2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Risk-Sensitive Optimal Control of Queues

Rahul Singh, Xueying Guo, and Eytan Modiano

***Abstract*— We consider the problem of designing risk-sensitive optimal control policies for scheduling packet transmissions in a stochastic wireless network. A single client is connected to an access point (AP) through a wireless channel. Packet transmission incurs a cost C, while packet delivery yields a reward of R units. The client maintains a finite buffer of size B, and a penalty of L units is imposed upon packet loss which occurs due to buffer overflow.**

We show that the risk-sensitive optimal control policy for such a simple set-up is of threshold type, i.e., it is optimal to carry out packet transmissions only when $Q(t)$, i.e., the queue length at time t exceeds a certain threshold τ. It is also shown that the value of the threshold τ increases upon increasing the cost per unit packet transmission C. Furthermore, it is also shown that a threshold policy with threshold equal to τ is optimal for a set of problems in which cost C lies within an interval $[C_l, C_u]$. Equations that need to be solved in order to obtain C_l, C_u are also provided.

I. Introduction

In this work we consider the risk-sensitive optimal control of a one-hop *stochastic* wireless network that comprises of a single client. Networked control systems are becoming increasingly susceptible to attacks [1], and tools such as risk-sensitive and robust control can play an important role in securing these systems. Employement of a risk-sensitive control policy can serve as a mechanism to protect the network against attacks such as denial-of-service attacks.

Consider a denial-of-service attack carried out by a stochastic adversary that expends power in order to jam the communication channel between the client and the AP. Utilizing a risk-sensitive network control policy will make the closed system more robust to errors in the modelling assumptions made on the adversarial attack. The risk-sensitive optimal control policy hedges against the uncertainty by placing a greater emphasis on system trajectories that incur higher operation costs. If $c(t), t = 1, 2, \ldots, T$ denotes the instantaneous cost incurred during time t, then the risk-sensitive cost with risk-sensitivity parameter $\gamma > 0$ incurred during time period T is given by

$$\mathbb{E}e^{\gamma \sum_{t=1}^{T} c(t)},$$

where the expectation is taken with respect to the arrival process, the control policy used for scheduling packets, and the departure process. In the large-risk limit, i.e., $\gamma \to \infty$, the risk sensitive cost approaches the minimax cost objective, see [2]. Since the minimax objective seeks to minimize the system cost for the worst case scenario, a risk sensitive controller designed with risk parameter γ set to a large value, has a good performance in case the system dynamics are "adversarial" in nature. The framework provides flexibility by allowing the network operator to choose between the two competing objectives of having low risk-neutral cost, and that of making the system safe against attacks by tuning the risk-sensitivity parameter γ. Risk-sensitive control theory builds upon the ideas of Dynamic games and robust control [3]–[6] and allows the system operator to generate control actions that reflect his confidence about the uncertainty in the model. It also generlizes the risk neutral approach towards dynamic optimization [7]. Risk-sensitive control approach provides a link between the stochastic and deterministic approaches to model system uncertainty [8], [9].

Risk-sensitive optimization places emphasis on higher order moments of the system cost [10], and thus risk-sensitive optimal control reduces undesirable stochastic variations in the system performance. This is highly desirable for network control systems in which the control loop is closed [11]–[21]. Risk sensitive system cost takes into account higher order moments of the (random) cost as well, as opposed to the risk neutral cost objective which inlcudes only the mean cost. Since risk sensitive cost objective penalizes higher order moments, it allows for designing a finer controller for the cost of interest.

We discuss past works dealing with results on risk-sensitive control, and their applications in security of network control systems in Section II. The set-up involving single client being served by an access point is introduced in Section III. We derive the structure of the optimal policy for single client scheduling problem in Section IV. Section V derives the set of transmission costs for which threshold policy with threshold equal to τ is optimal. Section VI discusses directions for future research, and also summarizes the key results of this paper.

