(f) := \sum_{p\in\mathcal{P}} f_p \ell_p(f) = \sum_{e\in E} f_e \ell_e(f_e)$.*

Although one could have measured total cost as the weighted sum of the costs $Q_p^\gamma(f)$ of all users, this captures users' utilities but not the system's benefit. Such a cost function was previously considered to compute the price of anarchy [Nikolova and Stier-Moses, 2014]. In contrast, our goal is to compare across different values of risk aversion so we want the various flow costs to be measured with the same units.

The next definition captures the increase in social cost at equilibrium introduced by user risk-aversion, compared to the social cost if users were risk-neutral.

Definition 3 ([Nikolova and Stier-Moses, 2015])**.** *Considering an instance family $\mathcal{F}$ of a routing game with uncertain delays, the* price of risk aversion *(PRA) associated with γ and κ (the risk-aversion coefficient and the variance-to-mean ratio) is defined by PRA$(\mathcal{F}, \gamma, \kappa) := \sup_{G,d,\ell,v} \Big\{ \frac{C(x)}{C(z)} : (G, d, \ell, v, \gamma) \in \mathcal{F}$ and $v(x) \leq \kappa \ell(x) \Big\}$, where x and z are the risk-averse and the risk-neutral Wardrop equilibria of the corresponding instance.*

3 Structural Lower Bounds

In this section, we prove two lower bounds on the price of risk aversion, both matching the upper bounds presented by Nikolova and Stier-Moses [2015]. The first bound for PRA is with respect to the minimum number of alternations among all alternating paths as defined below, while the second bound is with respect to the number of vertices in the graph. In fact, the same bounds hold in the mean-standard deviation model, but we defer that discussion to Section 5.

Given an instance, we denote a RNWE flow associated with it by z, and a RAWE flow by x. To define alternating paths, we partition the edge-set E into $A = \{e \in E : x_e \leq z_e\}$ and $B = \{e \in E : z_e < x_e\}$. Conceptually, an alternating path is an s-t-path in the graph where edges in B are reversed (see Fig. 3(c) for an illustration of the definition).

Definition 4 ([Nikolova and Stier-Moses, 2015])**.** *A generalized path $\pi = A_1$-B_1-A_2-B_2-$\cdots$-A_t-B_t-A_{t+1}, composed of a sequence of subpaths, is an* alternating path *when every edge in $A_i \subseteq A$ is directed in the direction of the path, and every edge in $B_i \subseteq B$ is directed in the opposite direction from the path. We say that such a path has $t + 1$ disjoint forward subpaths, and t* alternations.

The definition of alternating paths was motivated by the following result.

Theorem 1 ([Nikolova and Stier-Moses, 2015])**.** *Considering the set of instances with arbitrary mean and variance latency functions that admit an alternating path with up to η disjoint forward subpaths, PRA $\leq 1 + \eta\gamma\kappa$.*

The theorem implies that for the set of instances on graphs with n vertices, PRA $\leq 1 + \gamma\kappa\lceil \frac{n-1}{2} \rceil$ since in that case an alternating path cannot have more than $\lceil (n-1)/2 \rceil$ disjoint forward subpaths. We are going to prove that those upper bounds are tight. To get there, we first prove a more general result that shows how instances with high price of risk aversion can be constructed.

Theorem 2. *For an $i \in \mathbb{N}_{>0}$, consider $r_A^i, r_N^i \in \mathbb{R}_{\geq 0}$ such that $2^i r_A^i > (2^i - 1) r_N^i$. There exists an instance based on a graph $G^i(r_A^i, r_N^i)$ that satisfies the following two properties.*

- *If r_A^i risk-averse players are routed through $G^i(r_A^i, r_N^i)$, then the path cost along used paths at the RAWE flow x,*

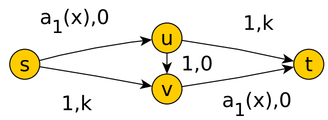

Figure 1: The base case $G^1(r_A^1, r_N^1)$ is a Braess graph.

as well as the expected latency, is $1 + 2^i\gamma\kappa$. The social cost is $C(x) = (1 + 2^i\gamma\kappa)r_A^i$.

- *If r_N^i risk-neutral players are routed through $G^i(r_A^i, r_N^i)$, then the expected latency along each used path at the RNWE flow z is 1. The social cost is $C(z) = r_N^i$.*

The proof is by induction on i. We will recursively construct the instance for i by forming a Braess instance with the graph resulting for the $i-1$ case. At each step we will need to find a mean latency function that makes the properties in the statement work.

