Quickest Detection of Adversarial Attacks against Correlated Equilibria

Kiarash Kazari[1], **Aris Kanellopoulos** [1], **György Dán**[1]

[1]Division of Network and Systems Engineering, School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Stockholm, Sweden
kkazari@kth.se, arisk@kth.se, gyuri@kth.se

Abstract

We consider correlated equilibria in strategic games in an adversarial environment, where an adversary can compromise the public signal used by the players for choosing their strategies, while players aim at detecting a potential attack as soon as possible to avoid loss of utility. We model the interaction between the adversary and the players as a zero-sum game and we derive the maxmin strategies for both the defender and the attacker using the framework of quickest change detection. We define a class of adversarial strategies that achieve the optimal trade-off between attack impact and attack detectability and show that a generalized CUSUM scheme is asymptotically optimal for the detection of the attacks. Our numerical results on the Sioux-Falls benchmark traffic routing game show that the proposed detection scheme can effectively limit the utility loss by a potential adversary.

1 Introduction

Correlated Equilibrium (CE) proposed by (Aumann 1987) is a solution concept in game theory that extends Nash Equilibria (NE) to correlated strategies among players. Unlike NE, where players choose their strategies independently, a CE represents a distribution over the space of joint actions of all players.

Compared to NE, CE offers a more efficient and practical framework for analyzing the strategic behaviour of players. Firstly, CE accommodates scenarios that result in a higher utilitarian social welfare than what can be achieved in any NE (Duffy and Feltovich 2010; Roughgarden and Tardos 2002). Additionally, CE are computationally more tractable. As shown by (Chen and Deng 2006; Daskalakis, Goldberg, and Papadimitriou 2009), computing a NE is a PPAD-complete problem even for two-player games. In contrast, a CE can be computed in polynomial time by solving a linear program (Papadimitriou and Roughgarden 2008). Moreover, there are efficient online learning dynamics that converge to the set of correlated equilibria (Cesa-Bianchi and Lugosi 2006; Anagnostides et al. 2022), or to the set of coarse correlated equilibria (Daskalakis, Fishelson, and Golowich 2021; Sessa et al. 2019), a related but weaker solution concept.

In a CE, coordination is facilitated by an external source of information, also called a *mediator*, which helps players align their strategies. The mediator samples from a known joint distribution over the players' actions and privately communicates each player's component without revealing the others' components. If the recommendation signals come from a CE, then no player has an interest to deviate from the mediator's suggested action, provided that the other players adhere to the recommendation. An example for a mediator is a map server in traffic routing, which recommends paths to the travelers (Ning and Du 2023). Another example is recommendations issued by stock analysts for the investors in financial markets (Leung 2007).

A fundamental assumption for the implementation of a CE in a game is the trustworthiness of the mediator. Nonetheless, the mediator could be compromised by an adversary, with the intention to cause damage to one or more players through manipulating the signal. For instance, routing software that recommends paths to users can be vulnerable to cyber attacks. Under attack, the distribution of the recommended action profiles would deviate from the original CE. From a player's perspective, this deviation must be detected promptly, as the assumption of following recommendations is based on the distribution being known and stable. A change can impact the utilities of the players, and upon detection, players may decide to adjust their strategies accordingly. In other words, timely attack detection is crucial for ensuring situational awareness, enabling the implementation of further countermeasures after detection.

In this paper, we demonstrate the vulnerability of the mediator's shared recommendation signal to adversarial attacks and propose a detection scheme to counteract them. We focus on the repeated play of a a CE in a strategic-form game, where an adversary manipulates the recommendation signals sent to the players by the mediator. Our key contributions are as follows:

- We model the interaction between the adversary and a victim player affected by the attack as a zero-sum game; the victim's objective is to detect the attack promptly while keeping the probability of false alarms low.
- We analyze the asymptotic maxmin strategies of the adversary and the defender as the cost of false alarms approaches infinity. This analysis introduces a family of distributions that enables the adversary to achieve the optimal trade-off between impact and detectability. For the defender, the optimal detection strategy is derived using

the framework of quickest change detection, where the goal is to identify an optimal stopping rule.

- We show the effectiveness of the proposed detection scheme in limiting the impact of the considered attacks in a tabular game as well as in a traffic routing game on the Sioux-Falls network benchmark.

To our knowledge, the concept of adversarial attacks on correlated equilibria is unexplored in the literature. Closest to our work in the game theory literature are (Roth 2008; Gadjov and Pavel 2023), which address resilience and the "price of malice" in the presence of a Byzantine player acting out of malice rather than self-interest. External attacks on a game were considered in (Feng and Hu 2023), in the form of compromised communication channels between agents in a graph. Additionally, our work connects to (Lin et al. 2020; Kazari, Shereen, and Dán 2023), which address attack and detection in multi-agent reinforcement learning (MARL). We frame our problem as a special case of a decentralized partially-observed Markov decision problem, where state transitions are not affected by actions and where the mediator's signal comprises players' partial observations. Since no analytical framework exists for solving the defender-attacker interaction in the general MARL case, our approach could be seen as a first step in this direction.

The rest of the paper organized as follows: In Section 2, we describe the preliminaries regarding CE and the framework of quickest change detection. Section 3 presents the problem formulation. In Section 4 we analyze the best responses of the adversary and the defender and propose a detection scheme. In Section 5 we evaluate the performance of the proposed detector. Section 6 concludes the paper.

2 Background

2.1 Correlated Equilibrium

Consider a strategic game $\mathcal{G} = (\mathcal{N}, \{\mathcal{A}^i\}_{i\in\mathcal{N}}, \{u^i\}_{i\in\mathcal{N}})$ where $\mathcal{N}$ is the set of the players, $\mathcal{A}^i$ is player $i \in \mathcal{N}$'s action set, and $u^i(a^i, a^{-i})$ is the utility of player i when she plays a^i and other players play a^{-i}. A CE of the game is a distribution $\boldsymbol{\pi}$ over $\mathcal{A} = \times_{i\in\mathcal{N}} \mathcal{A}^i$ such that for any player i and any function $\sigma^i : \mathcal{A}^i \to \mathcal{A}^i$ we have

$$\sum_{\mathbf{a}\in\mathcal{A}} \pi(\mathbf{a}) u^i(a^i, a^{-i}) \geq \sum_{\mathbf{a}\in\mathcal{A}} \pi(\mathbf{a}) u^i(\sigma^i(a^i), a^{-i}). \quad (1)$$

Coarse Correlated Equilibrium (CCE) is a closely related concept, defined as a distribution $\boldsymbol{\pi}$ over $\mathcal{A} = \times_{i\in\mathcal{N}} \mathcal{A}_i$ such that for any player i and any deviation $a^{i'} \in \mathcal{A}^i$ we have

$$\sum_{\mathbf{a}\in\mathcal{A}} \pi(\mathbf{a}) u^i(a^i, a^{-i}) \geq \sum_{\mathbf{a}\in\mathcal{A}} \pi(\mathbf{a}) u^i(a^{i'}, a^{-i}). \quad (2)$$

The actual realization of a CE in a game can be obtained by introducing a mediator who picks a sample $\mathbf{s} \in \mathcal{A}$ according to $\boldsymbol{\pi}$ and recommends the corresponding s^i to player $i \in \mathcal{N}$. Accordingly, player i's strategy can be represented by a function σ that maps s^i to an action in $\mathcal{A}^i$. Then, (1) can be interpreted as follows: assuming that all other players follow the mediator's recommendation, the strategy of "following the mediator's recommendation", i.e., $\sigma(s^i) = s^i$, is optimal for any player i. The difference between CE and CCE is that if the recommendation is drawn from a CCE, then it is a best response in expectation only before the player observes the recommendation signal. It can be shown that the set of correlated equilibria of a strategic game with finite action set is a non-empty, convex set (Maschler, Zamir, and Solan 2020). Moreover, the set of correlated equilibria is a subset of coarse correlated equilibria.

2.2 Stopping Variables and Quickest Change Detection

Let $\mathbf{X} = \{X_t, t = 1, 2, ...\}$ be a stochastic process. A stopping time T with respect to $\boldsymbol{X}$ is a random variable taking values in $\{1, 2, ...\}$ such that the event $\{T = t\}$ is a function of (only) $X_1, X_2, ..., X_t$. Consider now a stochastic process $\mathbf{X}$ where observations $X_1, X_2, \ldots, X_{\nu-1}$ are independently generated from a distribution with pmf (or pdf) f_0, while $X_\nu, X_{\nu+1}, \ldots$ are independently generated from a different distribution f_1. Both f_0 and f_1 are known distributions defined on the same space, but the change point ν is unknown. The objective of quickest change detection is to find a stopping time $T \geq \nu$ to detect the change as quickly as possible after it occurs. It is common to impose a constraint on the mean time between false alarms (MTBFA), i.e., $\mathbb{E}^{(\infty)}[T] \geq \gamma$, where γ is a predefined threshold and $\mathbb{E}^{(\infty)}$ denotes the expectation when $\nu = \infty$, i.e., when there is no change.

