Cost-aware Defense for Parallel Server Systems against Reliability and Security Failures ⋆

Qian Xie [a,d], Jiayi Wang [b,c], Li Jin [c,d]

[a] *School of Operations Research and Information Engineering, Cornell University, USA.*

[b] *Department of Electrical and Computer Engineering, University of California, San Diego, USA.*

[c] *UM Joint Institute, Shanghai Jiao Tong University, China.*

[d] *Tandon School of Engineering, New York University, USA.*

Abstract

Parallel server systems in transportation, manufacturing, and computing heavily rely on dynamic routing using connected cyber components for computation and communication. Yet, these components remain vulnerable to random malfunctions and malicious attacks, motivating the need for resilient dynamic routing that is both traffic-stabilizing and cost-efficient. In this paper, we consider a parallel server system with dynamic routing subject to reliability and stability failures. For the reliability setting, we consider an infinite-horizon Markov decision process where the system operator strategically activates a protection mechanism upon each job arrival based on traffic state observations. We prove that an optimal deterministic threshold protecting policy exists based on the dynamic programming recursion of the Hamilton-Jacobi-Bellman equation. For the security setting, we extend the model to an infinite-horizon stochastic game where the attacker strategically manipulates routing assignments. We show that both players follow a threshold strategy at every Markov perfect equilibrium. For both failure settings, we also analyze the stability of the traffic queues. Finally, we develop approximate dynamic programming algorithms to compute the optimal/equilibrium policies and present numerical examples/experiments for validation and illustration.

Key words: Queuing systems, cyber-physical security, stochastic games, Markov decision processes, HJB equation, Lyapunov function.

1 Introduction

The parallel server system is a classical model characterizing a service system of multiple servers, each with a waiting queue. Real-world instances include web server farms (Gupta et al., 2007), production lines (Govil and Fu, 1999), and transportation facilities (Jin and Amin, 2018). These systems use feedback from state observations to generate dynamic routing decisions, enhancing stability and throughput. However, their reliance on connected cyber components for data collection and transmission emposes them to persistent threats from malfunctions and manipulations (Cardenas et al., 2009). Malfunctions can arise from technical issues including network congestion, server unresponsiveness, packet loss, firewall restriction, signal interference, and authentication errors (Alpcan and Başar, 2010), or malicious attacks such as Denial-of-Service (DoS) (Wang et al., 2007; Al-Kahtani, 2012) that overwhelm servers with excessive traffic and cut off state observations. These disruptions may lead to the system operator's failure to deliver correct instructions. To illustrate the cause and impact of the malfunctions, consider two real-world motivating examples:

1. **Transportation:** Imagine a vehicle experiencing a failure in receiving routing information from a navigation app due to network connection issues. In this situation, drivers often resort to independent routing decisions based on personal preferences, such as route types, tolls, scenery, and familiarity.
2. **Manufacturing:** Similarly, in a production line, where production units are supposed to be routed to the shortest queue based on real-time routing information, a communication failure or breakdown can trigger a fallback mechanism (Fraile et al., 2018),

⋆ This work was in part supported by US NSF Award CMMI-1949710, C2SMART University Transportation Center, NYU Tandon School of Engineering, SJTU-UM Joint Institute, J. Wu & J. Sun Endowment Fund, and Cornell University McMullen Fellowship.

Email addresses: qx66@cornell.edu (Qian Xie), jiw139@ucsd.edu (Jiayi Wang), li.jin@sjtu.edu.cn (Li Jin).

leading to a random assignment to a default queue.

The above examples demonstrate two consequences of failures[1]: (i) routing based on individual preferences, historical data, or random selection; (ii) joining a default queue following a fallback mechanism. From the system perspective, the routing choices in the former outcome exhibit a random nature.

Manipulations, on the other hand, describe strategic attacks from adversaries with selfish or malicious intent. These include (i) *spoofing attacks* that directly send deceptive routing instructions to arrivals by impersonating the system operator and (ii) *falsification attacks* that inject misleading queue length data or create fictitious traffic, indirectly influencing the system operator's routing decisions (Feng et al., 2022; Al-Kahtani, 2012; Sakiz and Sen, 2017). For instance, a simulated traffic jam can cause motorists to deviate from their planned routes (Gravé-Lazi, 2014). Transportation infrastructure information (e.g., traffic sensors, traffic lights) and vehicle communications can also be intruded and manipulated (Feng et al., 2022; Sakiz and Sen, 2017; Al-Kahtani, 2012). Similar security risks also exist in production lines (Barrère et al., 2020; Fraile et al., 2018) and communication systems (Alpcan and Başar, 2010; Manshaei et al., 2013; De Persis and Tesi, 2015).

Real-world service systems will not be accepted by authorities, industry, and the public unless security issues are well addressed. However, cyber security risks have not been sufficiently studied in conjunction with the physical queuing dynamics. Furthermore, perfectly avoiding cyber failures is economically infeasible and technically unnecessary. Therefore, it is crucial to understand the impact of such threats and to design practical defense mechanisms. In practice, these defense mechanisms can be implemented with dynamic activation/deactivation of prevention/detection measures such as robust data validation, instruction encryption, and strict security protocol adherence (Cardenas et al., 2009; Manshaei et al., 2013). Nonetheless, these actions, while active, entail technological costs on computational resources, network bandwidth, energy consumption, maintenance efforts, etc.

In response to the above concerns, we aim to address the following two research questions in this paper:

(i) *How to model the security vulnerabilities and quantify the security risks for parallel queuing systems?*
(ii) *How to design traffic-stabilizing, cost-efficient defense strategies against failures?*

For the first question, we consider two scenarios of failures, viz. *reliability failures* and *security failures.* We formulate the security risks in terms of failure-induced queuing delays and defending costs. For the second question, we analyze the stability criteria of the failure-prone system with defense and characterize the structure of the cost-efficient strategies. We also develop algorithms to compute such strategies and discuss how to incorporate the stability condition. Our results are also demonstrated via numerical examples and simulations.

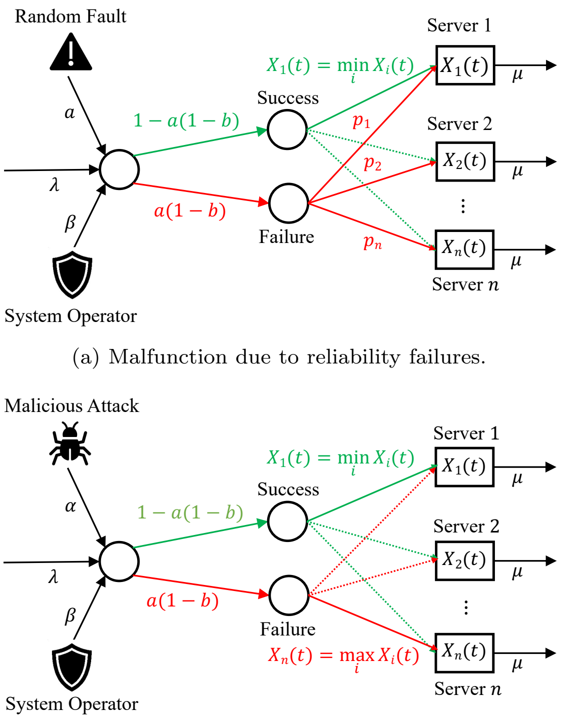
This paper is related to two lines of work: queuing control and game theory. On the queuing side, the majority of the existing analysis and design are based on perfect observation of the states (i.e., queue lengths) and perfect implementation of the control (Ephremides et al., 1980; Halfin, 1985; Eschenfeldt and Gamarnik, 2018; Knessl et al., 1986). Besides, researchers have noted the impact of delayed (Kuri and Kumar, 1995; Mehdian et al., 2017), erroneous (Beutler and Teneketzis, 1989; Xie and Jin, 2020), or decentralized (Ouyang and Teneketzis, 2015) information. Although these results provide hints for our problem, they do not directly apply to the security setting with failures such as imperfect sensing (state observation) and imperfect control implementation. On the game side, a variety of game-theoretic models have been applied to studying cyber-physical security in transportation (Tang et al., 2020; Laszka et al., 2019) and communication (Bohacek et al., 2007; Alpcan and Başar, 2010; Manshaei et al., 2013). However, to the best of our knowledge, the security risks of queuing systems have not been well studied from a combined game-theoretic and queuing-control perspective, which is essential for capturing the coupling between the queuing dynamics and the attacker-defender interactions.

Our model includes two parts: the physical part (parallel servers) and the cyber part (dynamic routing with failures). Specifically, we investigate two failure scenarios:

(1) **Reliability failures.** A fault may occur with a constant probability and the system operator can choose to activate protection for each arrival. In the event of a routing malfunction and the absence of activated protection, the arrival joins a random queue following certain probabilities; see Fig. 1a.
(2) **Security failures.** An adversary can launch an attack on each arrival using a feedback strategy and the system operator can choose to activate defense for each arrival. In the event of an effective attack and the absence of activated defense, the arrival joins an adversary-desired queue, with the worst-case scenario being the longest queue; see Fig. 1b.

To study the stability of a queuing system, previous works typically relied on characterization or approximation of the steady-state distribution of queuing states (Foley and McDonald, 2001). However, this approach is hard to integrate with failure models. Additionally, the steady-state distribution of queuing systems with state-dependent transition rates is intricate. In response to these challenges, we adopt a *Lyapunov function*-based approach which has been applied to queuing systems in no-failure scenarios (Kumar and Meyn, 1995; Dai and Meyn, 1995; Eryilmaz and Srikant, 2007; Xie and Jin, 2022) and enables us to derive *stability criteria* for queuing systems under control and to obtain upper bounds for the long-term average number of jobs in the system (Meyn and Tweedie, 1993).

[1] This paper does not consider scenarios in which arrivals leave or get rejected, say packet loss in computer networks.



(a) Malfunction due to reliability failures.

(b) Tampering due to security failures.

Fig. 1. An n-queue system with shortest-queue routing under failures. See Section 2 for definitions of notations.