Proof. For the base case $i = 1$, we let $G^1(r_A^1, r_N^1)$ be the Braess graph, shown in Fig. 1. Indeed, consider any r_A^1, r_N^1 such that $r_A^1 > r_N^1/2$ as indicated in the statement of the result. We define the mean latency function $a_1(x)$ to be any function that is strictly increasing for $x \geq r_N^1/2$ and such that $a_1(r_N^1/2) = 0$ and $a_1(r_A^1) = \gamma\kappa$. Note that in order for $a_1(x)$ to be strictly increasing, it is necessary that $r_A^1 > r_N^1/2$, which holds by hypothesis.

The RAWE flow x routes the r_A^1 risk-averse players along the zig-zag path. Hence, the mean-var objective in the upper-left and the lower-right edges, as well as the mean latency, will each be $\gamma\kappa$, totalling $1 + 2\gamma\kappa$ for each player in both cases. Hence, $C(x) = (1 + 2^1\gamma\kappa)r_A^1$. Instead, the RNWE flow z routes the r_N^1 risk-neutral players along the top and bottom paths, half and half. Hence, the cost for each player is 1 and $C(z) = r_N^1$, proving the base case.

Let us consider the inductive step where we assume we have an instance satisfying the properties for $i-1$ and construct the instance for step i. Starting from r_A^i and r_N^i satisfying the condition in the statement for case i, we set $r_A^{i-1} = (2^i r_A^i - r_N^i)/2^{i+1}$ and $r_N^{i-1} = r_N^i/2$. We first verify that these values satisfy the hypothesis for the case $i-1$. Indeed, $r_A^{i-1} > \frac{2^{i-1}-1}{2^{i-1}} r_N^{i-1}$ because by hypothesis $r_A^i > \frac{2^i-1}{2^i} r_N^i \Leftrightarrow \frac{r_A^i}{2} > \frac{2^i-1}{2^{i+1}} r_N^i$ which implies that $\frac{r_A^i}{2} - \frac{r_N^i}{2^{i+1}} > \frac{2^i-1}{2^{i+1}} r_N^i - \frac{1}{2^{i+1}} r_N^i = \frac{2^{i-1}-1}{2^{i-1}} \frac{r_N^i}{2}$.

Using the graph corresponding to step $i-1$ and the values of r_A^{i-1} and r_N^{i-1} specified previously, we construct graph $G^i(r_A^i, r_N^i)$ with those components as shown in Fig. 2. We define the mean latency function $a_i(x)$ to be any function that is strictly increasing for $x \geq r_N^i/2$ and such that $a_1(r_N^i/2) = 0$ and $a_i(\frac{r_A^i}{2} + \frac{r_N^i}{2^{i+1}}) = 2^{i-1}\gamma\kappa$. Note that in order for $a_i(x)$ to be strictly increasing, it is necessary that $\frac{r_A^i}{2} + \frac{r_N^i}{2^{i+1}} > \frac{r_N^i}{2}$, which actually holds because, by hypothesis, $r_A^i > \frac{2^i-1}{2^i} r_N^i \Leftrightarrow \frac{r_A^i}{2} > \frac{2^i-1}{2^{i+1}} r_N^i$.

The RAWE flow x routes the r_A^i risk-averse players as follows: $\frac{r_A^i}{2} - \frac{r_N^i}{2^{i+1}}$ units along the upper path, $\frac{r_N^i}{2^i}$ units along

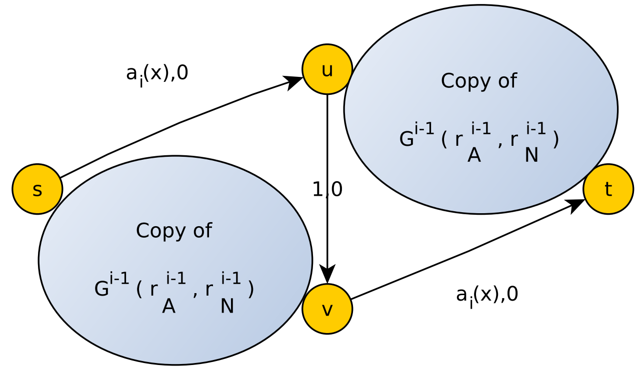

Figure 2: The recursive construction of $G^i(r_A^i, r_N^i)$ by forming a Braess graph topology using components of the earlier step.