There are various optimization formulations for quickest detection; one widely used formulation is to minimize the mean detection delay (Lorden 1971),

$$W(T) = \sup_{\nu\geq 1} \text{ess sup} \mathbb{E}^{(\nu)}[(T - \nu + 1)^+ | X_1, \ldots, X_{\nu-1}],$$

where $\mathbb{E}^{(\nu)}$ denotes the expectation when the change occurs at ν, and ess sup refers to the essential supremum with respect to the conditional expectation. This formulation takes into account the worst case in terms of pre-change samples. An alternative formulation (Pollak 1985; Lai 1998) is to minimize $\sup_{\nu\geq 1} \text{ess sup}\, \mathbb{E}^{(\nu)}[(T - \nu) \mid T \geq \nu]$.

The CUSUM procedure (Page 1954) is a well-known stopping time whose optimality has been shown in various cases (Xie et al. 2021). CUSUM is based on the Sequential Probability Ratio Test (SPRT). For a sequence of i.i.d. samples $X_1, X_2, \ldots$, the SPRT defines a stopping time to decide from which of two alternative distributions the samples X_i are generated. Let f_1 and f_0 be the pmfs (pdfs) of the two alternative hypotheses, SPRT then computes the log likelihood ratio at time t,

$$\begin{aligned} Y_t &= \log\left(\frac{f_1(X_1)\ldots f_1(X_t)}{f_0(X_1)\ldots f_0(X_0)}\right) \\ &= \sum_{k=1}^{t} \log\left(\frac{f_1(X_k)}{f_0(X_k)}\right) = \sum_{k=1}^{t} l(X_k), \end{aligned} \quad (3)$$

and stops at the first t at which Y_t goes above or below prespecified thresholds. CUSUM relies on the likelihood ratio test and is defined as

$$T^{\text{CUSUM}}(\mu) = \inf\{t : R_t = \max_{1\leq k\leq t} \sum_{j=k}^{t} l(X_j) \geq \mu\}, \quad (4)$$

where μ is a threshold such that $\mathbb{E}^{(\infty)}[T_c] = \gamma$. (4) can be interpreted as applying a set of SPRT tests assuming that the change has occurred at time $k = 1, 2, \ldots, t$ and stop as soon as any of the likelihood ratios goes above μ. Another useful representation of R_t is by using the recursion

$$R_0 = 0, \qquad R_t = (R_{t-1} + l(X_t))^+. \tag{5}$$

3 Problem Formulation

We consider a strategic game $\mathcal{G} = (\mathcal{N}, \{\mathcal{A}^i\}_{i\in\mathcal{N}}, \{u^i\}_{i\in\mathcal{N}})$, with $\mathcal{N} = \{1, 2, \ldots, N\}$. Let $\boldsymbol{\pi}$ denote a CE of $\mathcal{G}$. We assume that $\mathcal{G}$ is played repeatedly at time steps $t = 1, 2, \ldots$, and at each time step, a mediator picks a sample $\mathbf{s}_t = (s_t^1, \ldots, s_t^N) \sim \boldsymbol{\pi}$, and recommends s_t^i to player $i \in \mathcal{N}$. Without loss of generality, we assume that $\forall i \in \mathcal{N}$ and $\forall \mathbf{a} \in \mathcal{A}$, the utility $u^i(\mathbf{a}) \geq 0$ (all utilities in $\mathcal{G}$ can be increased by an arbitrary constant without affecting the players' preferences). Players play a CE and all players follow the mediator's recommendation, so $a_t^i = s_t^i$. We assume that player i can only observe its own utility.

3.1 Attack and Defense Model

We consider a powerful adversary that manipulates the public signal $\mathbf{s}_t$ sent to the players, with the objective to reduce the utility that a particular victim player, say Player 1, obtains in $\mathcal{G}$. The adversary starts to manipulate the signal at time ν, and uses a stationary distribution $\boldsymbol{\tau}$ for the manipulation, i.e., the distribution $\boldsymbol{\tau}$ of adversarial manipulation at time $t \geq \nu$ is unchanged. Accordingly, $\mathbf{s}_t \sim \boldsymbol{\tau}$ for $t \geq \nu$. At the same time, the victim player aims to detect a potential manipulation of the signal based on the sequence of utilities $U_1, U_2, U_3, \ldots$ it observes, using an appropriately chosen stopping rule. From now on, we use the terms "Player 1" and "the defender" interchangeably. The adversary should thus remain undetectable as much as possible.

Note that with this attack and defense model, we can assume that different action profiles result in distinct utilities for the defender. This means that for any $a_{(1)}^1$, $a_{(2)}^1 \in \mathcal{A}^1$ and any $a_{(1)}^{-1}$, $a_{(2)}^{-1} \in \mathcal{A}^{-1}$, we have $u^1(a_{(1)}^1, a_{(1)}^{-1}) \neq u^1(a_{(2)}^1, a_{(2)}^{-1})$, where we use the game theoretic notation $\mathcal{A}^{-1}$ to denote the action set of all players but Player 1. This assumption can be made because action profiles with the same utility can be grouped and treated as a single outcome. From the defender's perspective, these action profiles are indistinguishable, and from the adversary's perspective, there is no incentive to alter the distribution over them.

3.2 Attacker-Defender Game

Observe that we can model the interaction between the attacker and the defender as a two-person zero-sum game $\mathcal{G}_S$. The defender chooses a stopping rule, which stops the game based on the sequence of observations. The adversary chooses a manipulation strategy $\boldsymbol{\tau}$ and a start time ν. The defender's cost (or the adversary's utility) is defined based on the reduction in the total utility that Player 1 obtains in $\mathcal{G}$, as well as the cost of potential false alarms. In the game, Player 1's utilities $U_1, U_2, U_3, \ldots$ are the observations of the defender. Hence, for any $t \in \{1, 2, \ldots\}$, $U_t \in \mathcal{U} = \{u^1(\mathbf{a}) : \mathbf{a} \in \mathcal{A}\}$. Note that Player 1 does not change her strategy in $\mathcal{G}$; she follows the recommended public signal. Player 1 receives s_t^1, plays $a_t^1 = s_t^1$, and observes u_t^1. Based on the sequence of her utilities and assuming that all other players follow the public signal, Player 1 tries to detect a potential attack on the public signal. The attacker-defender game $\mathcal{G}_S$ can thus be defined as follows.

Defender's Strategy: Let $\mathcal{T}$ be the set of all stopping times for the sequence $U_1, U_2, \ldots$. The defender chooses a $T \in \mathcal{T}$ as its strategy.

Adversary's Strategy : The adversary chooses a (possibly stochastic) start time ν via a specification $P(\nu = t | \nu \geq t, U_1, \ldots, U_{t-1})$. A formal definition of $\mathcal{V}$, the class of all possible such start times, can be found in (Ritov 1990): Let $X_1, X_2, \ldots$ be independent random variables with $X_t \sim \text{Unif}(0, 1)$. $\mathcal{V}$ is the class of all random variables ν such that $\mathbb{I}\{\nu = 1\}$ is a measurable function of X_1, and for $t > 1$, $\mathbb{I}\{\nu = t\}$ is a measurable function of X_t, $\mathbb{I}\{\nu < t\}$ and $U_1, \ldots, U_{t-1}$.

The adversary also chooses a manipulation strategy $\boldsymbol{\tau} \in \Delta^{|\mathcal{A}|-1}$, where Δ^{k-1} denotes the $(k-1)$-dimensional probability simplex in $\mathbb{R}^k$. Since we assumed that all players follow the mediator recommendations and because Player 1's utilities in $\mathcal{G}$ for any two action profiles are different, there is a one-to-one correspondence between $\mathcal{A}$ and $\mathcal{U}$. Thus, we can consider $\boldsymbol{\pi}$ and $\boldsymbol{\tau}$ to be distributions on $\mathcal{U}$ as well.

Cost (Utility) Function : We consider that a false alarm has cost C and we define the defender's cost (the adversary's utility) as

$$\begin{aligned} c(T, \nu, \boldsymbol{\tau}) =& \mathbb{I}\{T < \nu\} \left[C - \sum_{t=1}^{T} U_t \right] \\ &+ \mathbb{I}\{T \geq \nu\} \left[\sum_{t=\nu}^{T} V_t - \sum_{t=1}^{\nu-1} U_t \right], \end{aligned} \tag{6}$$

where $V_t = u_{\boldsymbol{\pi}} - U_t$, and $u_{\boldsymbol{\pi}}$ is the expected per-step utility in $\mathcal{G}$ under $\boldsymbol{\pi}$, $u_{\boldsymbol{\pi}} = \sum_{\mathbf{a}\in\mathcal{A}} \boldsymbol{\pi}(\mathbf{a}) u^1(\mathbf{a})$. V_t represents the cost imposed by the attack at time t. This cost is the difference between the utility that Player 1 actually obtains and the utility she would have expected to obtain had there been no attack. The negative terms in (6) correspond to the utility that Player 1 obtains before the change or before stopping.