To analyze the optimal/equilibrium strategies, we extend the dynamic programming recursion technique in (Bertsekas, 2012, Chapter 4) and (Hajek, 1984) from two queues to n queues and failure-prone settings. This approach allows us to prove the threshold properties of the optimal protection under reliability failures and the equilibria of the security game based on *Hamilton-Jacobi-Belmman (HJB) equations.*

Our theoretical analysis provides practical insights for cost-aware strategic defense design. A key finding is that the operator has a higher incentive to protect/defend if the queues are more "imbalanced". In addition, numerical analysis indicates that 1) the incentive to protect grows with failure probability, declines with technological cost, and grows with traffic intensity; 2) the optimal protecting policy outperforms static policies such as never protect and always protect. We also note that the optimal policy is not always stabilizing, leading to the proposal of a stability-constrained optimal policy.

Our contributions lie in the following three aspects:

(1) **Modeling:** We model the cyber-physical vulnerabilities of parallel servers with dynamic routing under reliability/security failures. We also formulate the trade-off between queuing and technological costs in the two failure scenarios as a *Markov decision process* and a *stochastic game* respectively.

(2) **Analysis:** We establish stability criteria for dynamic routing with failures using Lyapunov functions. We also show the threshold properties of the optimal protection and the game equilibria on multidimensional state space using DP recursion.

(3) **Design:** Our theoretical results offer insights into the design of cost-aware defense mechanisms. We also propose algorithms for estimating stability-constrained optimal policies and game equilibria.

This paper is structured as follows. Section 2 presents the cyber-physical model. Section 3 and 4 study protection under reliability failures and defense against security failures, respectively. Section 5 gives a conclusion.

2 Parallel servers and failure models

2.1 Parallel server system

Consider a queuing system with n identical parallel servers. Jobs (e.g., vehicles, customers, production units) arrive according to a Poisson process of rate $\lambda > 0$. Each server serves jobs at an exponential rate of $\mu > 0$. The number of jobs either waiting or being served in the n servers at time t is denoted by $X(t) = \begin{bmatrix} X_1(t) \ X_2(t) \ \cdots \ X_n(t) \end{bmatrix} \in \mathbb{Z}^n_{\geq 0}$.

We use $x + (-)e_i$ to denote adding 1 to (subtracting 1 from) the i-th element x_i. Since the queue lengths are always non-negative, i.e., $x_i \geq 0$, we use $(x - e_i)^+ = \max(x - e_i, 0)$ to avoid the case that subtracting 1 makes the element negative. Let $x_{\min} = \min_i x_i$ and $x_{\max} = \max_i x_i$. We use x_{-i} to denote variables in x other than x_i, and we use the notation $x + e_{\min}$ when adding 1 to $x_{\min}$ while keeping x_{-i} the same. We call x a diagonal vector if $x_1 = x_2 = \cdots = x_n$ and a non-diagonal vector otherwise. Denote the one-norm of the vector x as $||x||_1 := x_1 + x_2 + \cdots + x_n$. Then $||X(t)||_1$ means the total number of jobs in the system at time t. We use $x \succ \mathbf{0}$ to denote that x is not a zero vector, i.e., $||x||_1 > 0$.

Without any failures, any arriving job should be allocated to the shortest queue. If there are multiple shortest queues, then the job is randomly allocated to one of them with (not necessarily equal) probabilities.

2.2 Reliability failures

Suppose that upon the arrival of a job to the system, a fault may occur with a constant probability $a \in (0, 1]$ and lead to a routing instruction malfunction. Consequently, the job joins a random queue with respective probabilities [2] $p_1, p_2, \cdots, p_n \in [0, 1]$, where $\sum_{i=1}^n p_i = 1$. For convenience, we define $p_{\max} := \max(p_1, p_2, \cdots, p_n)$.

The system operator can decide whether to protect an arriving job to ensure its optimal routing, i.e., the shortest-queue routing, as illustrated in Fig. 1a. However, such protection comes at the cost of a rate $c_b > 0$.

[2] The operator is assumed to know the values of random routing probabilities p, which can be either estimated using historical data and statistical techniques or predetermined by a fallback mechanism, contingent upon specific contexts.

In this scenario, the system operator faces a trade-off between queuing and protection costs. We formulate this problem as an *infinite-horizon continuous-time Markov decision process* with queue lengths as states. The operator adopts a *Markovian policy*, denoted as $\beta : \mathbb{Z}^n_{\geq 0} \to \Delta(\{\mathrm{NP}, \mathrm{P}\})$, with $\Delta(\{\mathrm{NP}, \mathrm{P}\}) := \{(1-b, b) : b \in [0, 1]\}$ representing the probability distribution over the action set {not protect, protect}. For simplicity, when the policy is deterministic, we write the mapping as $\beta : \mathbb{Z}^n_{\geq 0} \to \{\mathrm{NP}, \mathrm{P}\}$. The transition matrix $P_{\mathrm{R}} : \mathbb{Z}^n_{\geq 0} \times \{\mathrm{NP}, \mathrm{P}\} \mapsto \Delta\left(\mathbb{Z}^n_{\geq 0}\right)$ capturing queuing dynamics under the protection against reliability failures is given by $P_{\mathrm{R}}(x + e_{\min} \mid x, \mathsf{b}) = (1 - a\mathbb{1}\{\mathsf{b} = \mathrm{NP}\})\,\lambda$, $P_{\mathrm{R}}(x + e_i \mid x, \mathsf{b}) = a\mathbb{1}\{\mathsf{b} \neq \mathrm{NP}\}p_i\lambda$, and $P_{\mathrm{R}}((x - e_i)^+ \mid x, \cdot) = \mu$, $\forall i \in [n]$, where b represents the chosen action at state x.

The objective of the operator is to find an optimal protecting policy β that minimizes the expected cumulative discounted cost $J(x; \beta)$ given initial state $X(0) = x$:

$$\begin{aligned} J^*(x) &:= \min_{\beta} J(x; \beta) \\ &:= \min_{\beta} \mathbb{E}\Big[\int_0^{\infty} e^{-\gamma t} C(X(t),\ B(t))dt \Big| X(0) = x, B(t) \\ &\sim \beta(X(t)),\ X(t + dt) \sim P_{\mathrm{R}}(\cdot \mid X(t), B(t))\Big], \end{aligned} \quad (1)$$

where $\gamma \in (0, 1)$ is the discounted factor, $B(t) \in \{\mathrm{NP}, \mathrm{P}\}$ denotes the action chosen by the operator at time t, and $C : \mathbb{Z}^n_{\geq 0} \times \{\mathrm{NP}, \mathrm{P}\} \to \mathbb{R}_{\geq 0}$ is the net cost rate defined as

$$C(\xi, \mathsf{b}) := ||\xi||_1 + c_b\mathbb{1}\{\mathsf{b} = \mathrm{P}\}.$$

Denote the optimal protecting policy as β^*, then

$$\beta^*(x) := \operatorname*{argmin}_{\beta} J(x; \beta), \quad \forall x \in \mathbb{Z}^n_{\geq 0}.$$

2.3 Security failures

When each job arrives, a malicious attacker can manipulate the routing, directing the job to any desired queue. For simplicity, we consider the attacker's best action (and thus the system operator's worst case) – sending to the longest queue, as shown in Fig. 1b. Attacking a job also incurs a technological cost rate, denoted by $c_a > 0$. The operator's action is akin to the reliability scenario. The only difference is in this security scenario, the operator is aware of a simultaneous strategic attacker.

We formulate the interaction between the attacker and the operator (called defender in this scenario) as an *infinite-horizon stochastic game*. The attacker selects a (possibly mixed) *Markov strategy* $\alpha : \mathbb{Z}^n_{\geq 0} \to \Delta(\{\mathrm{NA}, \mathrm{A}\})$ with $\Delta(\{\mathrm{NA}, \mathrm{A}\}) := \{(1-a, a) : a \in [0, 1]\}$ representing the probability distribution over the action set {not attack, attack}. The transition matrix $P_{\mathrm{S}} : \mathbb{Z}^n_{\geq 0} \times \{\mathrm{NA}, \mathrm{A}\} \times \{\mathrm{NP}, \mathrm{P}\} \mapsto \Delta\left(\mathbb{Z}^n_{\geq 0}\right)$ capturing the queuing dynamics of the attacker-defender game is given by $P_{\mathrm{S}}(\xi + e_{\max} \mid \xi, \mathsf{a}, \mathsf{b}) = \mathbb{1}\{\mathsf{a} = \mathrm{A}\}\mathbb{1}\{\mathsf{b} = \mathrm{NP}\}\lambda$, $P_{\mathrm{S}}(\xi + e_{\min} \mid \xi, \mathsf{a}, \mathsf{b}) = \mathbb{1}\{\mathsf{a} \neq \mathrm{A}\}\mathbb{1}\{\mathsf{b} \neq \mathrm{NP}\}\lambda$, $P_{\mathrm{S}}((\xi - e_i)^+ \mid \xi, \cdot, \cdot) = \mu$, $\forall i \in [n]$.