II. Past Works

The work [22] is one of the first to consider the problem of dynamic optimization of risk-sensitive cost within the Markov Decison Process (MDP) framework. For

Rahul Singh and Eytan Modiano are with the Laboratory of Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139, USA. rahulsiitk@gmail.com, modiano@mit.edu

Xueying Guo is postdoctoral researcher at Computer Science Department, University of California, Davis guoxueying@outlook.com

978-1-5090-2873-3/17/$31.00 ©2017 IEEE

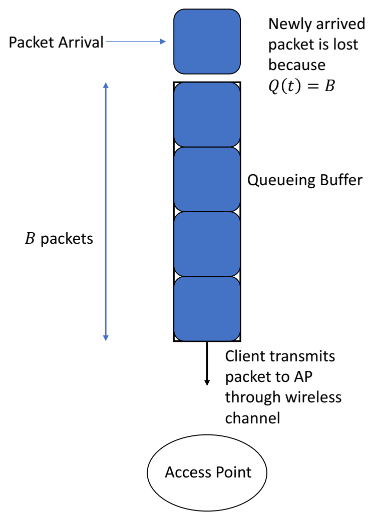

Fig. 1. A single client uses the stochastic wireles channel for carrying out packet transmissions. Size of the queueing buffer is equal to B packets, and if a packet arrives at a time t when the queue length $Q(t) = B$, then it is lost.

linear systems driven by Gaussian noise and quadratic one-step cost, [4] shows that the risk-sensitive controller depends upon the variance of noise, which is unlike the case of risk-sensitive LQG control. For a detailed treatment of risk-sensitive control of LQG systems, see [23]. Results concerned with risk-sensitive control of finite-state discrete-time controlled Markov chains can be found in [2], while [7] provides an overview of key results in risk-sensitive control.

In recent years, the problem of designing protocols and control policies for networked systems and crucial infrastructure such as sensor networks, electric power grids etc. has gained much attention [24], [25]. The authors of [26] considers the design of risk sensitive controller for a networked control system that is susceptible to denial-of-service attacks. The dynamical system of interest is assumed to be linear. Risk-sensitive control in the context of denial of service attacks in network, is studied in [27] and [28] derives scheduling policies that perform a mean versus variance trade-off with respect to packet interdelivery times.

Existing literature on stochastic control of queueing networks has mainly focused on a risk-neutral cost objective. Works such as [29]–[34] have derived optimal control policy and its structure under various assumptions regarding the stochastic queueing network. However there seems to be a gap with regards to the design of risk-sensitive control in the context of queueing networks.

III. Single Client Scheduling Problem

We begin by describing the risk-sensitive queue control problem involving a single client being served by an unreliable channel.

Continuous Time Model The system begins operation at time $t = 0$, and the packet arrivals to the client are governed by a Poisson process with rate λ. Let $Q(t)$ denote the queue length of the buffer at time t. If the client decides to carry out packet transmission at time t, then the time taken to complete packet transmission is exponentially distriibuted with mean $1/\mu$. During the time of packet transmission, cost is incurred at the rate of C units per unit time. The cost C models the amount of power utilized for packet transmission through the wireless medium. A reward of R units is generated upon a successful packet transmission, or equivalently the delivered packet is counted towards the network throughput [35]. The client maintains a queueing buffer of size B packets. A packet loss occurs at time t if a packet arrives and $Q(t) = B$, i.e. the queue buffer is full. The system is penalized L units upon a packet loss.

Equivalent Discrete-Time Model The continuous-time discrete space Markov process described above can be converted into an equivalent discrete-time Markov chain by sampling the embedded Markov chain at time epochs when a packet arrival or departure occurs. Such technique is commonly utilized in the analysis of queueing systems, see [29] or [36] Ch:10 for a detailed discussion. We now describe the discrete-time system in detail.

Let $Q(t)$ denote the queue length of the buffer at time t. The queue length $Q(t)$ of the client evolves over discrete time-slots $t = 1, 2, \ldots$. At each time t, the client can choose to either attempt packet transmission, i.e., $U(t) = 1$, or stay idle $U(t) = 0$. If $Q(t) > 0$ and the client attempts a packet transmission at time t, then the queue length at time $t+1$ is equal to $Q(t)-1$ with a probability p, while it is equal to $(Q(t) + 1) \wedge B$ with a probability $1 - p$. The quantity p is equal to the probability with which the packet transmission is completed before a new packet arrives in the original continuous-time model and is equal to $\mu/(\lambda + \mu)$. The client is charged $C > 0$ units for attempting to transmit a packet, and is provided a reward of $R > 0$ units upon successful packet delivery.