Observe that with this formulation, it might be desirable for the adversary to select a distribution $\boldsymbol{\tau}$ that leads to an infinitesimally small per-step cost V_t and an infinitely large detection time, thereby maximizing the total cost. However, in a more realistic scenario, where the time horizon is finite, such a strategy is not feasible. As such, we consider that there is a lower bound $\epsilon > 0$ on the expected per-step cost that the adversary wants to cause. The set of distributions the attacker can choose τ from is then

$$\mathcal{D}_\epsilon = \left\{ \boldsymbol{\tau} \in \Delta^{|\mathcal{A}|-1} : \mathbb{E}_{\boldsymbol{a}_t \sim \boldsymbol{\tau}} \left[\frac{1}{T - \nu + 1} \sum_{t=\nu}^{T} V_t \right] \geq \epsilon \right\}. \tag{7}$$

Thus, the adversary's strategy is a pair $(\nu, \boldsymbol{\tau}) \in \mathcal{V} \times \mathcal{D}_\epsilon$.

4 Best Response Characterization

In this section, we analyze the asymptotic behavior of the maximin strategies of the players in $\mathcal{G}_S$ as the false alarm cost $C \to \infty$. In this case the defender aims to keep the probability of a false alarm low. More precisely, $P(T < \infty | \nu = \infty) \to 0$ as $C \to \infty$, as we show later.

We begin by defining the components that form our proposed strategies for the players.

Definition 1. *Let*

$$\boldsymbol{\tau}_\theta(u) = \frac{\boldsymbol{\pi}(u)\exp(-\theta u)}{\sum_{m\in\mathcal{U}} \boldsymbol{\pi}(m)\exp(-\theta m)}, \quad u \in \mathcal{U} \tag{8}$$

Moreover, let $\theta_{\min}$ be such that $\mathbb{E}_{U\sim\boldsymbol{\tau}_{\theta_{\min}}}[(u_{\boldsymbol{\pi}} - U)] = \epsilon$. Then, for the parameter space $\Theta = [\theta_{\min}, \infty)$, we define the family of distributions $\mathcal{F}_\Theta \triangleq \{\boldsymbol{\tau}_\theta : \theta \in \Theta\}$.

Definition 2. *(Lorden 1971) Let $l^\theta(U) = \log(\frac{\boldsymbol{\tau}_\theta(U)}{\boldsymbol{\pi}(U)})$. The generalized sequential probability ratio test of $\mathcal{F}_\Theta$ against $\boldsymbol{\pi}$ with threshold μ for sequence $U_1, U_2, \dots$ is defined as*

$$\bar{T}(\mu) = \inf\left\{ t \geq 1 : \sup_{\theta\in\Theta} \left(\sum_{k=1}^{t} l^\theta(U_k) \right) > \mu \right\} \tag{9}$$

Definition 3. *(Lorden 1971) Let $\bar{T}(\mu)$ be as defined in Definition 2. We define the stopping time $T^*(\mu)$ as*

$$T^*(\mu) = \min_{k>1}\{\bar{T}_k(\mu) + k - 1\}, \tag{10}$$

where $\bar{T}_k(\mu)$ is $\bar{T}(\mu)$ applied to the sequence $U_k, U_{k+1}, \dots$.

Our main results are in terms of two minmax theorems based on $\boldsymbol{\tau}_{\theta_{\min}}$ as the adversary's strategy and $T^*(\mu)$ as the defender's strategy.

Loosely speaking, the general idea we pursue is as follows: If the adversary's strategy was known to the defender, then the CUSUM procedure, whose performance is directly related to the KL-divergence of $\boldsymbol{\tau}$ and π, would be the optimal choice. The distributions in $\mathcal{F}_\Theta$ establish the optimal trade-off between the expected imposed cost and the mentioned KL-divergence from the adversary's point of view, motivating the definition of $\mathcal{F}_\Theta$. Then, with a specific set of parameterized distributions, like $\mathcal{F}_\Theta$, the stopping time defined in Definition 3 is asymptotically optimal. Finally, the attacker can choose the strategy that maximizes its payoff against the mentioned stopping rule. In the sequel, the claims stated above are presented and proved formally. Section 4.3 includes our main results.

4.1 The Defender's Strategy

First we start by restating Lorden's asymptotic optimality results on $T^*(\mu)$. Consider a sequence $U_1, U_2, \dots$ where $U_1, U_2, \dots, U_{\nu-1}$ are i.i.d samples from F_0 and $U_\nu, U_{\nu+1}, \dots$ are i.i.d. samples from a parameterized distribution F_θ with an unknown parameter $\theta \in \Theta$. Proposition 1 shows how to construct an optimal change detection scheme from one-sided tests, which clarifies the relation between Definition 2 and Definition 3, and explains the motivation behind defining T^* in Definition 3. Proposition 2 presents sufficient conditions for one-sided tests to satisfy the requirements of Proposition 1. Note that here the optimality criteria are in Lorden's sense, i.e., minimizing the mean detection delay

$$W_\theta(T) = \sup_{\nu\geq 1} \operatorname{ess\,sup} \mathbb{E}_\theta^{(\nu)}[(T - \nu + 1)^+ | U_1, ..., U_{\nu-1}],$$

where $\mathbb{E}_\theta^{(\nu)}$ denotes the expectation when the change of distribution to F_θ happens at time ν. In the sequel, $\mathbb{E}_\theta$ is equivalent to $\mathbb{E}_\theta^{(0)}$.

Proposition 1. *(Theorem 1 in (Lorden 1971)) Assume that there exists a class of one-sided tests $\{\bar{T}^\alpha\}$ such that for all $0 < \alpha < 1$,*

$$P(\bar{T}^\alpha < \infty | \nu = \infty) \leq \alpha \tag{11}$$

Moreover, assume that for all $\theta \in \Theta$, as $\alpha \to 0$ we have

$$\mathbb{E}_\theta[\bar{T}^\alpha] \sim \frac{|\log\alpha|}{D_{KL}(f_\theta||f_0)}, \;\; D_{KL}(f_\theta||f_0) < \infty \tag{12}$$

For $\gamma \geq 1$, let $\alpha = \gamma^{-1}$ and define $T^\gamma = min_{k>1}\{\bar{T}_k^\alpha + k - 1\}$, where $\bar{T}_k^\alpha$ is $\bar{T}^\alpha$ applied to $U_k, U_{k+1}, \dots$. Then T^γ is a stopping variable with $\mathbb{E}^{(\infty)}[T] \geq \gamma$ that minimizes $W_\theta(T)$ for all $\theta \in \Theta$ as $\gamma \to \infty$, and we have

$$\mathbb{E}_\theta[T^\gamma] \sim \frac{\log\gamma}{D_{KL}(f_\theta||f_0)} \quad \text{as } \gamma \to \infty. \tag{13}$$

Proposition 2. *Suppose F_θ's are members of an exponential family of distributions such that*

$$dF_\theta(x) = \exp(\theta x - b(\theta))dF_0(x), \quad \theta \in \Theta \cup 0 \tag{14}$$

where $\Theta = [\theta_{min}, \infty)$ is an interval in $\mathbb{R}^+$, and $b(0) = 0$. Furthermore, let us define

$$\mu_\alpha = \log(3(D_{KL}(f_{\theta_{min}}||f_0)+1)^2) - \log(\alpha|\log\alpha|). \tag{15}$$

Then $\bar{T}(\mu_\alpha)$ (defined by Definition 2 with $l^\theta(U) = \log(\frac{f_\theta(U)}{f_0(U)})$ satisfies (11) and (12).

Proposition 3. *If the adversary's strategy $\boldsymbol{\tau}$ belongs to $\mathcal{F}_\Theta$, then $T^*(\mu_\alpha)$ defined in Definition 3 is asymptotically optimal with respect to $W_\theta(T)$ as $\alpha \to 0$.*

Proof. Observe that by considering $x = -u$, $f_0 \equiv \boldsymbol{\pi}$, and by defining $b(\theta) \triangleq \log(\sum_{m\in\mathcal{U}} \boldsymbol{\pi}(m)\exp(-\theta m))$, the members of $\mathcal{F}_\Theta$, defined in Definition 1, can be represented as in (14). Thus, Proposition 2 and consequently Proposition 1 can be applied to $\bar{T}$ and T^*, respectively. □

As (13) suggests, the asymptotic performance of T^* is inversely proportional to the KL-divergence of the pre-change and post-change distributions. Next, we identify the optimal trade-off between the cost and the KL-divergence of the adversary's strategy, which leads to Definition 1.