In this security game, the objective of the attacker is to maximize the expected cumulative discounted reward $V(x; \alpha, \beta)$ given the operator's Markov strategy β:

$$\begin{aligned} V^*_A(x; \beta) &:= \max_{\alpha} V(x; \alpha, \beta) \\ &:= \max_{\alpha}\ \mathbb{E}\Big[\int_0^{\infty} e^{-\gamma t} R(X(t), A(t), B(t))dt \Big| \\ &\qquad X(0) = x,\ A(t) \sim \alpha(X(t)),\ B(t) \sim \beta(X(t)), \\ &\qquad X(t + dt) \sim P_{\mathrm{S}}(\cdot \mid X(t), A(t), B(t))\Big], \end{aligned}$$

where $A(t) \in \{\mathrm{NA}, \mathrm{A}\}$ and $B(t) \in \{\mathrm{NP}, \mathrm{P}\}$ denote the actions chosen by the attacker and the defender respectively at time t, and $R : \mathbb{Z}^n_{\geq 0} \times \{\mathrm{NA}, \mathrm{A}\} \times \{\mathrm{NP}, \mathrm{P}\} \to \mathbb{R}$ is the net reward rate defined as

$$R(\xi, \mathsf{a}, \mathsf{b}) := ||\xi||_1 + c_b\mathbb{1}\{\mathsf{b} = \mathrm{P}\} - c_a\mathbb{1}\{\mathsf{a} = \mathrm{A}\}.$$

Here we model the attacker-defender game as a zero-sum game, which aligns with established security game literature (Alpcan and Başar, 2010). The attacker's reward comprises queuing attacking costs, along with a deduction for defending costs. This is motivated by the attacker's potential interest in maximizing the operator's total operating cost, akin to competitive motives in business contests. Similarly, the defender aims to minimize the expected cumulative discounted loss given the attacker's Markov strategy α:

$$V^*_B(x; \alpha) := \min_{\beta} V(x; \alpha, \beta).$$

In such an attacker-defender game, we define the Markov perfect equilibrium (α^*, β^*) as: for each state $x \in \mathbb{Z}^n_{\geq 0}$,

$$\alpha^*(x) = \operatorname*{argmax}_{\alpha} V(x; \alpha, \beta^*),\ \beta^*(x) = \operatorname*{argmin}_{\beta} V(x; \alpha^*, \beta).$$

3 Protection against reliability failures

In this section, we consider the design of operator's protecting policy from two aspects: stability and optimality.

It is well known that a parallel n-server system is stabilizable if and only if the demand is less than the total capacity, i.e., $\lambda < n\mu$. In the following results, we will see that even if this condition is met, in the absence of protection, reliability failures can still destabilize the queuing system, especially when the probability of failures is high and when the random faulty routing is highly heterogeneous; the following summarizes the above insights.

Proposition 1 *The unprotected n-server system with faulty probability a is stable if and only if*

$$\lambda < n\mu, \quad (2a)$$

$$ap_{max}\lambda < \mu. \quad (2b)$$

Furthermore, when the system is stable, we have the following upper bound of the long-time average number of jobs (denoted by $\bar{X}$):

$$\limsup_{t\to\infty} \frac{1}{t} \int\limits_{\tau=0}^{t} \mathbb{E}[||X(\tau)||_1]d\tau \leq \frac{\lambda + n\mu}{2\left(\mu - \max(ap_{\max}, \frac{1}{n})\lambda\right)}.$$

The next result provides a stability criterion for an n-server system with reliability failures and a given protecting policy. Its proof is presented in Section 3.1.

Theorem 1 (Stability under reliability failures) *Consider an n-server system with reliability failure probability $a > 0$. Suppose the operator selects a Markovian policy $\beta : \mathbb{Z}^n_{\geq 0} \to \Delta(\{\mathrm{NP}, \mathrm{P}\})$ with protection probability $b(x) := \beta(\mathrm{P} \mid x) \in [0, 1]$ at state $x \in \mathbb{Z}^n_{\geq 0}$. Then we have:*

(i) The system is stable if, for every non-diagonal state vector x, the protection probability $b(x)$ satisfies

$$b(x) > 1 - \frac{\mu||x||_1 - \lambda x_{\min}}{a\lambda \left(\sum\limits_{i=1}^{n} p_i x_i - x_{\min} \right)}. \qquad (3)$$

(ii) When (2a) *holds, there must exist a policy satisfying* (3)*. When* (2a)-(2b) *hold, every policy satisfies* (3).

(iii) If (3) *holds, the long-time average number of jobs in the system is upper-bounded by*

$$\bar{X} \leq \frac{\lambda + n\mu}{2c}, \qquad (4)$$

where

$$c = \min_{x \succ \mathbf{0}} \left\{ \mu - \lambda \frac{x_{\min}}{||x||_1} - a(1 - b(x))\lambda \frac{\sum_{i=1}^{n} p_i x_i - x_{\min}}{||x||_1} \right\}.$$

The next result implies a key finding: protection should be activated when queue lengths are more "imbalanced".

Theorem 2 (Optimal protecting policy) *Consider an n-server system subject to reliability failures. An optimal deterministic protecting policy β^* exists. This deterministic policy is also a threshold policy characterized by n threshold functions f_m $(m = 1, 2, \cdots, n)$ via*

$$b^*(x) := \beta^*(\mathrm{P} \mid x) = \mathbb{1}\left\{ \bigwedge\nolimits_{m=1}^{n} (f_m(x) > 0) \right\},$$

where for each $m = 1, 2, \cdots, n$,

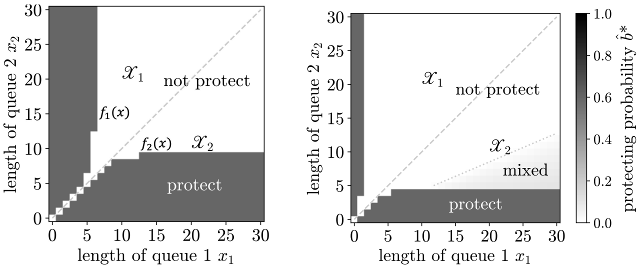
(i) the threshold function $f_m : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}$ partitions the polyhedron $\mathscr{X}_m = \{x \in \mathbb{Z}^n_{\geq 0} \mid x_i \geq x_m,\ \forall 1 \leq i \leq n\}$ into two subsets: $\{x \in \mathscr{X}_m \mid \beta^*(x) = NP\}$ and $\{x \in \mathscr{X}_m \mid \beta^*(x) = P\}$ by means of

$$b^*(x) = \mathbb{1}\{f_m(x) > 0\}, \quad \forall x \in \mathscr{X}_m;$$

(ii) within the polyhedron $\mathscr{X}_m$, the optimal protection probability $b^*(x)$ is monotonically non-decreasing (resp. non-increasing) in x_i $(\forall i \neq m)$ (resp. x_m) while other variables x_{-i} (resp. x_{-m}) are fixed.

Here (ii) supplements (i), demonstrating that the threshold functions characterize the degree of "imbalancedness". They partition the state space into $n + 1$ subsets: one "inner subset" with "balanced" states corresponding to the action "not protect", and the other n "outer subsets" with "imbalanced" states for action "protect". See the white and black areas in Fig. 2a. The concept "threshold function" has appeared in prior works (Bertsekas, 2012; Hajek, 1984; Stidham and Weber, 1993).

The rest of this section is devoted to the proofs, discussions, and numerical analysis of Theorem 1-2.



(a) Optimal (b) Stability-constrained optimal

Fig. 2. The characterization of the optimal protecting policy β^* and the stability-constrained optimal policy $\hat{\beta}^*$ for a two-server system ($p_1 = 0.1$, $p_2 = 0.9$, $\rho = 0.5$, $a = 0.9$).

3.1 Proof of Theorem 1

In this subsection, we provide a proof of the stability condition under the protected case (Theorem 1) and leave the proof of stability under the unprotected case (Proposition 1) to Appendix A.1. Both proofs use the following classical result (Meyn and Tweedie, 1993, Theorem 4.3):

Foster-Lyapunov drift criterion: Consider a countable-state continuous-time Markov chain X with state space S. Let $W : S \to \mathbb{R}_{\geq 0}$ be a qualified Lyapunov function, and let $\mathcal{L}$ denote the infinitesimal generator for W, with the drift $\mathcal{L}W$ given by

$$\mathcal{L}W(x) := \lim_{t \to 0} \frac{1}{t} \mathbb{E}[W(X(t)) \mid X(0) = x] - W(x).$$

For a non-negative function $f : S \to \mathbb{R}_{\geq 0}$, if there exists $c > 0$ and $d < \infty$ and a compact set $C \subset S$ such that for every $x \notin C$, the following drift condition holds:

$$\mathcal{L}W(x) \leq -cf(x) + d,$$

then for any initial condition $X(0) = x \in S$, we have

$$\limsup_{t \to \infty} \frac{1}{t} \int\limits_{\tau=0}^{t} \mathbb{E}[f(X(\tau))] d\tau \leq d/c.$$

In this paper, we care about *mean boundedness*, i.e., the upper bound of the long-time average number of jobs. Thus, when applying this theorem, we select $S = \mathbb{Z}^n_{\geq 0}$, $f(x) = ||x||_1$, and the quadratic Lyapunov function

$$W(x) = \frac{1}{2} \sum_{i=1}^{n} x_i^2. \qquad (5)$$

Proof of Theorem 1. (i) Apply the infinitesimal generator to the countable-state continuous-time MDP $\{X(t)\}_{t \geq 0}$ under the protecting policy β (denoted as $\mathcal{L}^\beta$) and the Lyapunov function W given by (5), we have

$$\begin{aligned} \mathcal{L}^\beta W(x) =& a(1 - b(x))\lambda \sum_{i=1}^{n} p_i \left(W(x + e_i) - W(x) \right) \\ &+ \left(1 - a(1 - b(x))\right) \lambda \left(W(x + e_{\min}) - W(x) \right) \\ &+ \mu \sum_{i=1}^{n} \mathbb{1}\{x_i > 0\} \left(W(x - e_i) - W(x) \right) \end{aligned}$$

$$=a(1-b(x))\frac{\lambda}{2}\sum_{i=1}^{n} p_i\left((x_i+1)^2-x_i^2\right)$$

$$+(1-a(1-b(x)))\frac{\lambda}{2}\left((x_{\min}+1)^2-x_{\min}^2\right)$$

$$+\frac{\mu}{2}\sum_{i=1}^{n}\mathbb{1}\{x_i>0\}\left((x_i-1)^2-x_i^2\right)$$

$$=a(1-b(x))\lambda\sum_{i=1}^{n}p_ix_i+(1-a(1-b(x)))\,\lambda x_{\min}$$

$$-\mu\sum_{i=1}^{n}x_i+\frac{1}{2}\lambda+\frac{1}{2}\sum_{i=1}^{n}\mathbb{1}\{x_i>0\}\mu$$

$$\leq a(1-b(x))\lambda\left(\sum_{i=1}^{n}p_ix_i-x_{\min}\right)+(\lambda x_{\min}$$

$$-\mu||x||_1)+\frac{1}{2}(\lambda+n\mu).$$

By (3) there exists constants

$$c=\min_{x\succ\mathbf{0}}\left\{\mu-\lambda\frac{x_{\min}}{||x||_1}-a(1-b(x))\lambda\frac{\sum_{i=1}^{n}p_ix_i-x_{\min}}{||x||_1}\right\}>0$$

and $d=\frac{1}{2}(\lambda+n\mu)$ such that

$$\mathcal{L}^{\beta}W(x)\leq -c||x||_1+d,\quad \forall x\in\mathbb{Z}^n_{\geq 0}. \tag{6}$$

(ii) When $\lambda<n\mu$, for every non-diagonal vector x, we have $\mu||x||_1-\lambda x_{\min}>\mu||x||_1-\lambda\frac{||x||_1}{n}=\left(\mu-\frac{\lambda}{n}\right)||x||_1>0$ and $\sum_{i=1}^{n}p_ix_i-x_{\min}>0$, then

$$1-\frac{\mu||x||_1-\lambda x_{\min}}{a\lambda\left(\sum\limits_{i=1}^{n}p_ix_i-x_{\min}\right)}<1.$$

Thus, $b(x)\equiv 1$ satisfies the stability condition (3) and $\beta(x)\equiv\mathrm{P}$ is a stabilizing policy that exists.