If at time t either the client decides to not carry out packet transmission, or if $Q(t) = 0$, then the queue length $Q(t+1)$ is equal to $(Q(t)+1) \vee B$ with probability 1. If an arriving packet at time t finds the queueing buffer full, i.e., $Q(t) = B$, then the packet is lost and the system is penalized $L > 0$ units. Figure 1 depicts the wireless network of interest. A history dependent scheduling policy π, for each time $t = 1, 2, \ldots$ maps the history of the system until time t to an action $U(t) \in \{0, 1\}$. A Markov policy π maps the queue length $Q(t)$ at time t to a decision $U(t) \in \{0, 1\}$. The infinite-horizon risk-sensitive cost incurred by the system is given by

$$\limsup_{T\to\infty} \frac{1}{\gamma T} \log \mathbb{E} \left\{ \exp \gamma \left(\sum_{t=1}^{T} CU(t) - R(t) + L(t) \right) \right\}, \tag{1}$$

where the random process $R(t)$ assumes the value R

if a packet is deliverd at time t, while is 0 otherwise, and the process $L(t)$ assumes the value L if a packet is lost at time t, and is 0 otherwise. The parameter $\gamma > 0$ controls the sensitivity of the client towards the risk, and is called risk-sensitivity parameter [10], [23]. If for any Markov policy π, the process $Q(t)$ is irreducible and aperiodic, the $\limsup$ in the above definition can be replaced by $\lim$ [37]. We briefly discuss the existing results on infinite horizon risk-sensitive control for finite-state Markov chains.

Results on Infinite Horizon Risk-Sensitive Control: Let us denote by $\pi^\star$ the policy that is optimal for the risk-sensitive MDP (1). It can be shown that ([7], [37], [38]) there exists a value function $V : [0, B] \mapsto \mathbb{R}$, and a scalar α, such that

$$\alpha V(i) = \min_{u \in \{0,1\}} \sum_{j \in [0,B]} e^{Co(i,j,u)} p(j|i,u) V(j), i \in [0, B] \tag{2}$$

where $p(j|i, u)$ is the transition probability associated with state i to state j under the application of control action u, and $Co(i, j, u)$ is the one-step cost associated with the state-action pair (i, a) and transition to state j. $\pi^\star(i)$ corresponds to the action u that minimizes the r.h.s. in the above equation for evaluation of $V(i)$.

Relative Value Iteration Algorithm: The fixed point equation (2) can be solved by carrying out the following fixed point iterations. Denote the estimate of the value function at iteration k by V_k. Then, the value function is updated according to

$$\tilde{V}_{k+1}(i) = \min_{u \in \{0,1\}} \sum_{j \in [0,B]} e^{C(i,j,u)} p(j|i,u) V_k(j), i \in [0, B]. \tag{3}$$

Thereafter normalize the iterates so that,

$$V_{k+1}(i) = \frac{\tilde{V}_{k+1}(i)}{\tilde{V}_{k+1}(0)}, \forall i \in [0, B]. \tag{4}$$

The policy generated at iteration k by the RVI algorithm applies the action that minimizes the r.h.s. of (3). It can be shown that for the RVI iterations, we have that $V_k \to V$, thereby yielding optimal policy [37]. Throughout, for $m \le n$, we denote by $[m, n]$ the set $\{m, m+1, \dots, n\}$.

IV. Structure of the Optimal Policy

We will show that the optimal policy for the single client scheduling problem is of threshold-type, i.e. it is optimal to carry out packet transmissions only when the queue length $Q(t)$ exceeds a certain threshold τ. The value of threshold τ depends on the system parameters p, and transmission cost C. We also show that τ increases with C.

Definition 1: A threshold policy with threshold τ, denoted as π_τ schedules packet transmissions at time $t = 1, 2, \dots$ only if the queue length $Q(t) \ge \tau$.

The Relative Value Iteration (RVI) algorithm discussed in the previous section converges, thus yielding optimal policy $\pi^\star$. We will show that at each iteration of the RVI algorithm, the produced policy is a threshold policy. This will prove that the optimal policy is of threshold-type.

Let V_k denote the value function at iteration k of the RVI algorithm. Thus, $V_k(n)$ denotes the relative cost associated with system state being in state n. Let $J_{k+1}(n, 1), J_{k+1}(n, 0)$ denote the costs associated with applying the actions $U(k+1) = 1$ and $U(k+1) = 0$ respecively when the system is in state n at stage $k+1$ of the RVI algorithm, i.e.,

$$J_{k+1}(n, 0) = \begin{cases} V_k(n+1), n \in [0, B-1] \\ e^{\gamma L} V_k(n), \text{ if } n = B, \end{cases} \tag{5}$$

$$J_{k+1}(n, 1) = \begin{cases} e^{\gamma C} V_k(n+1), \text{ if } n = 0, \\ p e^{\gamma(C-R)} V_k(n-1) + (1-p) e^{\gamma C} V_k(n+1), \\ \text{ if } n \in [1, B-1], \\ p e^{\gamma(C-R)} V_k(n-1) + (1-p) e^{\gamma(C+L)} V_k(n), \\ \text{ if } n = B. \end{cases} \tag{6}$$