4.2 Adversarial Strategy

For a given $\delta > 0$, let us consider the problem faced by the adversary,

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & \mathbb{E}_{\boldsymbol{\tau}}[U] \\ s.t. \quad & \boldsymbol{\tau} \in \Delta^{|\mathcal{A}|-1}, \quad D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi}) \leq \delta. \end{aligned} \tag{16}$$

In the next proposition, we show that the solution of this problem is a distribution in $\mathcal{F}_\Theta$.

Proposition 4. *Let* $\mathbf{a}_{min} \triangleq \min_{\mathbf{a}} u^1(\mathbf{a})$. *If* $\delta < -\log \boldsymbol{\pi}(\mathbf{a}_{min})$, *then, the minimizer of (16) is* $\boldsymbol{\tau}_{\theta^*}$, *where* $\theta^* \in (0, \infty)$ *is such that* $D_{KL}(\boldsymbol{\tau}_{\theta^*}||\boldsymbol{\pi}) = \delta$. *Moreover, if* $\delta \geq -\log \boldsymbol{\pi}(\mathbf{a}_{min})$, *then (16) attains its infimum at* $\boldsymbol{\tau}_{\boldsymbol{\theta}}$ *as* $\theta \to \infty$.

Remark 1. There is a one-to-one correspondence between θ, $d(\theta) \triangleq D_{KL}(\boldsymbol{\tau}_\theta||\boldsymbol{\pi})$ and also $u_\theta \triangleq \mathbb{E}_{\boldsymbol{\tau}_\theta}[U]$. In other words, as θ varies from 0 to ∞, $d(\theta)$ increases monotonically from 0 to $-\log \boldsymbol{\pi}(\mathbf{a}_{min})$, and u_θ decreases monotonically from $u_{\boldsymbol{\pi}}$ to $u_{\min}$, where $u_{\min} = u^1(\mathbf{a}_{min})$. Consequently, if $\epsilon \in (0, u_{\boldsymbol{\pi}} - u_{\min})$, then F_Θ as defined in Definition 1 is well-defined, and for all $\theta \in \Theta$, $u_{\boldsymbol{\pi}} - u_\theta \geq \epsilon$.

4.3 Main Results

We present our main results in the next two theorems, the proofs are presented in the Appendix. For $C > 1$, let $\mu(C)$ be such that

$$C = u_{\boldsymbol{\pi}} \mathbb{E}^{(\infty)}[T^*(\mu(C))] + \epsilon \mathbb{E}_{\theta_{\min}}[T^*(\mu(C))], \quad (17)$$

where $\theta_{\min}$ is as defined in Definition 1.

Theorem 1. *There is a* $\nu^*(C) \in \mathcal{V}$, *such that*

$$\sup_{\boldsymbol{\tau} \in \mathcal{D}_\epsilon} \sup_{\nu \in \mathcal{V}} \inf_{T \in \mathcal{T}} \mathbb{E}[c(T, \nu, \boldsymbol{\tau})] \sim \mathbb{E}[c(T^*(\mu(C)), \nu^*(C), \boldsymbol{\tau}_{\theta_{min}})] \quad as \;\; C \to \infty. \quad (18)$$

Theorem 2. *With* $\mu(C)$ *defined in (17), we have*

$$\inf_{T \in \mathcal{T}} \sup_{\nu \in \mathcal{V}} \sup_{\boldsymbol{\tau} \in \mathcal{F}_\Theta} \mathbb{E}[c(T, \nu, \boldsymbol{\tau})] \sim \mathbb{E}[c(T^*(\mu(C)), \nu_1, \boldsymbol{\tau}_{\theta_{min}})] \quad as \; C \to \infty, \quad (19)$$

where ν_1 *is defined as* $P(\nu_1 = 1) = 1$.

Theorem 1 implies that, without having any prior knowledge of the defender strategy, the highest possible cost that the adversary can guarantee is achieved by choosing $\boldsymbol{\tau}_{\theta_{\min}}$ as the manipulation strategy. Additionally, if the adversary chooses this strategy the best response of the defender is $T^*(\mu(C))$. Furthermore, Theorem 2 characterizes the maxmin strategy from the defender's perspective. In Theorem 2, the supremum over $\boldsymbol{\tau}$ is taken only within $\mathcal{F}_\theta$, meaning that if the attack is constrained to $\mathcal{F}_\theta$ then choosing $T^*(\mu(C))$ leads to the lowest cost that the defender can guarantee, asymptotically. However, if the adversary is aware that $T^*(\mu(C))$ is the chosen strategy, such guarantee cannot be made, and the adversary might be able to impose a higher cost by selecting some $\boldsymbol{\tau} \in \mathcal{D}_\epsilon \setminus \mathcal{F}_\theta$.

Remark 2. For $0 < \alpha < 1$, define μ_α as in (15) and $\gamma = \alpha^{-1}$. Then $T^*(\mu(C))$ defined by (17) as $C \to \infty$ is equivalent to $T^*(\mu_\alpha)$ for $\alpha \to 0$ (see the proof of Theorem 1 in the Appendix). Moreover, we have $P(T^*(\mu_\alpha) < \infty | \nu = \infty) \to 0$, $\mathbb{E}^{(\infty)}[T^*(\mu_\alpha)] \sim \gamma$ and $\mathbb{E}_\theta[T^*(\mu_\alpha)] \sim \frac{\log \gamma}{d(\theta)}$.

Compared to solving (17), Remark 2 provides an easier method for implementing T^* based on α (or γ) as input value that determines the threshold μ according to (15).

		Player 2		
		A_2	B_2	C_2
Player 1	A_1	0,0	6,1	9,3
	B_1	1,6	5,5	4,2
	C_1	3,9	2,4	7,7

Table 1: Payoffs in the Extended Game of Chicken.

4.4 Implementation of the Detection Mechanism

For a practical implementation of $T^*(\mu)$, we adapt the procedure proposed by (Lorden 1971). Note that $T(\mu)$ in (9) can be considered as stopping at the first time for which

$$\sup_{\theta \in \Theta} \{-\theta S_t - t b(\theta)\} > \mu, \quad (20)$$

where $S_t = U_1 + \cdots + U_t$. Equivalently, (20) can be rewritten as

$$S_t < -\inf_{\theta \in \Theta} \left\{ \frac{\mu}{\theta} + t \frac{b(\theta)}{\theta} \right\}. \quad (21)$$

It can be shown that the infimum in (21) is attained at $\theta_{\min}$ for $t \geq \frac{\mu}{d(\theta_{\min})}$, and at $\theta^{(t)}$ satisfying $d(\theta^{(t)}) = \frac{\mu}{t}$ for $t < \frac{\mu}{d(\theta_{\min})}$. Then for a recursive implementation of $T^*(\mu)$ in (10) we define a set of $M = \lfloor \frac{\mu}{d(\theta_{\min})} \rfloor$ thresholds as

$$z^{(k)} = -\frac{\mu}{\theta^{(k)}} - k \frac{b(\theta^{(k)})}{\theta^{(k)}}, \quad k = 1, 2, \ldots, M \quad (22)$$

Moreover, we define $Q^{(k)} = U_t + \cdots + U_{t-k+1}, \quad k = 1, \ldots, \min\{M, t\}$. Then, $T^*(\mu)$ is equivalent to the following procedure: at each step, perform a CUSUM test against $\theta_{\min}$ with threshold μ by computing R_t (see (4)). Additionally, we compute $Q^{(k)}$ and we stop if $Q^{(k)} < z^{(k)}$. Note that $Q^{(k)}$ can be updated recursively as

$$Q^{(k)\prime} = \begin{cases} U_t + Q^{(k-1)} & , t - t_1 \geq k \\ 0 & , t - t_1 < k \end{cases} \quad (23)$$

where $Q^{(k)\prime}$ denotes the updated value of $Q^{(k)}$ after observing U_t, and t_1 is the last time that $R_{t_1} = 0$. Whenever $R_t = 1$, all previous observations can be neglected and a new cycle be started by resetting t_1 and setting $Q^{(k)} = 0$.