the zig-zag path, and $\frac{r_A^i}{2} - \frac{r_N^i}{2^{i+1}}$ units along the lower path. The mean-var objective of the upper-left and the lower-right edges, as well as the mean latency, will each be $2^{i-1}\gamma\kappa$ since the flow through them is equal to $\frac{r_A^i}{2} + \frac{r_N^i}{2^{i+1}}$. The flow inside each of the copies of $G^{i-1}(r_A^{i-1}, r_N^{i-1})$ is a RAWE for which we know, by induction, that all players perceive a path cost of $1 + 2^{i-1}\gamma\kappa$, which additionally, by induction, is the mean latency of all used paths. Thus, the path cost that players perceive in $G^i(r_A^i, r_N^i)$ under the RAWE flow x is $1 + 2^i\gamma\kappa$, which additionally is the mean latency of all used paths, and the social cost is $C(x) = (1 + 2^i\gamma\kappa)r_A^i$.

The RNWE flow z routes the r_N^i risk-neutral players along the top and bottom paths, half and half. Hence, the path cost for perceived by each player is 1, as the mean-var objective in the upper-left and lower-right edges is equal to 0, and, by induction, passing through either of both copies of $G^{i-1}(r_A^{i-1}, r_N^{i-1})$ has a mean-var objective of 1. This implies that $C(z) = r_N^i$, which completes the proof. $\square$

The previous result provides a constructive way to generate instances with high price of risk aversion. Notice that the paths of the instances resulting from these constructions have at most one edge with non-zero variance. This fact is useful to extend our lower bounds to the mean-stdev model, since in that case summing and taking square roots is not needed.

Another useful observation is that the prevailing value for mean latency functions a_j under the RNWE flow z is 0, and under the RAWE flow x is 2^{j-1}. This can be easily proved by induction and will be used when establishing functional lower bounds on the PRA in the next section.

We now use the previous result to get lower bounds for PRA matching the upper bound specified earlier.

Corollary 1. *For any $n_0 \in \mathbb{N}$, there is an instance on a graph G with $n \geq n_0$ vertices such that its equilibria satisfy $C(x) \geq (1 + \gamma\kappa\lceil(n-1)/2\rceil)C(z)$.*

Proof. Consider an arbitrary demand d, and apply Theorem 2 with $r_A^i = r_N^i = d$ and $i = \min\{j \in \mathbb{N} : n_0 \leq 2^j\}$ to get instance $G^i(r_A^i, r_N^i)$. Consequently, x and z satisfy: $\frac{C(x)}{C(z)} =$

$\frac{(1+2^i\gamma\kappa)d}{d} = 1+\gamma\kappa\frac{n}{2}$, because $G^i(r_A^i, r_N^i)$ has 2^{i+1} vertices by construction. The result holds, as n is a power of two. □

The previous lower bound together with the upper bound given in the paragraph after Theorem 1 imply that the PRA with respect to the set of instances on graphs with up to n vertices is *exactly* equal to $1 + \gamma\kappa\lceil(n-1)/2\rceil$ when n is a power of 2. From there, the bound is tight infinitely often. Although for other values of n the bounds are not tight, they are close together so these results provide an understanding of the asymptotic growth of the PRA. We now refine this observation to the bound in Theorem 1.

Theorem 3. *The upper bound for the price of risk-aversion shown in Theorem 1 and the lower bound shown in Corollary 1 coincide for graphs of size that is a power of 2. Otherwise, the gap between them is less than 2.*

In conclusion, PRA $= 1 + \eta\gamma\kappa$ when the family of instances is defined as graphs with arbitrary mean and variance functions that admit alternating paths with up to η disjoint forward subpaths, for η equal to a power of 2. We have equality because the supremum in the definition of PRA is attained by the instance constructed previously.

4 Functional Bounds

In this section, we turn our attention to instances with mean latency functions restricted to be in a certain family (as, e.g., affine functions). We prove upper and lower bounds for the PRA that are asymptotically tight as $\gamma\kappa$ increases. The results rely on the variational inequality approach that was first used by Correa *et al.* [2008] to prove price of anarchy (POA) bounds for fixed families of functions. This approach was based on the properties of the allowed functions. Since then, these properties have been successively refined [Harks, 2011; Roughgarden and Schoppmann, 2015], and they are now usually referred to as the *local smoothness* property. Although not really needed for the results here, we use the latter terminology since it has become standard by now. To characterize a family of mean latency functions, we rely on the smoothness property, defined below.