Let $\partial J_{k+1}(n) := J_{k+1}(n, 0) - J_{k+1}(n, 1)$ denote the differential between the costs associated with taking the actions 0 and 1 if the queue length $Q(k+1)$ at iteration $k+1$ is equal to n. The differential ∂J_{k+1} is given as,

$$\partial J_{k+1}(n) = \begin{cases} (1 - e^{\gamma C}) V_k(n+1) \text{ if } n = 0, \\ V_k(n+1) \left[1 - (1-p) e^{\gamma C}\right] \\ \qquad - p e^{\gamma(C-R)} V_k(n-1), \text{ if } n \in [1, B-1], \\ V_k(n) \left[1 - (1-p) e^{\gamma C}\right] e^{\gamma L} \\ \qquad - p e^{\gamma(C-R)} V_k(n-1) \text{ if } n = B. \end{cases} \tag{7}$$

We clearly have,

Lemma 1: If the differential $\partial J_{k+1}(n), n \in [0, B]$ is a non-decreasing function of n, then the optimal policy produced at iteration $k+1$ by the RVI algorithm (3)-(4) is of threshold type.

Let us assume that ∂J_{k+1} is non-decreasing in n, and try to prove that the function ∂J_{k+2} is non-decreasing in n. This result will then imply that the optimal policy produced at iteration $k+2$ is also of threshold type.

Lemma 2: Let the optimal policy produced by the RVI algorithm at iteration $k+1$ be of threshold-type, with threshold value equal to τ. Then the differential ∂J_{k+1} satisfies

$$\partial J_{k+1}(n) \le 0, n \in [0, \tau - 1], \text{ and} \tag{8}$$
$$\partial J_{k+1}(n) \ge 0, n \in [\tau, B]. \tag{9}$$

The unscaled value function $\tilde{V}_{k+1}$ produced at iteration

$k+1$ is given by,

$$\tilde{V}_{k+1}(n) = \begin{cases} V_k(n+1), \text{ if } n \in [0, \tau-1], \\ pe^{\gamma(C-R)}V_k(n-1) + (1-p)e^{\gamma C}V_k(n+1), \\ \text{ if } n \in [\tau, B-1], \\ pe^{\gamma(C-R)}V_k(n-1) + (1-p)e^{\gamma(C+L)}V_k(n), \\ \text{ if } n = B. \end{cases} \tag{10}$$

We can now use the expression of ∂J_{k+2} derived in Lemma 3 of [39] in order to show that it is non-decreasing function of n.

Lemma 3: Assume that the differential ∂J_{k+1} at iteration $k+1$ is non-decreasing function of n. Then, the differential ∂J_{k+2} at iteration $k+2$ is also non-decreasing in n.

Proof: It follows from Lemma 3 of [39] that for $n \in [\tau+1, B-1]$, the function $\partial J_{k+2}(n)$ is a linear combination of the functions $\partial J_{k+1}(n-1)$ and $\partial J_{k+1}(n+1)$, both of which are assumed to be non-decreasing functions of n. Thus, the claim is true for $n \in [\tau+1, B-1]$. Similar reasoning proves the claim for $n \in [1, \tau-2]$.

We now verify whether the following two inequalities are true,

$$\partial J_{k+2}(\tau+1) \geq \partial J_{k+2}(\tau) \text{ and } \partial J_{k+2}(\tau) \geq \partial J_{k+2}(\tau-1).$$

We note that,

$$\begin{aligned} \partial J_{k+2}(\tau+1) &= \partial J_{k+1}(\tau)pe^{\gamma(C-R)} \\ &\quad + \partial J_{k+1}(\tau+2)(1-p)e^{\gamma C} \\ &\geq \partial J_{k+1}(\tau+2)(1-p)e^{\gamma C} \\ &\geq \partial J_{k+1}(\tau+1)(1-p)e^{\gamma C} \\ &= \partial J_{k+2}(\tau), \end{aligned}$$

where the first inequality follows since the optimal policy at iteration $k+1$ is of threshold type, and from Lemma 2 we have that $\partial J_{k+1}(\tau) \geq 0$. The second inequality follows from our assumption that ∂J_{k+1} is non-decreasing in n, i.e., $\partial J_{k+1}(\tau+2) \geq \partial J_{k+1}(\tau+1)$.