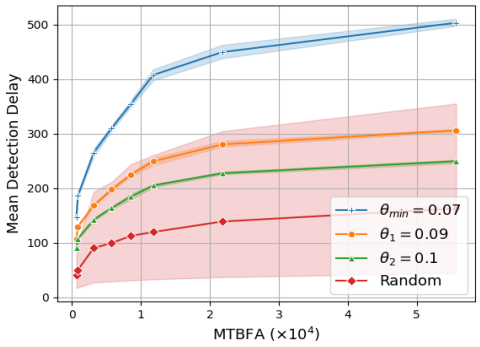
5 Numerical Results

In this section we evaluate the proposed detection scheme against several adversarial strategies in two games.[1]

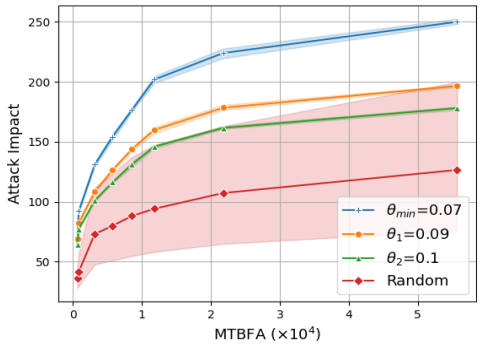
5.1 Toy Example

We start with considering an extension of the Game of Chicken game (Duffy and Feltovich 2010), with three pure strategies. The utilities are shown in Table 1. It is straightforward to verify that the following distribution is a correlated equilibrium of the game: $\boldsymbol{\pi}(A_1B_2) = \boldsymbol{\pi}(B_1A_2) = \boldsymbol{\pi}(B_1B_2) = \frac{1}{36}$, $\boldsymbol{\pi}(A_1C_2) = \boldsymbol{\pi}(C_1A_2) = \frac{1}{3}$, $\boldsymbol{\pi}(C_1C_2) = \frac{1}{4}$, $\boldsymbol{\pi}(A_1A_2) = \boldsymbol{\pi}(C_1B_2) = \boldsymbol{\pi}(B_1C_2) = 0$. The corresponding expected pay-off for both players is $u_{\boldsymbol{\pi}} = 6.08$,

[1] Code available at https://github.com/kiarashkaz/Detection-of-Adversarial-Attacks-against-CE

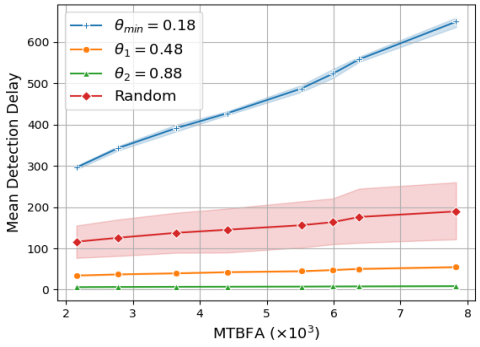


(a) Mean detection delay

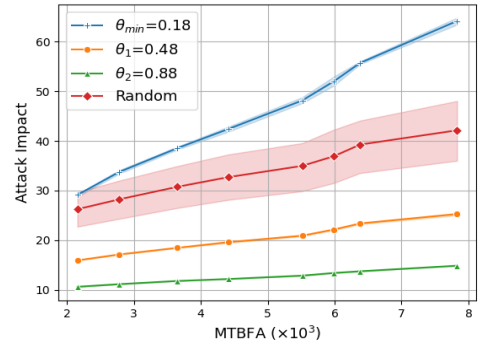


(b) Attack impact

Figure 1: Mean detection delay and attack impact with $\epsilon = 0.5$ in the Extended Game of Chicken with $u_{\boldsymbol{\pi}} = 6.08$. Confidence intervals are based on 5 runs.



(a) Mean detection delay



(b) Attack impact

Figure 2: Mean detection delay and attack impact with $\epsilon = 0.1$ in the Routing Game with $u_{\boldsymbol{\pi}} = 4.77$. Confidence intervals are based on 5 runs.

which results in a higher utilitarian social welfare than any pure Nash equilibrium.

We consider four attack strategies. $\boldsymbol{\tau}_{\theta_{\text{min}}}$ is the maxmin strategy of the adversary with $\theta_{\text{min}} = 0.07$. $\boldsymbol{\tau}_{\theta_1}$ and $\boldsymbol{\tau}_{\theta_2}$ are two other adversarial strategies with $\theta_{\text{min}} < \theta_1 = 0.09 < \theta_2 = 0.1$. As a baseline we use an attack where $\boldsymbol{\tau}$ is chosen randomly from the probability space D_ϵ (over elements with a non-zero probability of selection by the mediator). We refer to the different attacks as scenarios, hence we have five scenarios including the no attack scenario. As performance metrics we use the *MTBFA* $\mathbb{E}^{(\infty)}[T^*]$, and the mean detection delay. Additionally, we define the *attack impact* as the difference between the total utility obtained by the victim player during the period the attack is active before its detection and the expected total utility that the agent would have obtained if there had been no attack during that period.

For the evaluation, we consider a minimum per step attack impact of $\epsilon = 0.5$ and episodes with a maximum length of 10^5 time steps. Each episode ends when either a detection happens by the defender or the maximum number of time steps is reached. For each scenario we run 1000 episodes. In each episode of each scenario, the public signal is generated according to the corresponding adversarial strategy after the attack starts or $\boldsymbol{\pi}$ if there is no attack.

Figure 1a shows the mean detection delay vs. MTBFA for the considered attacks, obtained by varying the detector threshold μ_α. The figure shows that the time-to-detect increases as a with the logarithm of the mean time between false alerts, in accordance with Remark 2. Importantly, the figure shows that the detector is robust to the attacker's choice of θ; a higher value of θ allows the victim player to detect the attack sooner. Figure 1b shows the attack impact during the attack as a function of the average MTBFA. The figure shows that the detection scheme effectively limits the impact the adversarial impact. For example, at an average MTBFA of 5×10^4 the total utility the player obtains if there is no limit on the number of time steps is around 3×10^5, while the attack impact with the minmax strategy

Algorithm 1: Detection Strategy

Input: ϵ, α
Initialization:
Find $\theta_{\min}$ such that $u_{\theta_{\min}} = u_{\boldsymbol{\pi}} - \epsilon$
Compute μ_α as in (15)
Find $\theta^{(k)}$ such that $d(\theta^{(k)}) = \frac{\mu_\alpha}{k}, \quad k = 1, \dots, M$
Compute $z^{(k)}$ as in (22), $\quad k = 1, \dots, M$
Set $R = 0,\ Q^{(k)} = 0, \quad k = 1, \dots, M$
Repeated Play Procedure:
while not STOP **do**
  Receive s_t, play $a_t = s_t$, and observe U_t
  Compute $l_t = \log(\frac{\boldsymbol{\tau}_\theta(U_t)}{\boldsymbol{\pi}(U_t)})$ with $\theta = \theta_{\min}$
  Update R with $R \leftarrow (R + l_t)^+$
  if $R > \mu_\alpha$ **then**
    STOP
  else if $R > 0$ **then**
    Update $Q^{(k)}$ as in (23), $\quad k = 1, \dots, M$
    if $Q^{(k)} < z^{(k)}$ for any k **then**
      STOP
    end if
  else
    Reset $R = 0,\ Q^{(k)} = 0, \quad k = 1, \dots, M$
  end if
end while

is less than 250. Furthermore, the figure indicates that $\boldsymbol{\tau}_{\theta_{\min}}$ is the most effective attack strategy. Also, it can be observed that even with a random attack distribution that does not lie within $\mathcal{F}_\Theta$, the detection remains effective.

A note on the selection of ϵ and α: In a practical scenario, the selection of ϵ and α is based on the player's tolerance for incurred costs and the acceptable false alarm rate. This choice is also closely tied to the time horizon over which the repeated game is intended to be played. For example, suppose that in a real scenario, the Extended Game of Chicken is to be played 500 times, and the desired false alarm rate is $\frac{1}{5\times10^4}$, which corresponds to a MTBFA of 5×10^4. According to Figure 1, $\epsilon = 0.5$ is a good choice for a victim player who can tolerate an adversarial attack impact of 250, i.e., around 8% of the total expected utility 6.08×500 (observe that in the figure, mean detection time and the highest attack impact at an MTBFA of 5×10^4 are around 500, 250, respectively). In this case, if the adversary selects $\theta > \theta_{\min}$ the total imposed cost would be lower. Also, selecting a $\theta < \theta_{\min}$ by the adversary would result in an attack with a higher expected detection time. However, since the total length of repeated play is fixed at 500, even if the attack is not detected, the total imposed cost would still be less than 250 because the per-step cost is less than 0.5. In the Appendix, we provide additional results showing the numerical values of the attack impact as ϵ varies.