Definition 5 ([Roughgarden and Schoppmann, 2015])**.** *A function $\ell : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be* (λ, μ)-smooth around $x \in \mathbb{R}_{\geq 0}$ *if $y\ell(x) \leq \lambda y\ell(y) + \mu x\ell(x)$ for all $y \in \mathbb{R}_{\geq 0}$.*

Using the previous definition, we construct an upper bound for the PRA when mean latency functions $\{\ell_e\}_{e\in E}$ are $(1,\mu)$-smooth around the RAWE flow x_e for all edges $e \in E$. Meir and Parkes [2015] proved a similar bound using a related approach in which they generalize the smoothness definition to *biased smoothness* which holds with respect to a modified latency function. In our case, the modified latency function would be $\ell_e + \gamma v_e$. One advantage of our approach is its simplicity; it is a straightforward generalization of the POA proof given in Correa *et al.* [2008]. We provide a proof corresponding to our assumptions, matching what is needed to get our asymptotically-tight lower bounds.

Theorem 4. *Consider the set of general instances with mean latency functions $\{\ell_e\}_{e\in E}$ that are $(1,\mu)$-smooth around any RAWE flow x_e for all $e \in E$. Then, with respect to that set of instances, PRA $\leq (1+\gamma\kappa)\frac{1}{1-\mu}$.*

Proof. We consider an instance within the family, a corresponding RAWE flow x, and a RNWE flow z. Further, we let $A = \{e \in E | x_e \leq z_e\}$ and $B = \{e \in E | z_e < x_e\}$. Using a variational inequality formulation for the RAWE, we get $\sum_{e\in E} x_e(\ell_e(x_e) + \gamma v_e(x_e)) \leq \sum_{e\in E} z_e(\ell_e(x_e) + \gamma v_e(x_e))$. Partitioning the sum over E at both sides into terms for A and B, subtracting from it the inequality $\sum_{e\in A} x_e v_e(x_e) + \sum_{e\in B} x_e v_e(x_e) \geq \sum_{e\in B} z_e v_e(x_e)$, and further bounding $v_e(x_e)$ by $\kappa\ell_e(x_e)$, we get that $C(x) = \sum_{e\in A} x_e\ell_e(x_e) + \sum_{e\in B} x_e\ell_e(x_e) \leq \sum_{e\in A}(1+\gamma\kappa)z_e\ell_e(x_e) + \sum_{e\in B} z_e\ell_e(x_e)$. Applying the definition of A to the first term and the $(1,\mu)$-smoothness condition to the second, we get: $C(x) \leq \sum_{e\in A}(1+\gamma\kappa)z_e\ell_e(z_e) + \sum_{e\in B}(z_e\ell_e(z_e) + \mu x_e\ell_e(x_e)) \leq (1+\gamma\kappa)C(z) + \mu C(x)$. □

The bound in the previous result is similar to that for the POA for nonatomic games with no uncertainty. Indeed, the result there is that POA $\leq (1-\mu)^{-1}$, the same without the $1+\gamma\kappa$ factor. The values of $(1-\mu)^{-1}$ have been computed for different families of functions in previous work. On the other hand, for unrestricted functions this value is infinite so the bound becomes vacuous in that case, which provides support for the structural analysis of Section 3.

To evaluate the tightness of our upper bounds, we now propose lower bounds for the PRA. More specifically, we provide a family of instances indexed by i whose latency functions are $(1,\mu_i)$-smooth, for $\mu_i = 1 - 2^{-i}$, for which the bound is approximately tight. These instances imply lower bounds equal to $1 + \gamma\kappa(1-\mu_i)^{-1} = 1 + \gamma\kappa 2^i$. First, notice that although the lower and upper bounds do not match, they are similar. The difference is whether the 1 is or is not multiplied by the μ factor. When the $\gamma\kappa$ term is large, both bounds are essentially equal. Second, notice that for large values of i, necessarily the number of alternations of the longest alternating path must grow exponentially large to simultaneously match the structural upper bound presented in Theorem 1.