Next, we have,

$$\begin{aligned} \partial J_{k+2}(\tau) &= (1-p)e^{\gamma C}\partial J_{k+1}(\tau+1) \\ &\geq (1-p)e^{\gamma C}\partial J_{k+1}(\tau) \\ &= \partial J_{k+2}(\tau-1), \end{aligned}$$

where the inequality follows from our assumption that $\partial J_{k+1}(\tau)$ is non-decreasing.

We now prove $\partial J_{k+2}(0) \leq \partial J_{k+2}(1)$. We substitute the values of $\partial J_{k+2}(0), \partial J_{k+2}(1)$ from Lemma 3 of [39], so that for $n=0$ we have,

$$\begin{aligned} \partial J_{k+2}(n) &= \left[1 - e^{\gamma C}(1-p)\right] V_k(n+2) - pe^{\gamma(C)}V_k(n) \\ &\leq \left[1 - e^{\gamma C}(1-p)\right] V_k(n+2) - pe^{\gamma(C-R)}V_k(n) \\ &\leq \left[1 - e^{\gamma C}(1-p)\right] V_k((n+1)+2) \\ &\quad - pe^{\gamma(C-R)}V_k(n+1) \\ &= \partial J_{k+2}(n+1), \end{aligned}$$

where the first inequality follows since $R > 0$, and the second inequality follows since ∂J_{k+1} is assumed to be non-decreasing in n.

Finally, we prove $\partial J_{k+2}(B) \geq \partial J_{k+2}(B-1)$. Substituting the values of $\partial J_{k+2}(B)$ from Lemma 3 of technical report [39], and the value of $\partial J_{k+2}(B-1)$ from (7), the condition $\partial J_{k+2}(B) \geq \partial J_{k+2}(B-1)$ reduces to,

$$\begin{aligned} &V_{k+1}(B)[1-(1-p)e^{\gamma C}]e^{\gamma L} - pe^{\gamma(C-R)}V_{k+1}(B-1) \\ &\geq V_{k+1}(B)\left[1-(1-p)e^{\gamma C}\right] - pe^{\gamma(C-R)}V_{k+1}(B-2), \end{aligned}$$

or equivalently

$$\begin{aligned} &V_{k+1}(B)\left[1-(1-p)e^{\gamma C}\right]\left[e^{\gamma L}-1\right] \\ &\geq pe^{\gamma(C-R)}\left[V_{k+1}(B-1) - V_{k+1}(B-2)\right] \end{aligned}$$

This concludes the proof. ∎

Theorem 1 (Optimality of Threshold Policy): For the single client risk-sensitive scheduling problem of minimizing the infinite-horizon cost (1), a threshold policy is optimal.

Proof: We will use induction on the iteration number k of the RVI algorithm in order to prove the theorem. For the RVI algorithm, let us initialize the $V_0(n) = 1, \forall n \in [0, B]$. It then follows that,

$$\partial J_1(n) = \begin{cases} (1-e^{\gamma C}) \text{ if } n = 0, \\ \left[1-(1-p)e^{\gamma C}\right] - pe^{\gamma(C-R)}, \\ \quad \text{if } n \in [1, B-1], \\ \left[1-(1-p)e^{\gamma C}\right]e^{\gamma L} - pe^{\gamma(C-R)} \text{ if } n = B. \end{cases} \tag{11}$$

It is easily verified that ∂J_1 is non-decreasing in n. Thus, it now follows from Lemma 3, that at each iteration k of the RVI algorithm, the function ∂J_k is non-decreasing in n. Thus, from Lemma 1 we have that the policy produced by the RVI algorithm at each iteration k is of threshold type. Since the RVI algorithm converges to the optimal policy, the optimal policy is also of threshold type. ∎

Next, we show that for the optimal policy $\pi^\star$, the threshold denoted by $\tau^\star$ increases with the transmission cost C. The following condition ensures that the threshold of the policy produced by the RVI algorithm at stage $k+1$ is an increasing function of transmission cost C.

Condition 1 (Monotonicity): If $C_1, C_2 > 0$ are such that $C_1 > C_2$, then $\partial J_{k+1}^{C_1}(m) \leq \partial J_{k+1}^{C_2}(m)$ for each $m \in [0, B]$.

Lemma 4: Assume that the Condition 1 is true for the RVI algorithm at iteration k. Then, the Condition 1 also holds true at iteration $k+1$ of the RVI algorithm, and hence for the policy produced at iteration $k+1$, the threshold value is an increasing function of the transmission cost C.