5.2 Routing Game

Second, we consider a traffic routing game based on the model described in (Sessa et al. 2020) using the Sioux Falls traffic network (LeBlanc, Morlok, and Pierskalla 1975). There are $N = 528$ players, each aiming to transfer specific units of goods between two nodes in the network. The network comprises 24 nodes and 76 edges. Each player has $|\mathcal{A}^i| = 5$ pure strategies, representing the 5 shortest paths between its source and destination. The travel time on each edge is determined by its capacity and the total volume of goods passing through it. Each player seeks to minimize the total time required to transport its goods, we thus use the negative of the total time as the utility of a player. For more details on the traffic and travel time models we refer to (Sessa et al. 2020).

A CE in this game is a distribution over $\mathcal{A} = 5^{528}$ outcomes, hence infeasible to compute. We thus used the no-regret learning algorithm GP-MW (Sessa et al. 2019) for computing an approximate CCE $\boldsymbol{\pi}$. We selected one of the agents as the victim and quantized its utility in 1000 bins, i.e., action profiles leading to utilities within the same quantization level were considered the same. Moreover, due to the large scale of utility values, we divided all the utilities by 1000. For the evaluation we ran episodes with a length of 10^4 time steps, and used $\epsilon = 0.1$. For each scenario we ran 1000 episodes. The considered adversarial strategies correspond to $\theta_{\min} = 0.18$, $\theta_1 = 0.48$, $\theta_2 = 0.88$, and a distribution chosen randomly from probability space D_ϵ.

Figure 2 shows the mean detection delay (2a) and the attack impact (2b) of the attack as a function of the MTBFA. Note that due to the limited range of the horizontal axis (1 order of magnitude less than in Figure 1), the logarithmic growth of the curves is not apparent. Nonetheless, the results demonstrate that despite the complexity of the game and although the public signal is an approximate CCE, the proposed adversarial attacks on the public signal are effective. Similarly, the proposed detection method proves effective in promptly identifying the attacks and in mitigating their impact. This observation highlights the scalability of the proposed method for real-world applications.

6 Conclusion

We considered the detection of adversarial attacks on the signals of a mediator in correlated equilibria of non-cooperative games. We proposed a detector based on a generalized CUSUM stopping rule designed for a specific set of adversarial distributions $\mathcal{F}_\Theta$. We showed that any attack following a distribution in $\mathcal{F}_\Theta$ results in the maximum possible impact, with the constrained KL-distance to the CE, which is the primary factor determining the detection delay of the generalized CUSUM scheme. We proved that the proposed detection scheme is the maxmin strategy of the defender given that the attack distribution belongs to $\mathcal{F}_\Theta$. Additionally, we showed that the strategy that guarantees the highest possible impact for the adversary is also lies within $\mathcal{F}_\Theta$. Through numerical evaluations, we showed that the proposed method effectively mitigates the impact of various attacks, with a detection delay that increases logarithmically with the mean time between false alarms. Our results demonstrated that the detection mechanism is robust against various attacker strategies, including random attacks that fall outside $\mathcal{F}_\Theta$, and remains effective even in a complex routing game with a large number of action profiles.

7 Acknowledgments

The work was partly funded by the Swedish Research Council through project 2020-03860 and by Digital Futures through the CLAIRE project. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at Linköping University partially funded by the Swedish Research Council through grant agreement no. 2022-06725 .

References

Anagnostides, I.; Daskalakis, C.; Farina, G.; Fishelson, M.; Golowich, N.; and Sandholm, T. 2022. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *Proc. of ACM Symposium on Theory of Computing (STOC)*, 736–749.

Aumann, R. J. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society*, 1–18.

Cesa-Bianchi, N.; and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge university press.

Chen, X.; and Deng, X. 2006. Settling the Complexity of Two-Player Nash Equilibrium. In *FOCS*, volume 6, 261–272.

Daskalakis, C.; Fishelson, M.; and Golowich, N. 2021. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems (NeurIPS)*, 34: 27604–27616.

Daskalakis, C.; Goldberg, P. W.; and Papadimitriou, C. H. 2009. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2): 89–97.

Duffy, J.; and Feltovich, N. 2010. Correlated equilibria, good and bad: an experimental study. *International Economic Review*, 51(3): 701–721.

Feng, Z.; and Hu, G. 2023. Resilient Distributed Algorithms for Exponential Nash Equilibrium Seeking. In *International Conference on Industrial Artificial Intelligence (IAI)*, 1–4.

Gadjov, D.; and Pavel, L. 2023. An algorithm for resilient Nash equilibrium seeking in the partial information setting. *IEEE Transactions on Control of Network Systems*, 10(4): 1645–1655.

Kazari, K.; Shereen, E.; and Dán, G. 2023. Decentralized Anomaly Detection in Cooperative Multi-Agent Reinforcement Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 162–170.

Lai, T. L. 1998. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information theory*, 44(7): 2917–2929.

LeBlanc, L. J.; Morlok, E. K.; and Pierskalla, W. P. 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation research*, 9(5): 309–318.

Leung, C. K. Y. 2007. Equilibrium correlations of asset price and return. *The Journal of Real Estate Finance and Economics*, 34: 233–256.

Lin, J.; Dzeparoska, K.; Zhang, S. Q.; Leon-Garcia, A.; and Papernot, N. 2020. On the robustness of cooperative multi-agent reinforcement learning. In *2020 IEEE Security and Privacy Workshops (SPW)*, 62–68. IEEE.

Lorden, G. 1971. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 1897–1908.

Maschler, M.; Zamir, S.; and Solan, E. 2020. *Game theory*. Cambridge University Press.

Ning, Y.; and Du, L. 2023. Robust and resilient equilibrium routing mechanism for traffic congestion mitigation built upon correlated equilibrium and distributed optimization. *Transportation research part B: methodological*, 168: 170–205.

Page, E. S. 1954. Continuous inspection schemes. *Biometrika*, 41(1/2): 100–115.

Papadimitriou, C. H.; and Roughgarden, T. 2008. Computing correlated equilibria in multi-player games. *Journal of the ACM (JACM)*, 55(3): 1–29.

Pollak, M. 1985. Optimal detection of a change in distribution. *The Annals of Statistics*, 206–227.

Ritov, Y. 1990. Decision theoretic optimality of the CUSUM procedure. *The Annals of Statistics*, 1464–1469.

Roth, A. 2008. The price of malice in linear congestion games. In *International Workshop on Internet and Network Economics*, 118–125. Springer.

Roughgarden, T.; and Tardos, É. 2002. How bad is selfish routing? *Journal of the ACM*, 49(2): 236–259.

Sessa, P. G.; Bogunovic, I.; Kamgarpour, M.; and Krause, A. 2019. No-regret learning in unknown games with correlated payoffs. *Advances in Neural Information Processing Systems (NeurIPS)*, 32.

Sessa, P. G.; Bogunovic, I.; Krause, A.; and Kamgarpour, M. 2020. Contextual games: Multi-agent learning with side information. *Advances in Neural Information Processing Systems (NeurIPS)*, 33: 21912–21922.

Xie, L.; Zou, S.; Xie, Y.; and Veeravalli, V. V. 2021. Sequential (Quickest) Change Detection: Classical Results and New Directions. *IEEE Journal on Selected Areas in Information Theory*, 2(2): 494–514.

Appendix

A Proof of Theorems and Propositions

A.1 Proof of Proposition 2

As stated in (Lorden 1971), if μ_α is chosen such that $\bar{T}(\mu_\alpha)$ satisfies (11) and $\mu_\alpha \sim |\log\alpha|$, then it satisfies (12) as well. Theorem 1 in (Lorden 1973) proves that choosing μ_α as mentioned in Proposition 2 satisfies (11). Moreover, we can write

$$\begin{aligned}\mu_\alpha &\sim -\log(\alpha) - \log(|\log(\alpha)|)\\ &\sim -\log(\alpha) = |\log(\alpha)|, \text{ as } \alpha \to 0\end{aligned} \tag{24}$$

A.2 Proof of Proposition 4

In this proof we treat $\boldsymbol{\tau}$, $\mathbf{u}$, and $\boldsymbol{\pi}$ as their representing vectors. In the following, $\mathbf{x}[k]$ denotes the k-th element of vector $\mathbf{x}$. (16) can be rewritten as

$$\begin{aligned}\min_{\boldsymbol{\tau}} \quad & \mathbf{u}^\top\boldsymbol{\tau}\\ s.t. \quad & ||\boldsymbol{\tau}||_1 = 1,\\ & \boldsymbol{\tau} \geq 0,\\ & D_{KL}(\boldsymbol{\tau},\boldsymbol{\pi}) = (\boldsymbol{\tau}^\top\log\boldsymbol{\tau} - \boldsymbol{\tau}^\top\log\boldsymbol{\pi}) \leq \delta\end{aligned} \tag{25}$$

If $\delta = \infty$, then there is no condition on the KL-divergence and the solution is $\boldsymbol{\tau}^* = \mathbf{e}_{\mathbf{a}_{min}}$, where $\mathbf{e}$ represents the unit vector. Otherwise we would have $D_{KL}(\boldsymbol{\tau}^*||\boldsymbol{\pi}) < \infty$, which implies that $\boldsymbol{\tau}^*[k] = 0$ iff $\boldsymbol{\pi}[k] = 0$. Let $\tilde{\boldsymbol{\pi}}$ be the vector representing the non-zero elements of $\boldsymbol{\pi}$. Also, $\tilde{\boldsymbol{\tau}}$ and $\tilde{\mathbf{u}}$ represent the same indices of $\boldsymbol{\tau}$ and $\mathbf{u}$, respectively. Then, problem (25) can be expressed with $\boldsymbol{\pi}$, $\boldsymbol{\tau}$, and $\mathbf{u}$ replaced by $\tilde{\boldsymbol{\pi}}$, $\tilde{\boldsymbol{\tau}}$, and $\tilde{\mathbf{u}}$.