Theorem 5. *For any $i > 0$, letting $\mu_i = 1 - 2^{-i}$, PRA $\geq 1 + \gamma\kappa(1-\mu_i)^{-1}$ for the family of instances satisfying the $(1,\mu_i)$-smoothness property.*

Proof sketch. To get the result we use the recursively constructed graph G^i of Theorem 2 for $r_A^i = r_N^i = 1$, but with the non-constant cost functions satisfying the $(1,\mu_i)$-smoothness condition around the RAWE: for $1 \leq j \leq i$ we let $a_j(x) = \max\left\{0, \frac{2^{j-1}\gamma\kappa}{\frac{2^j-1}{2^i-1} - \frac{2^{j-1}}{2^i}}\left(x - \frac{2^{j-1}}{2^i}\right)\right\}$.

In G^i there are 2^i parallel paths, i.e. paths not containing any vertical edge, and $2^i - 1$ zig-zag paths, i.e. the rest of the paths. The RNWE flow z splits the unit flow equally along the 2^i parallel paths and each a_j gets $2^{j-1}/2^i$ units of flow and has 0 cost, exactly as in Theorem 2, implying $C(z) = 1$. The RAWE flow x splits the unit flow equally along the $2^i - 1$ zig-zag paths and each a_j gets $2^{j-1}/(2^i-1)$ units of flow and has cost equal to $2^{j-1}\gamma\kappa$, exactly as in Theorem 2, implying $C(x) = 1 + 2^i\gamma\kappa$, and thus the bound for PRA follows.

To conclude, some technical work is needed for proving that the a_j's are $(1,\mu_i)$-smooth around x_e for all e. The other cost functions are constants, thus $(1,\mu_i)$-smooth. □

5 The Mean-Stdev Model

In this section, we turn to the mean-standard deviation model and prove upper and lower structural bounds on PRA, now assuming that κ is the maximum coefficient of variation $CV_e(f_e)$ among edges, where CV is defined as the ratio between the standard deviation (stdev) and the mean. The lower bounds follow from the instances used to prove the lower bounds in the mean-variance case in Section 3. Considering families of graphs with up to τ forward disjoint subpaths and general mean latency and stdev functions, we prove upper bounds for the cases $\tau = \{1, 2\}$, and lower bounds for arbitrary τ. For $\tau \leq 2$, both bounds coincide, so the PRA for the stdev case gets characterized exactly.

Since we are now dealing with standard deviations, we redefine $Q^\gamma_p(f) = \ell_p(f) + \gamma\sigma_p(f)$, where $\sigma_p(f) = (\sum_{e\in p} \sigma^2_e(f_e))^{1/2}$. Given an instance based on a graph G, we refer to a RNWE flow by z and to a RAWE flow by y. Further, we denote the path cost perceived by players at the RAWE y by $Q^\gamma(y)$ and the expected latencies perceived by players at the RNWE z by $Q^0(z)$. The definition for the price of risk aversion is analogous to that given in Section 2.

Our first result provides inequalities that relate the social cost at equilibrium with the perceived utilities. The first part is known and the second is a generalization.

Proposition 1. *For an arbitrary instance, $C(z) = d \cdot Q^0(z)$ and $C(y) \leq d \cdot Q^\gamma(y)$, where d is the traffic demand.*

As in Section 3, we partition the edge-set E into two: $A = \{e \in E : y_e \leq z_e\}$ and $B = \{e \in E : z_e < y_e\}$. We assume that all edges in E are used by either flow y or z. This is without loss of generality because unused edges can be deleted without any consequence. The definition of B implies that $y_e > 0$ for all $e \in B$, while the assumption implies that $z_e > 0$ for all $e \in A$. To prove an upper bound on PRA, we rely again on alternating paths, the existence of which is guaranteed by Lemma 4.5 of Nikolova and Stier-Moses [2015]. The next result bounds the PRA for graphs that admit simple alternating paths (i.e., actual s-t paths), or alternating paths with a single alternation.

Lemma 1. *For $\tau \in \{1, 2\}$, considering the set of instances on general topologies with arbitrary mean latency and stdev functions that admit alternating paths with $\tau - 1$ alternations, PRA $\leq 1 + \tau\gamma\kappa$.*

Proof sketch. For $\tau = 1$, we simply use the equilibrium conditions together with the monotonicity of the cost functions and the stdev-to-mean ratio bound. For $\tau = 2$, we let $\pi = A_1$-B_1-A_2 be the alternating path. Figure 3 (c) for $\eta = 1$ illustrates the topology of these subpaths.