When $\max(ap_{\max},1/n)\lambda<\mu$, for every non-diagonal vector x, we have $a\lambda\sum_{i=1}^{n}p_ix_i+(1-a)\lambda x_{\min}\leq\max(ap_{\max},1/n)\lambda||x||_1<\mu||x||_1$, and then

$$1-\frac{\mu||x||_1-\lambda x_{\min}}{a\lambda\left(\sum\limits_{i=1}^{n}p_ix_i-x_{\min}\right)}<0.$$

Thus, every policy satisfies the stability criterion (3).

(iii) By Foster-Lyapunov criterion, the drift condition (6) implies the upper bound (4) and thus the stability. □

Theorem 1 provides a stability criterion for any protecting policy. This implies that the operator needs to choose a positive protection probability to stabilize the system at certain states. We will use stabilizing threshold probabilities given by (3) to obtain a stability-constrained optimal policy. See Section 3.3 and Appendix A.4.

3.2 Proof of Theorem 2

A standard way to solve the discounted infinite-horizon minimization problem (1) is to write down its HJB equation for optimality (Chang, 2004, Chapter 4):

$$0=\min_{\beta}\{||x||_1+c_bb(x)-\gamma J^*(x)+\mathcal{L}^{\beta}J^*(x)\}. \tag{7}$$

We can rewrite it in the following recurrence form:

$$\begin{aligned}(\gamma+\lambda+n\mu)J^*(x)=\min_{\beta}\Bigg\{&||x||_1+c_bb(x)+\\ &\mu\sum_{i=1}^{n}J^*((x-e_i)^+)+\lambda J^*(x+e_{\min})+(1-b(x))a\\ &\lambda\Big(\sum_{i=1}^{n}p_iJ^*(x+e_i)-J^*(x+e_{\min})\Big)\Bigg\}.\end{aligned} \tag{8}$$

The optimal protecting policy β^* is essentially the solution to (8). By a standard result of discrete-state finite-action MDP (Puterman, 2014, Theorem 6.2.10), an optimal deterministic stationary policy exists. Furthermore, in the no-failure scenario ($a=0$), the operator never needs to protect (i.e., $\forall x,\ \beta^*(x)=\mathrm{NP}$); and when all queue lengths are equal, i.e., $x_1=x_2=\cdots=x_n$, the operator deterministically deactivates the protection.

Let $\tilde{\lambda}=\lambda/(\gamma+\lambda+n\mu)$, $\tilde{\mu}=\mu/(\gamma+\lambda+n\mu)$, and $\tilde{J}^*(\cdot)=(\gamma+\lambda+n\mu)J^*(\cdot)$, then we can rewrite (8) as

$$\begin{aligned}\tilde{J}^*(x)=\min_{b\in\{0,1\}}\Bigg\{&||x||_1+c_bb+\tilde{\mu}\sum_{i=1}^{n}\tilde{J}^*((x-e_i)^+)\\ &+\tilde{\lambda}\tilde{J}^*(x+e_{\min})+(1-b)a\\ &\tilde{\lambda}\Big(\sum_{i=1}^{n}p_i\tilde{J}^*(x+e_i)-\tilde{J}^*(x+e_{\min})\Big)\Bigg\}\\ =:\min_{b\in\{0,1\}}\Big\{&c(x,b)+\sum_{x'}p(x'|x,b)\tilde{J}^*(x')\Big\}.\end{aligned} \tag{9}$$

By applying the DP recursion technique (Bertsekas, 2012, Chapter 4.6), we can demonstrate (i) the existence of threshold functions and (ii) the monotonicity of the optimal protection probability, i.e., $\forall x\in\mathbb{Z}^n_{\geq 0}$, if we let $m=\mathrm{argmin}_i\,x_i$, then

$$\begin{aligned}b^*(x+e_i)&\geq b^*(x),\quad \forall i\neq m\\ b^*(x-e_m)&\geq b^*(x).\end{aligned} \tag{10}$$

Now we prove (10). Let $\Delta^*(x)=\sum\limits_{i=1}^{n}p_i\tilde{J}^*(x+e_i)-\tilde{J}^*(x+e_m)$. Note that by the definition of β^* and Theorem 2(i), $b^*(x)=1$ if $\Delta^*(x)>\frac{c_b}{a\tilde{\lambda}}$ and $b^*(x)=0$ if $\Delta^*(x)<\frac{c_b}{a\tilde{\lambda}}$. Then the monotonicity of b^* is essentially the monotonicity of Δ^*. Thus, (10) is equivalent to

$$\begin{aligned}\Delta^*(x+e_i)&\geq\Delta^*(x),\quad \forall i\neq m\\ \Delta^*(x-e_m)&\geq\Delta^*(x).\end{aligned} \tag{11}$$

We defer the proof of (11) to Appendix A.2. The high-level idea is to use induction based on value iteration. □

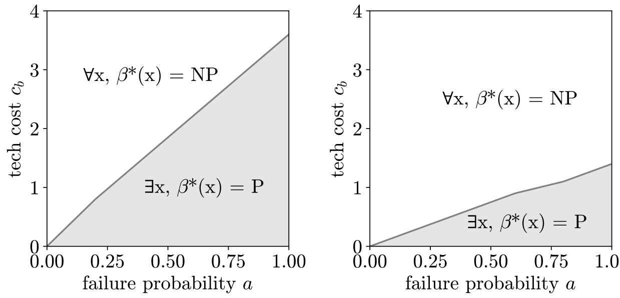
3.3 Numerical Analysis and Discussions

The optimal policy can be estimated using an algorithm called truncated policy iteration (TPI). See Algorithm 1 in Appendix A.4. It is adapted from the classic policy iteration algorithm (Sutton and Barto, 2018) and based on the following recursion form of the HJB equation (9):

$$\tilde{J}^{k+1}(x) = \min_b \left\{ c(x,b) + \gamma \sum_{x'} p(x'|x,b) \tilde{J}^k(x') \right\}. \quad (12)$$

Next, we conduct two numerical analysis. The first one is about the relationship between the incentive to protect and the system parameters. The second one compares the optimal policy with two benchmark policies: always protect and never protect.

In the first analysis, we explore the tipping points when the system operator starts protecting "riskier" states under the optimal policy β^*, i.e., $\exists x$ s.t. $\beta^*(x) = \mathrm{P}$, as failure probability a and technological cost c_b vary. Fig. 3 illustrates that the incentive to protect is non-decreasing in failure probability a, non-increasing in technological cost c_b and non-decreasing in *traffic intensity* (a.k.a. utilization ratio) $\rho = \lambda/(n\mu)$. In short, the operator has a higher incentive to protect when 1) the failure probability is higher; 2) the technological cost is lower; and 3) the traffic intensity is higher.



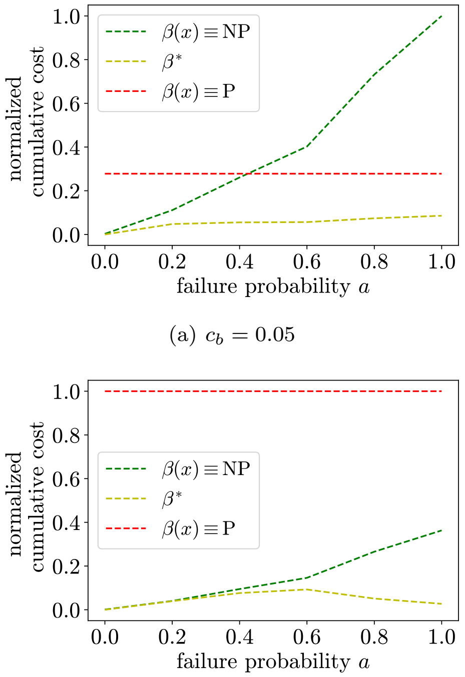
(a) $\rho = 0.8$

(b) $\rho = 0.2$

Fig. 3. The tipping points when the system operator starts to protect "riskier" states under the optimal policy as the failure probability and the technological cost change.

The second analysis involves running a Monte Carlo simulation on the cumulative discounted cost of the optimal policy and two benchmark policies under varying failure probabilities. The results show that the optimal policy β^* can significantly reduce the security risk, compared to two static policies: $\beta(x) \equiv \mathrm{P}$ (always protects) and $\beta(x) \equiv \mathrm{NP}$ (never protects). See Fig. 4 where the yellow curves are below the red and green curves. The cumulative discounted cost is calculated as the sum of the queuing costs and technological costs in the episode (50000s), normalized within the range of [0,1]. Note that under the policy $\beta(x) \equiv \mathrm{P}$, the job always joins the shortest queue regardless of the failure probability, rendering the cumulative discounted cost as a constant (red curve).