It is straightforward to verify that the optimization problem is convex and the Slater's condition holds. Accordingly, we use KKT theorem to derive optimality conditions. The Lagrangian can be written as:

$$\begin{aligned}\mathcal{L} = \tilde{\mathbf{u}}^\top\tilde{\boldsymbol{\tau}} + \eta\left((\tilde{\boldsymbol{\tau}}^\top\log\tilde{\boldsymbol{\tau}} - \tilde{\boldsymbol{\tau}}^\top\log\tilde{\boldsymbol{\pi}}) - \delta\right)\\ + \lambda(||\tilde{\boldsymbol{\tau}}||_1 - 1) - \tilde{\boldsymbol{\beta}}^\top\tilde{\boldsymbol{\tau}}\end{aligned} \tag{26}$$

where $\eta, \tilde{\boldsymbol{\beta}}[k] \geq 0$, and the complementary slackness implies that $\tilde{\boldsymbol{\beta}}^*[k]\tilde{\boldsymbol{\tau}}^*[k] = 0$ and $\eta^*(D_{KL}(\tilde{\boldsymbol{\tau}}^*||\tilde{\boldsymbol{\pi}}) - \delta) = 0$. We know that $\tilde{\boldsymbol{\tau}}^*[k] \neq 0$ for any k, so $\tilde{\boldsymbol{\beta}}^*[k] = 0$. Setting the gradient of $\mathcal{L}$ to zero leads to

$$\nabla_{\tilde{\boldsymbol{\tau}}}\mathcal{L} = \tilde{\mathbf{u}} + \eta(\log\tilde{\boldsymbol{\tau}} - \log\tilde{\boldsymbol{\pi}}) + \eta\mathbf{1} + \lambda\mathbf{1} = \mathbf{0} \tag{27}$$

If $D_{KL}(\tilde{\boldsymbol{\tau}}^*||\tilde{\boldsymbol{\pi}}) < \delta$, then η^* needs to be zero, leading to $\tilde{\boldsymbol{u}} + \lambda\mathbf{1} = \mathbf{0}$, which is impossible. Thus, $D_{KL}(\tilde{\boldsymbol{\tau}}^*||\tilde{\boldsymbol{\pi}}) = \delta$.

Solving (27) for $\tilde{\boldsymbol{\tau}}$, we get $\tilde{\boldsymbol{\tau}}^*[k] = \tilde{\boldsymbol{\pi}}[k]\exp(-\frac{\tilde{\mathbf{u}}[k]}{\eta^*} - \frac{\lambda^*}{\eta^*} - 1)$. The value of λ^* can be found according to the condition $||\tilde{\boldsymbol{\tau}}||_1 = 1$:

$$\lambda^* = \eta^*\log\left(e^{-1}\sum_k \tilde{\boldsymbol{\pi}}[k]\exp(-\frac{\tilde{\mathbf{u}}[k]}{\eta^*})\right) \tag{28}$$

Finally, by substituting (28) into the expression of $\tilde{\boldsymbol{\tau}}^*$, we obtain

$$\tilde{\boldsymbol{\tau}}^*[k] = \frac{\tilde{\boldsymbol{\pi}}[k]\exp(-\frac{\tilde{\mathbf{u}}[k]}{\eta^*})}{\sum_{k'}\tilde{\boldsymbol{\pi}}[k']\exp(-\frac{\tilde{\mathbf{u}}[k']}{\eta^*})} \tag{29}$$

By defining $\theta^* \triangleq \frac{1}{\eta^*}$, (29) takes the form of $\boldsymbol{\tau}_\theta$ as in Definition 1. θ^* should be such that $D_{KL}(\boldsymbol{\tau}^*, \boldsymbol{\pi}) = \delta$. For the rest of the proof we require the following lemma:

Lemma 1. *For any θ, $d(\theta) = D_{KL}(\boldsymbol{\tau}_\theta, \boldsymbol{\pi})$ can be expressed as:*

$$d(\theta) = \theta b'(\theta) - b(\theta)$$

Proof. we can write:

$$\begin{aligned}\theta b'(\theta) - b(\theta) &= \frac{-\sum_{u\in\mathcal{U}}\theta u\boldsymbol{\pi}(u)\exp(-\theta u)}{\sum_{u\in\mathcal{U}}\boldsymbol{\pi}(u)\exp(-\theta u)} - b(\theta)\\ &= -\sum_{u\in\mathcal{U}}\boldsymbol{\tau}_\theta(u)\theta u - \sum_{u\in\mathcal{U}}\boldsymbol{\tau}_\theta(u)b(\theta)\\ &= \sum_{u\in\mathcal{U}}\boldsymbol{\tau}_\theta(u)(-\theta u - b(\theta))\\ &= \sum_{u\in\mathcal{U}}\boldsymbol{\tau}_\theta(u)\log\frac{\boldsymbol{\tau}_\theta(u)}{\boldsymbol{\pi}(u)} = d(\theta)\end{aligned} \tag{30}$$

□

Note that as $\theta \to \infty$, we have $\sum_{u\in\mathcal{U}}\boldsymbol{\pi}(u)\exp(-\theta u) \sim \boldsymbol{\pi}(u_{min})\exp(-\theta u_{min})$. Accordingly, as $\theta \to \infty$:

$$\theta b'(\theta) = \frac{-\sum_{u\in\mathcal{U}}\theta u\boldsymbol{\pi}(u)\exp(-\theta u)}{\sum_{u\in\mathcal{U}}\boldsymbol{\pi}(u)\exp(-\theta u)} \sim -\theta u_{min} \tag{31}$$

$$b(\theta) \sim \log(\boldsymbol{\pi}(u_{min})e^{-\theta u_{min}}) \tag{32}$$

Thus, we have:

$$\lim_{\theta\to\infty} d(\theta) = -\log \boldsymbol{\pi}(u_{min}) \tag{33}$$

Therefore, if $\delta < -\log \boldsymbol{\pi}(u_{min})$, then there exists a θ^* such that $d(\theta^*) = \delta$, and $\boldsymbol{\tau}_{\theta^*}$ is the unique solution to (16). Otherwise, the optimization problem has no solution, and the objective function reaches an infimum at $\theta = \infty$.

A.3 Proof of Theorem 1

We can write

$$\begin{aligned}\mathbb{E}[c(T,\nu,\boldsymbol{\tau})] = P(T<\nu)\left(C - \mathbb{E}\left[\sum_{t=1}^{T} U_t \Big| T<\nu\right]\right)\\ +P(T\geq\nu)\left(\mathbb{E}\left[\sum_{t=\nu}^{T} V_t \Big| T\geq\nu\right]\right.\\ \left. - \mathbb{E}\left[\sum_{t=1}^{\nu-1} U_t \Big| T\geq\nu\right]\right)\end{aligned} \tag{34}$$

Since T is a stopping time, according to the Wald's identity (Wald 1946), we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} U_t \Big| T<\nu\right] &= \mathbb{E}[U_t|T<\nu]\ \cdot \mathbb{E}[T|T<\nu]\\ &= u_{\boldsymbol{\pi}}\mathbb{E}^{(\infty)}[T]\end{aligned} \tag{35}$$

$$\begin{aligned}\mathbb{E}\left[\sum_{t=\nu}^{T} V_t \Big| T\geq\nu\right] &= \mathbb{E}[V_t|T\geq\nu]\cdot\mathbb{E}^{(\nu)}_{\boldsymbol{\tau}}[T-\nu+1|T\geq\nu]\\ &= (u_{\boldsymbol{\pi}}-u_{\boldsymbol{\tau}})\mathbb{E}^{(\nu)}_{\boldsymbol{\tau}}[T-\nu+1|T\geq\nu]\end{aligned} \tag{36}$$

According to (35) and (36), applying Theorem 1 in (Ritov 1990) for a fixed $\boldsymbol{\tau}$ implies that

$$\sup_{\nu\in\mathcal{V}}\inf_{T\in\mathcal{T}} \mathbb{E}[c(T,\nu,\boldsymbol{\tau})] \;=\; \mathbb{E}[c(T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C)),\nu^*(C),\boldsymbol{\tau})] \tag{37}$$

where $\mu_1(C)$ is such that

$$\begin{aligned}&C - u_{\boldsymbol{\pi}}\mathbb{E}^{(\infty)}[T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C))]\\ &\quad= (u_{\boldsymbol{\pi}}-u_{\boldsymbol{\tau}})\mathbb{E}_{\boldsymbol{\tau}}[T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C))] \triangleq K(C,\boldsymbol{\tau}),\end{aligned} \tag{38}$$

and ν_* is defined as

$$P(\nu^*=t|\nu\geq t, U_1,\ldots,U_{t-1}) = p(C)(1-Q^{\boldsymbol{\tau}}_{t-1})^+. \tag{39}$$

In (39), $Q^{\boldsymbol{\tau}}_t = \exp(R^{\boldsymbol{\tau}}_t + l^{\boldsymbol{\tau}}_t)$ (see (5)), and $p(C)$ is some probability.