Consider a RAWE flow y and a flow-carrying path C_1-B_1-D_1 under y. Using the equilibrium conditions for y, by separately examining on whether $\sigma^2_{C_1}(y) + \sigma^2_{B_1}(y) + \sigma^2_{D_1}(y) \leq \sigma^2_{A_1}(y) + \sigma^2_{D_1}(y)$, and by using some properties of the square root we manage to derive in both cases that $\ell_{C_1}(y) + \ell_{B_1}(y) + \gamma\sqrt{\sigma^2_{C_1}(y) + \sigma^2_{B_1}(y)} \leq (1 + \gamma\kappa)\ell_{A_1}(y)$. Using a similar argument for the path C_1-A_2 instead of A_1-D_1, we further derive that $\ell_{D_1}(y) + \gamma\sigma_{D_1}(y) \leq (1+\gamma\kappa)\ell_{A_2}(y) - \ell_{B_1}(y)$. Since path C_1-B_1-D_1 is used under y, $Q^\gamma(y) = \ell_{C_1}(y) + \ell_{B_1}(y) + \ell_{D_1}(y) + \gamma\sqrt{\sigma^2_{C_1}(y) + \sigma^2_{B_1}(y) + \sigma^2_{D_1}(y)}$. Using the inequalities above, $Q^\gamma(y) \leq (1 + \gamma\kappa)(\ell_{A_1}(z) + \ell_{A_2}(z)) - \ell_{B_1}(z)$.

For the RNWE flow z using the equilibrium conditions we can derive $Q^0(z) \geq \ell_{A_1}(z) + \ell_{A_2}(z) - \ell_{B_1}(z)$. Combining it with the inequality above it, the fact that $Q^0(z)$ is an upper bound for both $\ell_{A_1}(z)$ and $\ell_{A_2}(z)$, and the definition of sets A and B we get $Q^\gamma(y) \leq (1 + 2\gamma\kappa)Q^0(z)$. The result follows since by Proposition 1: $C(y) \leq d \cdot Q^\gamma(y) \leq d \cdot (1 + 2\gamma\kappa)Q^0(z) = (1 + 2\gamma\kappa)C(z)$. ☐

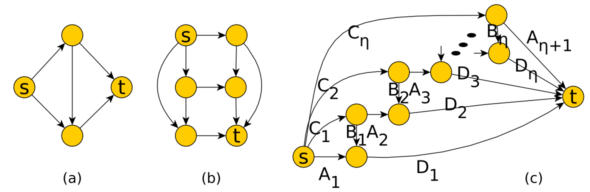

Figure 3: (a) The Braess graph. (b) The domino-with-ears graph. (c) A subgraph of a graph admitting an alternating path with η alternations. Since the edges of the alternating path receive flow, there must exist paths C_i's, and D_i's that bring flow from the source to the edges in A_i's and B_i's and take flow from them and deliver it to the sink, respectively.

Lemma 1 for $\tau = 1$ implies that series-parallel graphs satisfy that PRA $\leq 1 + \gamma\kappa$ and it is well known that those graphs can be characterized as not containing the Braess graph. Along that line, we provide a forbidden minor characterization of Lemma 1 for $\tau = 2$: the topology of instances for which PRA $\leq 1 + 2\gamma\kappa$ is characterized by graphs not containing the domino-with-ears graph as a minor (Fig 3(b)).

Theorem 6. *Instances on general topologies with arbitrary mean latency and stdev functions that do not have domino-with-ears as a minor satisfy PRA $\leq 1 + 2\gamma\kappa$.*

The proof is by contradiction: Assume that the statement is false. By Lemma 1, this implies that any alternating path has at least $\tau \geq 2$ alternations and thus the graph of Fig. 3(c) for $\eta = \tau$ lies inside the network. But this is a contradiction as this graph admits the domino-with-ears as a minor.

The previous proof implies that any graph that has the domino-with-ears graph as a minor cannot belong to the family of instances for which PRA $\leq 1 + 2\gamma\kappa$. Lemma 1 and Theorem 6 together imply that PRA $\leq 1 + \tau\gamma\kappa$ holds for the set of instances admitting alternating paths in which the number of disjoint forward subpaths is not more than τ, for $\tau \leq 2$. Although this does not fully generalize Theorem 1 to the standard deviations case, it is a first step in that direction.