Last, we address the integration of stability and optimality considerations. Notably, the optimal policy may not always be stabilizing. For instance, the optimal policy under the parameters $p_1 = 0.1$, $p_2 = 0.9$, $\rho = 0.5$, $a = 0.9$ fails to meet the stability condition (3). To address this issue, we can select an optimal policy satisfying the stability condition (3) by solving a stability-constrained MDP (Zanon et al., 2022). This involves adding an additional constraint (3) to the optimal control problem (1).



(a) $c_b = 0.05$

(b) $c_b = 0.5$

Fig. 4. Comparison of the normalized cumulative discounted costs between the optimal policy and the static policies ($p_1 = 0.1$, $p_2 = 0.9$, $\rho = 0.8$).

We call the solution *stability-constrained optimal policy* and denote it as $\hat{\beta}^*$. This policy, unlike the optimal one, involves randomization over actions {P, NP} at some states, see Fig. 2b. Appendix A.4 gives a corresponding modification of the TPI algorithm. Additionally, we can examine the existence of a stabilizing policy and the stabilizability of the optimal policy using stability conditions (2a)-(2b) as follows:

- When (2a)-(2b) hold, i.e., $\max(ap_{\max}, 1/n)\lambda < \mu$, the optimal protecting policy is also stabilizing.
- When only (2a) holds, i.e., $\lambda/n < \mu \leq ap_{\max}\lambda$, the optimal protecting policy may not be stabilizing.
- When (2a) does not hold, i.e., $\lambda \geq \mu n$, no stabilizing protecting policy exists.

4 Defense against security failures

In this section, we analyze the attacker's attacking strategy and the system operator's defending strategy from two aspects: stability and game equilibrium.

The following criterion can be used for checking the stability of the n-server system under any state-dependent attacking and defending strategies:

Theorem 3 (Stability under security failures) *Consider an n-server system facing security failures. Suppose at each state $x \in \mathbb{Z}^n_{\geq 0}$, the attacker (resp. system operator) selects a state-dependent Markov strat-*

egy α (resp. β) with attack (resp. defense) probability $a(x) := \alpha(\mathrm{A} \mid x)$ (resp. $b(x) := \beta(\mathrm{P} \mid x)$). Then we have:

(i) The system is stable when the attack and defense probabilities $a(x)$ and $b(x)$ satisfy the following for every non-diagonal state vector x,

$$a(x)\,(1-b(x)) < \frac{\mu||x||_1 - \lambda x_{\min}}{\lambda(x_{\max} - x_{\min})}. \tag{13}$$

(ii) When $\lambda < n\mu$, there must exist a Markov strategy β with defense probability $b(x)$ satisfying (13).

(iii) Furthermore, if (13) holds, then the long-time average number of jobs is upper-bounded by

$$\bar{X} \leq \frac{\lambda + n\mu}{2c}, \tag{14}$$

where

$$c = \min_{x \succ \mathbf{0}} \left\{ \mu - \lambda \frac{x_{\min}}{||x||_1} - a(x)(1-b(x))\lambda \frac{x_{\max} - x_{\min}}{||x||_1} \right\}.$$

The next result characterizes the structure of Markov perfect equilibria: the defense probability is higher when queue lengths are more "imbalanced".

Theorem 4 (Markov perfect equilibrium) *The Markov perfect equilibrium (MPE) of the attacker-defender stochastic game exists, and the following holds:*

(i) MPE (α^*, β^*) is qualitatively different over the following three subsets of the state space $\mathbb{Z}^n_{\geq 0}$:
  - *(a) $S_1 = \{x \in \mathbb{Z}^n_{\geq 0} \mid (\alpha^*(x), \beta^*(x)) = (NA, NP)\}$; ("low risk")*
  - *(b) $S_2 = \{x \in \mathbb{Z}^n_{\geq 0} \mid (\alpha^*(x), \beta^*(x)) = (A, NP)\}$; ("medium risk")*
  - *(c) $S_3 = \{x \in \mathbb{Z}^n_{\geq 0} \mid (\alpha^*(x), \beta^*(x))$ is mixed$\}$. ("high risk")*

(ii) The boundaries between S_1 and S_2, as well as those between S_2 and S_3 are characterized by threshold functions g_{ij}, h_{ij} $(1 \leq i \neq j \leq n)$ as follows:

$$S_1 = \left\{ x \in \mathbb{Z}^n_{\geq 0} \middle| \bigwedge\nolimits_{1 \leq i \neq j \leq n} (g_{ij}(x) < 0) \right\},$$

$$S_2 = \left\{ x \in \mathbb{Z}^n_{\geq 0} \middle| \bigwedge\nolimits_{1 \leq i \neq j \leq n} (g_{ij}(x) > 0 \wedge h_{ij}(x) < 0) \right\},$$

$$S_3 = \left\{ x \in \mathbb{Z}^n_{\geq 0} \middle| \bigwedge\nolimits_{1 \leq i \neq j \leq n} (h_{ij}(x) > 0) \right\}.$$

For each $i, j = 1, 2, \cdots, n$ $(i \neq j)$,
  - *(a) $g_{ij}, h_{ij} : \mathbb{Z}^n_{\geq 0} \to \mathbb{R}$ separate the polyhedron $\mathscr{X}_{ij} = \{x \in \mathbb{Z}^n_{\geq 0} \mid x_i = x_{\max},\ x_j = x_{\min}\}$ into three subsets: $S_1 \cap \mathscr{X}_{ij}$, $S_2 \cap \mathscr{X}_{ij}$ and $S_3 \cap \mathscr{X}_{ij}$;*
  - *(b) state x has a lower (resp. higher) or equal security level than state $x + e_i$ (resp. $x + e_j$).*

The threshold functions here also characterize the degree of "imbalancedness". Intuitively, S_1–S_3 correspond to different security risk levels, leading to varied defense decisions: when queues are more "imbalanced", the security risk is higher, and the incentive to defend is higher. See Fig. 5 for a visualization. For the relationship between the security levels and system parameters (e.g., technological costs, traffic intensity), see Section 4.2.

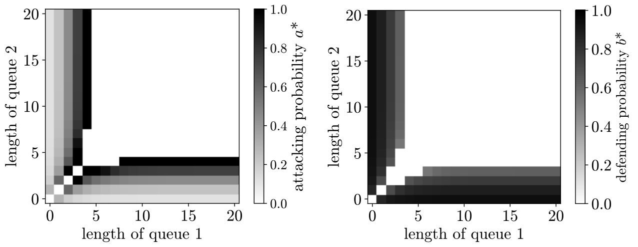


Fig. 5. The equilibrium attacking and defending strategies for a two-server system ($\rho = 0.5$, $c_a = 0.1$, $c_b = 0.2$).

The security game has four equilibrium regimes under different combinations of attack cost c_a and defense cost c_b; see Fig. 6. Each regime is labeled with corresponding subsets of Markov perfect equilibria and security levels.

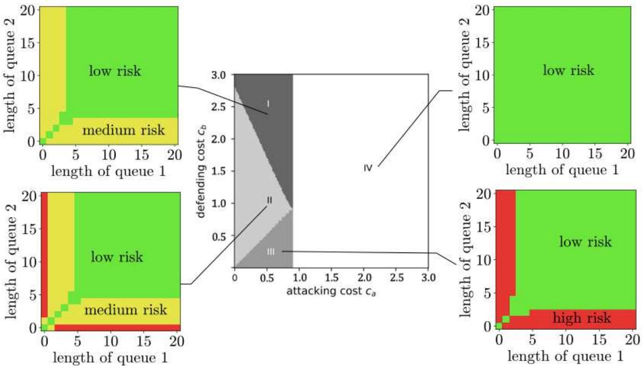


Fig. 6. Equilibrium regimes of the security game ($\rho = 0.8$).

The rest of this section is devoted to the proofs of Theorem 3-4, as well as an additional discussion on the computation and regimes of Markov perfect equilibria.

4.1 Proof of Theorem 3

(i) By applying the infinitesimal generator $\mathcal{L}^{\alpha,\beta}$ to the MDP $\{X(t)\}_{t \geq 0}$ under the attacking strategy α and the defending strategy β as well as the Lyapunov function (5), we have

$$\begin{aligned}
\mathcal{L}^{\alpha,\beta} W(x) = &(1 - a(x)(1-b(x)))\,\frac{\lambda}{2}\left((x_{\min}+1)^2 - x_{\min}^2\right) \\
&+ a(x)(1-b(x))\frac{\lambda}{2}\left((x_{\max}+1)^2 - x_i^2\right) \\
&+ \frac{\mu}{2}\sum_{i=1}^{n} \mathbb{1}\{x_i > 0\}\left((x_i - 1)^2 - x_i^2\right) \\
\leq &a(x)(1-b(x))\lambda(x_{\max} - x_{\min}) \\
&+ \lambda x_{\min} - \mu||x||_1 + \frac{1}{2}(\lambda + n\mu).
\end{aligned} \tag{15}$$

Hence, by (13) there exists $c = \min\limits_{x \succ \mathbf{0}} \{\mu - \lambda \frac{x_{\min}}{||x||_1} - a(x)(1 - b(x))\lambda \frac{x_{\max} - x_{\min}}{||x||_1})\} > 0$ and $d = \frac{1}{2}(\lambda + n\mu)$ such that

$$\mathcal{L}^{\alpha,\beta} W(x) \leq -c||x||_1 + d, \quad \forall x \in \mathbb{Z}^n_{\geq 0}.$$

(ii) When $\lambda < n\mu$, for every non-diagonal vector x, the inequalities $\mu||x||_1 \geq n\mu x_{\min} > \lambda x_{\min}$ and $x_{\max} > x_{\min}$ hold. This implies $\frac{\mu||x||_1 - \lambda x_{\min}}{\lambda(x_{\max} - x_{\min})} > 0$. Consequently, regardless of the attacking strategy, the defending strategy with $b(x) \equiv 1$ satisfies the stability condition (13).