Proof: In the ensuing discussion, we let $V_{k,1}$ denote the value function associated with k-th iteration of RVI applied to the risk-sensitive control problem (1) with transmission cost set at C_1, while $\partial J_k^{C_1}$ will denote the corresponding cost differential. Similarly for $V_{k,2}, \partial J_k^{C_2}$.

In order to prove the claim, we need to show that $\partial J_{k+2}^{C}(n)$ is increasing function of C for each $n \in [0, B]$. For an $n \in [\tau+1, B-1]$, and $C_1 > C_2 > 0$ we have,

$$\begin{aligned}
&\partial J_{k+2}^{C_1}(n) \\
&= (1-p)e^{\gamma C_1}\partial J_{k+1}^{C_1}(n+1) + pe^{\gamma(C_1-R)}\partial J_{k+1}^{C_1}(n-1) \\
&\leq (1-p)e^{\gamma C_2}\partial J_{k+1}^{C_2}(n+1) + pe^{\gamma(C_2-R)}\partial J_{k+1}^{C_2}(n-1) \\
&= J_{k+2}^{C_2}(n),
\end{aligned}$$

where the equalities follow from the relation (12) of [39] and the inequality follows from our assumption that the Condition 1 is satisfied at iteration $k+1$ of the RVI algorithm.

Next, we prove the claim for $n \in [1, \tau-1]$. For $C_1 > C_2 > 0$ and $n \in [1, \tau-2]$ we have

$$\begin{aligned}
&\partial J_{k+2}^{C_1}(n) \\
&= \left[1 - e^{\gamma C_1}(1-p)\right] V_{k,1}(n+2) - pe^{\gamma(C_1-R)}V_{k,1}(n) \\
&= \partial J_k^{C_1}(n+1) \\
&\leq \partial J_k^{C_2}(n+1) \\
&= \partial J_{k+2}^{C_2}(n),
\end{aligned}$$

where the inequality results from Condition 1.

Now we prove the desired condition for $n = 0$. It follows from (7) that the condition $\partial J_{k+2}^{C_1}(0) \leq \partial J_{k+2}^{C_2}(0)$ reduces to $(1-e^{\gamma C})V_{k+1,1}(1) \leq (1-e^{\gamma C_2})V_{k+1,2}(1)$. Since $C_1, C_2 > 0$ the condition is equivalent to $V_{k+1,1}(1) \geq V_{k+1,2}(1)$. Fix a time horizon $T > 0$, and a scheduling policy π, and consider the operation of two systems under the application of the policy π. The transition probabilities of the two controlled Markovian systems are taken to be the same, but their transmission costs are set at C_1 and C_2. Construct their sample paths on the same probability space. It now follows from stochastic coupling [40], that the sample path cost $\sum_{t=1}^{T} CU(t) - R(t) + L(t)$, or equivalently the cost $e^{\gamma \sum_{t=1}^{T} CU(t)-R(t)+L(t)}$ incurred by the system with cost set at C_1 is greater than or equal to the system with cost equal to C_2. Hence it follows that $V_{k+1,1}(1) \geq V_{k+1,2}(1)$.

For $n = \tau, \tau-1$, the differential $\partial J_{k+2}^{C_1}(n)$ yields us

$$\begin{aligned}
&\partial J_{k+2}^{C_1}(n) \\
&= (1-p)e^{\gamma C_1}\partial J_{k+1}^{C_1}(n+1) \\
&\leq (1-p)e^{\gamma C_1}\partial J_{k+1}^{C_2}(n+1) \\
&= \partial J_{k+2}^{C_2}(n),
\end{aligned}$$

where the equality follows from the relation (12) of [39], and the inequality results from Condition 1. ■

Theorem 2: Consider the problem of designing a scheduling policy that makes decisions regarding packet transmissions in order to minimize the infinite horizon risk-sensitive cost (1). For $C_1 > C_2 > 0$, let τ_{C_1} and τ_{C_2} denote the threshold values of the optimal policies when transmission costs are set at C_1 and C_2 respectively. We then have $\tau_{C_1} \geq \tau_{C_2}$.

Proof: Consider the optimal risk-sensitive control problem (1) with transmission cost set at C, and initialize $V_0(n) = 1, \forall n \in [0, B]$. We then have

$$\partial J_1(n) = \begin{cases} (1-e^{\gamma C}) \text{ if } n = 0, \\ \left[1-(1-p)e^{\gamma C}\right] - pe^{\gamma(C-R)}, \\ \qquad \text{if } n \in [1, B-1], \\ \left[1-(1-p)e^{\gamma C}\right] e^{\gamma L} - pe^{\gamma(C-R)} \text{ if } n = B. \end{cases} \tag{12}$$

It is easily verified that $\partial J_1(n)$ is non-increasing function of C, and hence Condition 1 holds true at iteration $k = 1$ of the RVI algorithm.