To provide matching lower bounds for the results of this section, we note that all paths in the proof of Theorem 2 have at most one edge with nonzero variance. Thus, all the lower bounds in Section 3 work by reinterpreting the variances as standard deviations and the variance-to-mean ratios as coefficients of variation. This is summarized below.

Corollary 2. *The upper bounds PRA $\leq 1 + \gamma\kappa$ and PRA $\leq 1 + 2\gamma\kappa$ corresponding to graphs that admit alternating paths with 1 and 2 disjoint forward subpaths, respectively, are tight.*

References

[Altman *et al.*, 2006] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter. A survey on networking games in telecommunications. *Computers and Operations Research*, 33(2):286–311, 2006.

[Angelidakis *et al.*, 2013] H. Angelidakis, D. Fotakis, and T. Lianeas. Stochastic congestion games with risk-averse players. In B. Vöcking, editor, *Proceedings of the 6th International Symposium on Algorithmic Game Theory* (SAGT). Volume 8146 of *Lecture Notes in Computer Science*, pages 86–97. Springer, 2013.

[Beckmann *et al.*, 1956] M. J. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT, 1956.

[Cominetti and Torrico, 2013] R. Cominetti and A. Torrico. Additive consistency of risk measures and its application to risk-averse routing in networks. Technical report, Universidad de Chile, 2013.

[Cominetti, 2015] R. Cominetti. Equilibrium routing under uncertainty. *Mathematical Programming B*, 151:117–151, 2015.

[Correa and Stier-Moses, 2011] J. R. Correa and N. E. Stier-Moses. Wardrop equilibria. In J. J. Cochran, editor, *Encyclopedia of Operations Research and Management Science*. Wiley, 2011.

[Correa *et al.*, 2008] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. A geometric approach to the price of anarchy in nonatomic congestion games. *Games and Economic Behavior*, 64:457–469, 2008.

[Fotakis *et al.*, 2015] D. Fotakis, D. Kalimeris, and T. Lianeas. Improving selfish routing for risk-averse players. In E. Markakis and G. Schäfer, editors, *Proceedings of the 11th International Conference on the Web and Internet Economics* (WINE). Volume 9470 of *Lecture Notes in Computer Science*, pages 328–342. Springer, 2015.

[Harks, 2011] T. Harks. Stackelberg strategies and collusion in network games with splittable flow. *Theory Comput Syst*, 48:781802, 2011.

[Meir and Parkes, 2015] R. Meir and D. Parkes. Playing the wrong game: Smoothness bounds for congestion games with behavioral biases. *Proceedings of the 10th Workshop on the Economics of Networks, Systems and Computation* (NetEcon), Portland, OR, 2015.

[Nie, 2011] Y. Nie. Multi-class percentile user equilibrium with flow-dependent stochasticity. *Transportation Research Part B*, 45:1641–1659, 2011.

[Nikolova and Stier-Moses, 2011] E. Nikolova and N. E. Stier-Moses. Stochastic selfish routing. *Proceedings of the 4th International Symposium on Algorithmic Game Theory* (SAGT), pages 314–325, Berlin, Heidelberg, 2011. Springer.

[Nikolova and Stier-Moses, 2014] E. Nikolova and N. E. Stier-Moses. A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research*, 62(2):366–382, 2014.

[Nikolova and Stier-Moses, 2015] E. Nikolova and N. E. Stier-Moses. The burden of risk aversion in mean-risk selfish routing. In T. Roughgarden, M. Feldman, and M. Schwarz, editors, *Proceedings of the Sixteenth ACM Conference on Economics and Computation* (EC), pages 489–506. ACM, 2015.

[Ordóñez and Stier-Moses, 2010] F. Ordóñez and N. E. Stier-Moses. Wardrop equilibria with risk-averse users. *Transportation Science*, 44(1):63–86, 2010.

[Piliouras *et al.*, 2013] G. Piliouras, E. Nikolova, and J. S. Shamma. Risk sensitivity of price of anarchy under uncertainty. *Proceedings of the Fourteenth ACM Conference on Electronic Commerce* (EC), pages 715–732, ACM, 2013.

[Roughgarden and Schoppmann, 2015] T. Roughgarden and F. Schoppmann. Local smoothness and the price of anarchy in splittable congestion games. *Journal of Economic Theory*, 156:317–342, 2015.

[Roughgarden, 2003] T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67(2):341–364, 2003.

[Wardrop, 1952] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers, Part II, Vol. 1*, pages 325–378, 1952.