(iii) By the Foster-Lyapunov criterion, the drift condition (15) implies the upper bound (14) and thus the stability of the system. □

4.2 Proof of Theorem 4

For the attacker-defender stochastic game, we first show the existence of Markov perfect equilibrium.

Proposition 2 *Markov perfect equilibrium* (α^*, β^*) *of the attacker-defender stochastic game always exists.*

Proof. Note that the state space $\mathbb{Z}^n_{\geq 0}$ is countable and the action space $\{0, 1\}$ is finite (and thus compact). By (Federgruen, 1978, Theorem 1), a MPE exists. □

Next, we discuss the derivation of Markov perfect equilibria. According to Shapley's extension on the minimax theorem for stochastic game (Shapley, 1953), the attacker and the defender have the same minimax value:

$$V^*_B(x; \alpha^*) = V^*_A(x; \beta^*) = V^*(x).$$

Thus, we only need to compute the minimax value V^* of the stochastic game. Similar to the derivation of (9), we obtain the following HJB equation of the minimax problem (letting $\tilde{V}^*(\cdot) = (\gamma + \lambda + n\mu)V^*(\cdot)$):

$$\begin{aligned}\tilde{V}^*(x) = \max_{\alpha} \min_{\beta} \Big\{ & ||x||_1 + c_b b(x) - c_a a(x) + \\ & \tilde{\mu} \sum_i \tilde{V}^*((x - e_i)^+) + \tilde{\lambda}\tilde{V}^*(x + e_{\min}) + a(x)(1 - b(x)) \\ & \tilde{\lambda}\Big(\tilde{V}^*(x + e_{\max}) - \tilde{V}^*(x + e_{\min})\Big)\Big\}. \qquad (16)\end{aligned}$$

For each state $x \in \mathbb{Z}^n_{\geq 0}$, let $\delta^*(x) = \tilde{\lambda}(\tilde{V}^*(x + e_{\max}) - \tilde{V}^*(x + e_{\min}))$ and build an auxiliary matrix game

$$\begin{aligned} M(x, \tilde{V}^*) = & \Big(||x||_1 + \tilde{\mu} \sum_{i=1}^{n} \tilde{V}^*((x - e_i)^+) + \tilde{\lambda}\tilde{V}^*(x + e_{\min})\Big) \\ & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & c_b \\ -c_a + \delta^*(x) & -c_a + c_b \end{bmatrix}. \qquad (17)\end{aligned}$$

Then $\delta^*(x)$ and $(\alpha^*(x), \beta^*(x))$ can be obtained by *Shapley-Snow method* (Shapley and Snow, 1952), a convenient algorithm for finding the minimax value and equilibrium strategies of any two-player zero-sum game.

Proof of Theorem 4(i). Consider the matrix game $M(x, \tilde{V}^*)$ defined as (17) where the attacker is the row player and the system operator is the column player. Based on the Shapley-Snow method, the equilibrium strategies $(\alpha^*(x), \beta^*(x))$ are in the following three cases depending on the relationship between $\delta^*(x)$ and the technological costs $c_a, c_b > 0$:

(a) When $\delta^*(x) \leq c_a$, it is obvious that $\alpha^*(x) =$ NA (i.e., $a^*(x) = 0$) is a dominant strategy. Then $c_b > 0$ implies $\beta^*(x) =$ NP (i.e., $b^*(x) = 0$). That is, the attacker has no incentive to attack, and thus the defender does not need to defend. At this pure strategy equilibrium, the security risk is low.

(b) When the defending cost c_b is higher then the attacking cost c_a, and $c_a < \delta^*(x) \leq c_b$, it is obvious that $\beta^*(x) =$ NP (i.e., $b^*(x) = 0$) is a dominant strategy. Then $c_b > -c_a + c_b$ implies $\alpha^*(x) =$ A (i.e., $a^*(x) = 1$). That is, the defender has no incentive to defend and consequently, the attacker prefers to attack. At this pure strategy equilibrium, the security risk is higher than in the first case but tolerable.

(c) When $\delta^*(x) > \max\{c_a, c_b\} > 0$, no saddle point exists. Then both players consider mixed strategies with $a^*(x) = \frac{c_b}{\delta^*(x)}$, $b^*(x) = 1 - \frac{c_a}{\delta^*(x)}$. Particularly, the operator needs to select positive protecting probability, and now the security risk is high.

The above three cases correspond to the three subsets of states. Note that the subset S_2 is empty when $c_a > c_b$. □

From the above proof, we observe that for fixed c_a and c_b, the security risk level is higher when δ^* is larger. Then as in the proof of Theorem 2, we use the equivalence of the monotonicity of security risk levels and the monotonicity of δ^* to show the threshold property of the equilibria.

Proof of Theorem 4(ii). For any $x \in \mathbb{Z}^n_{\geq 0}$, let $l = \operatorname{argmax}_i x_i$, $m = \operatorname{argmin}_i x_i$. Then $\delta^*(x) = \tilde{\lambda}(\tilde{V}^*(x + e_l) - \tilde{V}^*(x + e_m))$. Since the monotonicity of the security risk levels of the states is equivalent to the monotonicity of δ^*, and implies the existence of the threshold functions, it is sufficient to show that δ^* is monotonically non-decreasing (resp. non-increasing) in the largest variable $x_{\max}$ (resp. the smallest variable $x_{\min}$) when other variables are fixed; that is,

$$\delta^*(x + e_l) \geq \delta^*(x), \quad \delta^*(x - e_m) \geq \delta^*(x). \qquad (18)$$

The proof of (18) also uses induction based on value iteration and can be found in Appendix A.3. □

4.3 Equilibrium analysis and discussions

First, we develop an adaptation of Shapley's algorithm (Shapley, 1953; Alpcan and Başar, 2010)to compute the minimax value V^* and the equilibrium (α^*, β^*) based on the DP recursion of HJB equation (16). See Algorithm 2 in Appendix 2. In each iteration, let $\delta(x) = \tilde{\lambda}(\tilde{V}(x + e_{\max}) - \tilde{V}(x + e_{\min}))$ and build an auxiliary matrix game $M(x, \tilde{V})$ as in (17); then update $\tilde{V}(x)$ with the minimax value $val(M)$ given by the Shapley-Snow method:

- when $\delta(x) \leq c_a$, $val(M) = ||x||_1 + \tilde{\mu} \sum_{i=1}^{n} \tilde{V}((x - e_i)^+) + \tilde{\lambda}\tilde{V}(x + e_{\min})$;
- when $c_a < \delta(x) \leq c_b$, $val(M) = ||x||_1 - c_a + \tilde{\mu} \sum\limits_{i=1}^{n} \tilde{V}((x - e_i)^+) + \tilde{\lambda}\tilde{V}(x + e_{\max})$;

- when $\delta(x) > \max\{c_a, c_b\}$, $val(M) = ||x||_1{+}c_b + \tilde{\mu}\sum_{i=1}^n \tilde{V}((x-e_i)^+) + \tilde{\lambda}\tilde{V}(x+e_{\min}) - \frac{c_a c_b}{\delta(x)}$.

When $\tilde{V}(x)$ converges to $\tilde{V}^*(x)$, we again use Shapley-Snow method to solve the matrix game $M(x, \tilde{V}^*)$ and obtain the estimation of the equilibrium $(\alpha^*(x), \beta^*(x))$.

Next, we discuss the existence of different security levels under different combinations of c_a and c_b. We have seen that no medium-risk state exists when $c_a > c_b$. In Fig. 6, various regimes correspond to particular combinations of security levels. Under large attacking costs, the attacker has no incentive to attack, then only the low-risk states exist (see regime IV). When the attacking cost goes smaller but still greater than the defending cost ($c_a > c_b$), not only the low-risk states but also the high-risk states exist (see regime III) since the attacker has less incentive to attack. As the defending cost increases to be greater than the attacking cost ($c_b > c_a$), the defender has less incentive to defend, and now all risk levels including the medium risk exist (see regime II).

Last, we remark that deriving a stability-constrained equilibrium is not reasonable as constraints can only be imposed on the system operator, not the attacker. Nevertheless, we can derive stability-constrained best responses for the operator, given the attacker's strategy.

5 Concluding Remarks

In this work, we address reliability and security concerns in dynamic routing and propose cost-efficient protection/defense advice for service system operators. Our theoretical results can provide insights for real-world applications like vehicle navigation, signal-free intersection control, flight dispatch, and data packet routing.

Future directions include 1) detailed analysis of stability-constrained optimal policies/best responses; 2) extension to general queuing networks, note that stability for renewal arrival processes and general service times can be established using fluid model techniques (Dai and Meyn, 1995); 3) design of practical near-optimal heuristic policies and analysis of optimality gaps; and 4) design of efficient algorithms for numerous parallel servers.

Acknowledgements

We appreciate discussions with Yu Tang, Xi Xiong, Haoran Su, Ziv Scully, Siddhartha Banerjee, Zhengyuan Zhou, Manxi Wu, Dorothy Ng, and Nairen Cao, as well as the invaluable feedback and comments provided by the reviewers on our manuscript.

A Appendices

A.1 Proof of Proposition 1

Here we provide the proof of the stability condition for an unprotected system using standard results on Poisson process subdivision and the generalized join the shortest queue systems (Foley and McDonald, 2001, Theorem 1).

Stability of generalized JSQ: Let $N = \{1, 2, \cdots, n\}$ be the set of n exponential servers. For each nonempty subset $S \subset N$, define the traffic intensity on S as

$$\rho_S := \frac{\sum_{S' \subset S} \lambda_{S'}}{\mu_S} = \frac{\sum_{S' \subset S} \lambda_{S'}}{\sum_{i \in S} \mu_i}.$$

Let $\rho_{\max} := \max_{S \subset N} \rho_S$ be the traffic intensity of the most heavily loaded subset. The generalized join the shortest queue system is stable if and only if $\rho_{\max} < 1$.