The result now follows by using induction on iteration number k in conjunction with Lemma 4. ■

V. Computing the Optimal Policy

Having derived the structure of the optimal policy, we would like to compute the value of threshold τ corresponding to the optimal policy. In view of Theorem 2, we will derive the set of values of transmission cost C such that the policy π_τ is optimal when the transmission cost is set at C. We omit the details due to space limitations. A detailed discussion is provided in Section V of [39]. We directly state the main result.

Theorem 3: Consider the class comprising of optimal risk-sensitive control problems parametrized by transmission cost C, in which for each individual risk-sensitive MDP the cost incurred is given by (1). Then, the threshold policy π_τ is optimal for risk-sensitive MDPs for which the cost $C \in [C_l, C_u]$, where C_l, C_u can be obtained by solving the inequalities (24)-(25) of [39] .

VI. Conclusion and Future Works

We have derived the optimal risk-sensitive scheduling policy for a single client being served by a wireless channel. The otimal policy was shown to have a threshold structure, and hence is easily implementable. Furthermore we showed that the threshold increases with packet transmission cost, and hence the policy with threshold set at τ is optimal when the transmission cost lies within the interval $[C_l, C_u]$. The quantities C_l, C_u can be derived by solving two equations. We plan to extend the analysis to the case where multiple clients share a single wireless channel, and the AP has to prioritize the clients for packet transmissions, based on their queue lengths. We would also like to consider the scenario where the transmitter can choose to transmit from amongst various power levels, where a transmission involving higher power having a higher service rate.

References

[1] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *HotSec*, 2008.

[2] S. P. Coraluppi and S. I. Marcus, "Risk-sensitive and minimax control of discrete-time, finite-state markov decision processes," *Automatica*, vol. 35, no. 2, pp. 301–309, 1999.

[3] M. James, "Asymptotic analysis of nonlinear stochastic risk-sensitive control and differential games," *Mathematics of Control, Signals and Systems*, vol. 5, no. 4, pp. 401–417, 1992.

[4] D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *Automatic Control, IEEE Transactions on*, vol. 18, no. 2, pp. 124–131, Apr 1973.

[5] A. Bensoussan and J. Van Schuppen, "Optimal control of partially observable stochastic systems with an exponential-of-integral performance index," *SIAM Journal on Control and Optimization*, vol. 23, no. 4, pp. 599–613, 1985.

[6] M. R. James, J. S. Baras, and R. J. Elliott, "Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems," *IEEE transactions on automatic control*, vol. 39, no. 4, pp. 780–792, 1994.

[7] S. I. Marcus, E. Fernandez-Gaucherand, D. Hernandez-Hernandez, S. Coraluppi, and P. Fard, "Risk sensitive Markov decision processes," in *Systems and control in the 21st century*, 1997.

[8] W. Fleming and D. Hernandez-Hernandez, "Risk sensitive control of finite state machines on an infinite horizon. i," in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, vol. 4, Dec 1997, pp. 3407–3412 vol.4.

[9] W. H. Fleming and W. M. McEneaney, "Risk-sensitive control on an infinite time horizon," *SIAM Journal on Control and Optimization*, vol. 33, no. 6, pp. 1881–1915, 1995.

[10] P. Kumar and J. Van Schuppen, "On the optimal control of stochastic systems with an exponential-of-integral performance index," *Journal of mathematical analysis and applications*, vol. 80, no. 2, pp. 312–332, 1981.

[11] R. Singh, I.-H. Hou, and P. Kumar, "Fluctuation analysis of debt based policies for wireless networks with hard delay constraints," in *IEEE INFOCOM, 2014 Proceedings*, April 2014, pp. 2400–2408.

[12] Rahul Singh, I-Hong Hou and P.R. Kumar, "Pathwise performance of debt based policies for wireless networks with hard delay constraints," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, Dec 2013, pp. 7838–7843.

[13] X. Guo, R. Singh, P. Kumar, and Z. Niu, "A high reliability asymptotic approach for packet inter-delivery time optimization in cyber-physical systems," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '15. New York, NY, USA: ACM, 2015, pp. 197–206. [Online]. Available: http://doi.acm.org/10.1145/2746285.2746305

[14] R. Singh and A. Stolyar, "Maxweight scheduling: "smoothness" of the service process," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, April 2016, pp. 1–9.