Proof of proposition 1. The unprotected n-queue system has $n+1$ classes of jobs. The i-th class enters server i as a Poisson process of rate $ap_i\lambda$ ($1 \le i \le n$). The $(n+1)$-th class enters the n-queue system as a Poisson process of rate $(1-a)\lambda$; when a job of this class arrives, the job joins the shortest queue. According to the above theorem, the $(n+1)$-class, n-server system is stable if and only if

$$\rho_{\max} = \max\left(\max_{1 \le i \le n} ap_i\lambda/\mu, \lambda/(n\mu)\right) < 1,$$

which is equivalent to (2a)-(2b).

By applying the infinitesimal generator to the same Lyapunov function (5) we have

$$\begin{aligned}
\mathcal{L}W(x) =& a\frac{\lambda}{2}\sum_{i=1}^n p_i\left((x_i+1)^2 - x_i^2\right) \\
&+ (1-a)\frac{\lambda}{2}\left((x_{\min}+1)^2 - x_{\min}^2\right) \\
&+ \frac{\mu}{2}\sum_{i=1}^n \mathbb{1}\{x_i > 0\}\left((x_i-1)^2 - x_i^2\right) \\
=& a\lambda\sum_{i=1}^n p_i x_i + (1-a)\lambda x_{\min} - \mu\sum_{i=1}^n x_i \\
&+ \frac{\lambda}{2} + \frac{\mu}{2}\sum_{i=1}^n \mathbb{1}\{x_i > 0\} \\
\le& \left(\max(ap_{\max}, 1/n)\lambda - \mu\right)||x||_1 + \frac{1}{2}(\lambda + n\mu).
\end{aligned}$$

Hence, by (2a)–(2b) there exists a constant $c = \mu - \max(ap_{\max}, 1/n)\lambda > 0$ and $d = \frac{1}{2}(\lambda + n\mu)$ such that

$$\mathcal{L}W(x) \le -c||x||_1 + d, \quad \forall x \in \mathbb{Z}^n_{\ge 0}.$$

By the Foster-Lyapunov drift criterion, this drift condition implies the upper bound and thus the stability. □

A.2 Induction part of Theorem 2

In this subsection, we continue the proof of Theorem 2 by showing (11) using the DP recursion technique.

Let $\Delta^k(x) = \sum_{i=1}^n p_i \tilde{J}^k(x+e_i) - \tilde{J}^k(x+e_m)$, it is sufficient to show for all $k \in \mathbb{N}$,

$$\begin{aligned}
&\Delta^k(x+e_i) \ge \Delta^k(x), \quad \forall i \ne m \\
&\Delta^k(x-e_m) \ge \Delta^k(x).
\end{aligned}$$

One can verify that the above hold for $k = 0, 1, 2$. Here we consider multiple base cases to avoid reaching trivial conclusions, say all inequalities are just equalities.

Now we show the inductive step. Let $f^k(x) = \sum_{i=1}^n p_i \min\{c_b, a\tilde{\lambda}\Delta^k(x+e_i)\} - \min\{c_b, a\tilde{\lambda}\Delta^k(x+e_m)\}$. Then according to the recursion (12), we have $\forall j \neq m$,

$$
\begin{aligned}
&\Delta^{k+1}(x+e_j) - \Delta^{k+1}(x) \\
=&\tilde{\mu}\sum_{i=1}^n [\Delta^k((x+e_j-e_i)^+) - \Delta^k((x-e_i)^+)] && \text{(A.1)}\\
&+\tilde{\lambda}[\Delta^k(x+e_j+e_m) - \Delta^k(x+e_m)] && \text{(A.2)}\\
&+f^k(x+e_j) - f^k(x), && \text{(A.3)}\\
&\Delta^{k+1}(x-e_m) - \Delta^{k+1}(x) \\
=&\tilde{\mu}\sum_{i=1}^n [\Delta^k((x-e_m-e_i)^+) - \Delta^k((x-e_i)^+)] && \text{(A.4)}\\
&+\tilde{\lambda}[\Delta^k(x) - \Delta^k(x+e_m)] && \text{(A.5)}\\
&+f^k(x-e_m) - f^k(x). && \text{(A.6)}
\end{aligned}
$$

Based on the induction hypothesis, we have $\forall j \neq m$,

$$
\begin{aligned}
\Delta^k((x+e_j-e_i)^+) &\geq \Delta^k((x-e_i)^+), && \text{(A.7)}\\
\Delta^k(x+e_j+e_m) &\geq \Delta^k(x+e_m), && \text{(A.8)}\\
\Delta^k(x+e_j+e_i) &\geq \Delta^k(x+e_i), && \text{(A.9)}\\
\Delta^k((x-e_m-e_i)^+) &\geq \Delta^k((x-e_i)^+), && \text{(A.10)}\\
\Delta^k(x) &\geq \Delta^k(x+e_m), && \text{(A.11)}\\
\Delta^k(x-e_m+e_i) &\geq \Delta^k(x+e_i). && \text{(A.12)}
\end{aligned}
$$

Then (A.7) and (A.10) naturally give rise to (A.1) ≥ 0 and (A.4) ≥ 0 respectively. We can also use (A.8)-(A.9) to discuss the possibilities of (A.3) under the min operations and establish (A.2) + (A.3) ≥ 0. For example, when $\Delta^k(x+e_j+e_i) \geq \Delta^k(x+e_i) \geq \frac{c_b}{a\tilde{\lambda}}$ and $\Delta^k(x+e_j+e_m) \geq \Delta^k(x+e_m) \geq \frac{c_b}{a\tilde{\lambda}}$, we have (A.2) + (A.3) $= a\tilde{\lambda}\sum_i p_i \left(\Delta^k(x+e_j+e_i) - \Delta^k(x+e_i)\right) + (1-a)\tilde{\lambda}\left(\Delta^k(x+e_j+e_m) - \Delta^k(x+e_m)\right) \geq 0$. Similarly, we can derive (A.5) + (A.6) ≥ 0 using (A.11)-(A.12).

Thus, we can conclude that (A.1) + (A.2) + (A.3) ≥ 0 and (A.4) + (A.5) + (A.6) ≥ 0, which yield

$$
\begin{aligned}
\Delta^{k+1}(x+e_j) &\geq \Delta^{k+1}(x), \quad \forall j \neq m\\
\Delta^{k+1}(x-e_m) &\geq \Delta^{k+1}(x).
\end{aligned}
$$

A.3 Induction part of Theorem 4

In this subsection, we continue the proof of Theorem 4 by showing (18) using the DP recursion technique.

Let $\delta^k(x) = \tilde{\lambda}(\tilde{V}^k(x+e_l) - \tilde{V}^k(x+e_m))$, it is sufficient to show for all $k \in \mathbb{N}$,

$$\delta^k(x+e_l) \geq \delta^k(x), \quad \delta^k(x-e_m) \geq \delta^k(x).$$

For the base cases, one can verify that the above inequalities hold for $k = 0, 1, 2$.

Now we show the inductive step. According to the value iteration form of formula (16),

$$
\begin{aligned}
&\delta^{k+1}(x+e_l) - \delta^{k+1}(x)\\
=&\tilde{\mu}[\delta^k(x) - \delta^k((x-e_l)^+)]\\
&+\tilde{\mu}[\delta^k((x+e_l-e_m)^+) - \delta^k((x-e_m)^+)]\\
&+\tilde{\lambda}[\delta^k(x+e_l+e_m) - \delta^k(x+e_m)] + g^k(x+2e_l)\\
&-g^k(x+e_l+e_m) - g^k(x+e_l) + g^k(x+e_m),
\end{aligned}
$$

where

$$g^k(x) = \max\left\{0, \min\left\{\delta^k(x) - c_a, c_b - \frac{c_a c_b}{\delta^k(x)}\right\}\right\}.$$

Note that based on the induction hypothesis, we have

$$\delta^k((x+e_l-e_m)^+) \geq \delta^k((x-e_m)^+) \geq \delta^k(x) \geq \delta^k((x-e_l)^+),$$

$$\delta^k(x+2e_l) \geq \delta^k(x+e_l) \geq \delta^k(x+e_l+e_m) \geq \delta^k(x+e_m).$$

Then we can conclude that $\delta^{k+1}(x+e_l) \geq \delta^{k+1}(x)$ and prove $\delta^{k+1}(x-e_m) \geq \delta^{k+1}(x)$ in a similar way.

A.4 Truncated policy iteration

In this subsection, we present the truncated policy iteration algorithm (Algorithm 1) for estimating stability-constrained optimal policy. This algorithm is adapted from the classic policy iteration algorithm (Sutton and Barto, 2018) by combining the stability condition (3). Since the original state space is countably infinite, here we set a boundary to make the state space finite so that the algorithm can terminate in finite steps.

A.5 Adapted Shapley's algorithm

In this subsection, we present the adapted Shapley's algorithm (Algorithm 2) for computing the minimax value and equilibrium strategies of the attacker-defender stochastic game. In each iteration, it builds an auxiliary matrix game and obtains the minimax value using the Shapley-Snow method (Shapley and Snow, 1952).

References

Al-Kahtani, M.S. (2012). Survey on security attacks in vehicular ad hoc networks (vanets). In *2012 6th international conference on signal processing and communication systems*, 1–9. IEEE.

Alpcan, T. and Başar, T. (2010). *Network security: A decision and game-theoretic approach.* Cambridge University Press.

[3] For computing the stability-constrained optimal policy, we can modify the algorithm as follows. In the initialization stage, for every x, if $x_1 = \cdots = x_n$ then continue, else set $\theta(x) = \max\left(1 - \frac{\mu||x||_1 - \tilde{\lambda}x_{\min}}{a\tilde{\lambda}\left(\sum_{i=1}^n p_i x_i - x_{\min}\right)}, 0\right)$. In the policy improvement stage, substitute $b(x) = 0$ with $b(x) = \theta(x)$.

Algorithm 1 Truncated policy iteration for estimating optimal policy $\beta \approx \beta^*$ (continuing)

```
Parameters: small $\epsilon > 0$, queue length upper bound $B$
Input: arrival rate $\lambda$, service rate $\mu$, number of servers $n$, discounted factor $\gamma$, protection cost rate $c_b$, random routing probabilities $p_1, p_2, \cdots, p_n$
Initialization: set $J(x) \in \mathbb{R}_{\geq 0}$ and $b(x) \in \{0,1\}$ arbitrarily (e.g., $J(x) = 0$, $b(x) = 0$) for all $x \in \mathcal{X} = \{0, 1, \cdots, B\}^n$

$\tilde{\lambda} \leftarrow \lambda/(\gamma + \lambda + n\mu)$,   $\tilde{\mu} \leftarrow \mu/(\gamma + \lambda + n\mu)$
repeat
    # Policy evaluation
    repeat
        $\Delta \leftarrow 0$
        foreach $x \in \mathcal{X}$ do
            $v \leftarrow J(x)$
            Compute $c(x, b)$ and $p(\cdot|x, b)$ based on (9)
            $J(x) \leftarrow \min_{b \in \{0,1\}} \{c(x, b) + \sum_{x'} p(x'|x, b)J(x')\}$
            $\Delta \leftarrow \max(\Delta, |v - J(x)|)$
        end
    until $\Delta < \epsilon$;
    # Policy improvement
    $stable \leftarrow True$
    foreach $x \in \mathcal{X}$ do
        old-action$\leftarrow b(x)$
        if $a\tilde{\lambda} \left(\sum_{i=1}^{n} p_i J(x + e_i) - J(x + e_{\min})\right) > c_b$ then
            $b(x) = 1$ [3]
        end
        else
            $b(x) = 0$
        end
        if old-action$\neq b(x)$ then
            $stable \leftarrow False$
        end
    end
until $stable = True$;
Output $\beta = (1 - b)\mathrm{NP} + b\mathrm{P} \approx \beta^*$
```

Algorithm 2 Adapted Shapley's algorithm for estimating equilibrium strategies $\beta \approx \beta^*$, $\alpha \approx \alpha^*$ (continuing)

```
Parameters: small $\epsilon > 0$, queue length upper bound $B$
Input: arrival rate $\lambda$, service rate $\mu$, number of servers $n$, discounted factor $\gamma$, protection cost rate $c_b$, attack cost rate $c_a$
Initialization set $V(x) \in \mathbb{R}_{\geq 0}$ arbitrarily (e.g., $V(x) = 0$) for all $x \in \mathcal{X} = \{0, 1, \cdots, B\}^n$

$\tilde{\lambda} \leftarrow \lambda/(\gamma + \lambda + n\mu)$,   $\tilde{\mu} \leftarrow \mu/(\gamma + \lambda + n\mu)$
repeat
    $\Delta \leftarrow 0$
    foreach $x \in \mathcal{X}$ do
        $v \leftarrow V(x)$
        $\delta(x) \leftarrow \tilde{\lambda}(V(x + e_{\max}) - V(x + e_{\min}))$
        Build auxiliary matrix game $M(x, V)$ as in (17)
        Compute the minimax value $val(M)$ by the Shapley-Snow method described in Section 4.3
        $V(x) \leftarrow val(M)$
        $\Delta \leftarrow |v - V(x)|$
    end
until $\Delta < \epsilon$;
foreach $x \in \mathcal{X}$ do
    Build auxiliary matrix game $M(x, V)$ as in (17)
    Compute the equilibrium strategies $\alpha$ and $\beta$ by the Shapley-Snow method described in Section 4.2
end
Output $\beta \approx \beta^*$, $\alpha \approx \alpha^*$
```

Barrère, M., Hankin, C., Nicolaou, N., Eliades, D.G., and Parisini, T. (2020). Measuring cyber-physical security in industrial control systems via minimum-effort attack strategies. *Journal of information security and applications*, 52, 102471.

Bertsekas, D. (2012). Dynamic programming and optimal control, vol II: Approximate dynamic programming.

Beutler, F.J. and Teneketzis, D. (1989). Routing in queueing networks under imperfect information: Stochastic dominance and thresholds. *Stochastics: An International Journal of Probability and Stochastic Processes*, 26(2), 81–100.

Bohacek, S., Hespanha, J., Lee, J., Lim, C., and Obraczka, K. (2007). Game theoretic stochastic routing for fault tolerance and security in computer networks. *IEEE transactions on parallel and distributed systems*, 18(9), 1227–1240.

Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., Sastry, S., et al. (2009). Challenges for securing cyber physical systems. In *Workshop on future directions in cyber-physical systems security*, volume 5. Citeseer.

Chang, F.R. (2004). *Stochastic optimization in continuous time*. Cambridge University Press.

Dai, J.G. and Meyn, S.P. (1995). Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Transactions on Automatic Control*, 40(11), 1889–1904.

De Persis, C. and Tesi, P. (2015). Input-to-state stabilizing control under denial-of-service. *IEEE Transactions on Automatic Control*, 60(11), 2930–2944.

Ephremides, A., Varaiya, P., and Walrand, J. (1980). A simple dynamic routing problem. *IEEE transactions on Automatic Control*, 25(4), 690–693.

Eryilmaz, A. and Srikant, R. (2007). Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *IEEE/ACM Transactions on Networking (TON)*, 15(6), 1333–1344.

Eschenfeldt, P. and Gamarnik, D. (2018). Join the shortest queue with many servers. the heavy-traffic asymptotics. *Mathematics of Operations Research*, 43(3), 867–886.

Federgruen, A. (1978). On n-person stochastic games by denumerable state space. *Advances in Applied Probability*, 10(2), 452–471.

Feng, Y., Huang, S.E., Wong, W., Chen, Q.A., Mao, Z.M., and Liu, H.X. (2022). On the cybersecurity of traffic signal control system with connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*.

Foley, R.D. and McDonald, D.R. (2001). Join the shortest queue: stability and exact asymptotics. *The Annals of Applied Probability*, 11(3), 569–607.

Fraile, F., Tagawa, T., Poler, R., and Ortiz, A. (2018). Trustworthy industrial iot gateways for interoperability platforms and ecosystems. *IEEE Internet of Things Journal*, 5(6), 4506–4514.

Govil, M.K. and Fu, M.C. (1999). Queueing theory in manufacturing: A survey. *Journal of manufacturing systems*, 18(3), 214–240.

Gravé-Lazi, L. (2014). Technion students find way to hack waze, create fake traffic jams. *The Jerusalem Post Available at: https://www.jpost.com/enviro-tech/technion-students-find-way-to-hack-waze-create-fake-traffic-jams-346377.*

Gupta, V., Balter, M.H., Sigman, K., and Whitt, W. (2007). Analysis of join-the-shortest-queue routing for web server farms. *Performance Evaluation*, 64(9-12), 1062–1081.

Hajek, B. (1984). Optimal control of two interacting service stations. *IEEE transactions on automatic control*, 29(6), 491–499.

Halfin, S. (1985). The shortest queue problem. *Journal of Applied Probability*, 22(4), 865–878.

Jin, L. and Amin, S. (2018). Stability of fluid queueing systems with parallel servers and stochastic capacities. *IEEE Transactions on Automatic Control*, 63(11), 3948–3955.

Knessl, C., Matkowsky, B., Schuss, Z., and Tier, C. (1986). Two parallel queues with dynamic routing. *IEEE transactions on communications*, 34(12), 1170–1175.

Kumar, P. and Meyn, S.P. (1995). Stability of queueing networks and scheduling policies. *IEEE Transactions on Automatic Control*, 40(2), 251–260.

Kuri, J. and Kumar, A. (1995). Optimal control of arrivals to queues with delayed queue length information. *IEEE Transactions on Automatic Control*, 40(8), 1444–1450.

Laszka, A., Abbas, W., Vorobeychik, Y., and Koutsoukos, X. (2019). Detection and mitigation of attacks on transportation networks as a multi-stage security game. *Computers & Security*, 87, 101576.

Manshaei, M.H., Zhu, Q., Alpcan, T., Bacşar, T., and Hubaux, J.P. (2013). Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3), 25.

Mehdian, S., Zhou, Z., and Bambos, N. (2017). Join-the-shortest-queue scheduling with delay. In *2017 American Control Conference (ACC)*, 1747–1752. IEEE.

Meyn, S.P. and Tweedie, R.L. (1993). Stability of markovian processes iii: Foster–lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3), 518–548.

Ouyang, Y. and Teneketzis, D. (2015). Signaling for decentralized routing in a queueing network. *Annals of Operations Research*, 1–39.

Puterman, M.L. (2014). *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons.

Sakiz, F. and Sen, S. (2017). A survey of attacks and detection mechanisms on intelligent transportation systems: Vanets and iov. *Ad Hoc Networks*, 61, 33–50.

Shapley, L.S. (1953). Stochastic games. *Proceedings of the national academy of sciences*, 39(10), 1095–1100.

Shapley, L.S. and Snow, R. (1952). Basic solutions of discrete games. *Contributions to the Theory of Games*, 1, 27–35.

Stidham, S. and Weber, R. (1993). A survey of markov decision models for control of networks of queues. *Queueing systems*, 13(1), 291–314.

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction.* MIT press.

Tang, Y., Wen, Y., and Jin, L. (2020). Security risk analysis of the shorter-queue routing policy for two symmetric servers. In *2020 American Control Conference (ACC)*, 5090–5095. IEEE.

Wang, Y., Lin, C., Li, Q.L., and Fang, Y. (2007). A queueing analysis for the denial of service (dos) attacks in computer networks. *Computer Networks*, 51(12), 3564–3573.

Xie, Q. and Jin, L. (2020). Resilience of dynamic routing in the face of recurrent and random sensing faults. In *2020 American Control Conference (ACC)*, 1173–1178. IEEE.

Xie, Q. and Jin, L. (2022). Stabilizing queuing networks with model data-independent control. *IEEE Transactions on Control of Network Systems.*

Zanon, M., Gros, S., and Palladino, M. (2022). Stability-constrained markov decision processes using mpc. *Automatica*, 143, 110399.