[15] ——, "Maxweight scheduling: Asymptotic behavior of unscaled queue-differentials in heavy traffic," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 43, no. 1. ACM, 2015, pp. 431–432.

[16] ——, "Maxweight scheduling: Asymptotic behavior of unscaled queue-differentials in heavy traffic," in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '15. New York, NY, USA: ACM, 2015, pp. 431–432.

[17] R. Singh and E. Modiano, "Optimal routing for delay-sensitive traffic in overlay networks," *arXiv preprint arXiv:1703.07419*, 2017.

[18] R. Singh and P. Kumar, "Throughput optimal decentralized scheduling of multi-hop networks with end-to-end deadline constraints: Unreliable links," *arXiv preprint arXiv:1606.01608*, 2016.

[19] I. Hou, R. Singh *et al.*, "Capacity and scheduling of access points for multiple live video streams," *arXiv preprint arXiv:1306.2360*, 2013.

[20] ——, "Scheduling of access points for multiple live video streams," in *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2013, pp. 267–270.

[21] R. Singh, P. Kumar, and E. Modiano, "Throughput optimal decentralized scheduling of multi-hop networks with end-to-end deadline constraints: Ii wireless networks with interference," *arXiv preprint arXiv:1709.01672*, 2017.

[22] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Management Science*, vol. 18, no. 7, pp. pp. 356–369, 1972.

[23] P. Whittle, "Risk-sensitive optimal control," 1990.

[24] E. Bompard, C. Gao, R. Napoli, A. Russo, M. Masera, and A. Stefanini, "Risk assessment of malicious attacks against power systems," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 39, no. 5, pp. 1074–1085, 2009.

[25] G. N. Ericsson, "Toward a framework for managing information security for an electric power utility?cigré experiences," *IEEE transactions on power delivery*, vol. 22, no. 3, pp. 1461–1469, 2007.

[26] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2009, pp. 31–45.

[27] G. K. Befekadu, V. Gupta, and P. J. Antsaklis, "Risk-sensitive control under a class of denial-of-service attack models," in *American Control Conference (ACC), 2011*. IEEE, 2011, pp. 643–648.

[28] R. Singh, X. Guo, and P. R. Kumar, "Index policies for optimal mean-variance trade-off of inter-delivery times in real-time sensor networks," in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 505–512.

[29] Z. Rosberg, P. Varaiya and J. Walrand, "Optimal control of service in tandem queues," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 600–610, Jun 1982.

[30] C. Buyukkoc, P. Variaya, and J. Walrand, "c mu rule revisited." *Adv. Appl. Prob.*, vol. 17, no. 1, pp. 237–238, 1985.

[31] W. Lin and P. Kumar, "Optimal control of a queueing system with two heterogeneous servers," *IEEE Transactions on Automatic control*, vol. 29, no. 8, pp. 696–703, 1984.

[32] S. Stidham Jr and R. R. Weber, "Monotonic and insensitive optimal policies for control of queues with undiscounted costs," *Operations Research*, vol. 37, no. 4, pp. 611–625, 1989.

[33] J. M. George and J. M. Harrison, "Dynamic control of a queue with adjustable service rate," *Operations Research*, vol. 49, no. 5, pp. 720–731, 2001.

[34] J. M. Harrison, "Dynamic scheduling of a multiclass queue: Discount optimality," *Operations Research*, vol. 23, no. 2, pp. 270–282, 1975.

[35] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[36] L. I. Sennott, *Stochastic dynamic programming and the control of queueing systems*. John Wiley & Sons, 2009, vol. 504.

[37] V. S. Borkar, "Q-learning for risk-sensitive control," *Mathematics of operations research*, vol. 27, no. 2, pp. 294–311, 2002.

[38] A. Brau-Rojas, R. Cavazos-Cadena, and E. Fernandez-Gaucherand, "Controlled Markov chains with risk-sensitive criteria: some (counter) examples," in *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on*, vol. 2, Dec 1998, pp. 1853–1858 vol.2.

[39] R. Singh, X. Guo, and E. Modiano, "Risk-sensitive optimal control of queues," *arXiv preprint arXiv:1708.08178*, 2017.

[40] H. Thorisson, "Coupling methods in probability theory," *Scandinavian journal of statistics*, pp. 159–182, 1995.