$K(C,\boldsymbol{\tau})$ is the expected post-attack cost that the adversary can guarantee. Then, choosing $\boldsymbol{\tau}$ that maximizes this cost, leads to the highest cost the adversary can guarantee. Let $\gamma_1 = \mathbb{E}^{(\infty)}[T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C))]$ .Note that as $C\to\infty$, $\mu_1(C)$ and also $\mathbb{E}^{(\infty)}(T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C)))$ go to ∞. As a result,

$$\mathbb{E}_{\boldsymbol{\tau}}[T^{\text{CUSUM}}_{\boldsymbol{\tau}}(\mu_1(C))] \sim \frac{\log\gamma_1}{D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi})},\quad \text{as } C\to\infty \tag{40}$$

Thus, according to (38), as $C\to\infty$ we have

$$u_{\boldsymbol{\pi}}\gamma_1 + (u_{\boldsymbol{\pi}}-u_{\boldsymbol{\tau}})\frac{\log\gamma_1}{D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi})} \sim C, \tag{41}$$

which implies that $K(C,\boldsymbol{\tau}) \sim (u_{\boldsymbol{\pi}}-u_{\boldsymbol{\tau}})\frac{\log C}{D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi})}$ as $C\to\infty$. Hence, optimal $\boldsymbol{\tau}$ for the adversary, is the one that maximizes $g(\boldsymbol{\tau}) = \frac{(u_{\boldsymbol{\pi}}-u_{\boldsymbol{\tau}})}{D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi})}$.

Lemma 2. *The function $g(\theta) = \frac{u_{\boldsymbol{\pi}}-u_\theta}{d(\theta)}$ is monotonically decreasing in θ.*

Proof. We show that $g'(\theta) < 0$. Note that we can write

$$u_\theta = \sum_{u\in\mathcal{U}} \boldsymbol{\tau}_\theta(u)u = \frac{\sum_{u\in\mathcal{U}} u\boldsymbol{\pi}(u)\exp(-\theta u)}{\sum_{u\in\mathcal{U}}\boldsymbol{\pi}(u)\exp(-\theta u)} = -b'(\theta) \tag{42}$$

Therefore, by using Lemma 1, we have

$$\begin{aligned}g'(\theta) &= \frac{-u_{\boldsymbol{\pi}}\theta b''(\theta)}{d^2(\theta)} - \left(\frac{-b'(\theta)}{\theta b'(\theta)-b(\theta)}\right)'\\ &= \frac{-b''(\theta)(\theta u_{\boldsymbol{\pi}} + b(\theta))}{d^2(\theta)}.\end{aligned} \tag{43}$$

$b(\theta)$ is strictly convex, and $b''(\theta) > 0$ (Lorden 1971). Thus, it is sufficient to show that $\theta u_{\boldsymbol{\pi}} + b(\theta) > 0$, which is equivalent to $\log(e^{\theta u_{\boldsymbol{\pi}}}\sum_{u\in\mathcal{U}}\pi(u)e^{-\theta u}) > 0$. Therefore, it is sufficient to show that

$$\sum_{u\in\mathcal{U}} \pi(u)e^{\theta(u_{\boldsymbol{\pi}}-u)} > 1 \tag{44}$$

Now, note that Jensen's inequality implies that

$$e^{\theta\mathbb{E}_{U\sim\boldsymbol{\pi}}[u_{\boldsymbol{\pi}}-U]} \leq \mathbb{E}_{U\sim\boldsymbol{\pi}}[e^{\theta(u_{\boldsymbol{\pi}}-U)}]. \tag{45}$$

We have $\mathbb{E}_{U\sim\boldsymbol{\pi}}[u_{\boldsymbol{\pi}}-U] = 0$. Thus, $\mathbb{E}_{U\sim\boldsymbol{\pi}}[e^{\theta(u_{\boldsymbol{\pi}}-U)}] \geq 1$. Since $\theta > 0$ and different values of $u\in\mathcal{U}$ are nonidentical, the equality does not hold, and we have $\mathbb{E}_{U\sim\boldsymbol{\pi}}[e^{\theta(u_{\boldsymbol{\pi}}-U)}] > 1$, which is essentially equivalent to (44). □

Proposition 5. *The maximizer of $g(\boldsymbol{\tau})$ is $\boldsymbol{\tau}_{\theta_{min}}$.*

Proof. Observe that according to Remark 1 for every $\boldsymbol{\tau}\in\mathcal{D}_\epsilon$, there exists a $\theta\in\Theta$ such that $u_{\boldsymbol{\tau}} = u_\theta$. For this θ, Proposition 4 implies that $D_{KL}(\boldsymbol{\tau}||\boldsymbol{\pi}) \geq D_{KL}(\boldsymbol{\tau}_\theta||\boldsymbol{\pi})$ (otherwise, $\boldsymbol{\tau}$ would have been the solution to (16)). Consequently, we have $g(\boldsymbol{\tau}) \leq g(\boldsymbol{\tau}_\theta)$. As a result, the maximizer of $g(\boldsymbol{\tau})$ lies within $\mathcal{F}_\Theta$.

Lemma 2 implies that $\theta_{\min}$ is the maximizer of $g(\boldsymbol{\tau}_\theta)$ within $\mathcal{F}_\theta$, and thus within $\mathcal{D}_\epsilon$. □

Now suppose that the adversary chooses $(\boldsymbol{\tau}_{\theta_{\min}}, \nu^*)$. As stated before, a Cusum stopping time with threshold $\mu_1(C)$ (defined specifically with respect to $\theta_{\min}$) is optimal. We claim that choosing $T^*(\mu(C))$ leads to the same cost for the defender asymptotically. For $\gamma > 1$, let μ^*_γ be defined as in (15) with $\alpha = 1/\gamma$. According to Proposition 1 as $\gamma\to\infty$, we have $\mathbb{E}^{(\infty)}[T^*(\mu^*_\gamma)] \sim \gamma$ and $\mathbb{E}_\theta[T^*(\mu^*_\gamma)] \sim \frac{\log\gamma}{d(\theta)}$. Consider the function $C(\gamma) \triangleq u_{\boldsymbol{\pi}}\gamma + \frac{\epsilon}{d(\theta_{\min})}log\gamma$. $C(\gamma)$ is a one-to-one monotonically increasing function of γ. Thus, as

γ varies monotonically from 1 to ∞, $C(\gamma)$ varies monotonically from $u_{\boldsymbol{\pi}}$ to ∞. Therefore, comparing $C(\gamma)$ and (17) implies that, $T^*(\mu(C))$ as $C \to \infty$ should be the same as $T^*(\mu^*_\gamma)$ as $\gamma \to \infty$. As a result, we have

$$\mathbb{E}_\theta[T^*(\mu(C))] \sim \frac{\log \mathbb{E}^{(\infty)}[T^*(\mu(C))]}{d(\theta_{\text{min}})}, \text{ as } C \to \infty \quad (46)$$

On the other hand, applying (41) for θ_{min}, we get: $C \sim u_{\boldsymbol{\pi}}\gamma_1 + \epsilon\frac{\log \gamma_1}{d(\theta_{\text{min}})}$ as $C \to \infty$. Hence, according to (17) and (46), we must have $\mathbb{E}^{(\infty)}[T^*(\mu(C))] \sim \gamma_1$. In other words, as $C \to \infty$:

$$\begin{aligned}\mathbb{E}^{(\infty)}[T^*(\mu(C))] &\sim \mathbb{E}^{(\infty)}[T_\theta^{CUSUM}(\mu_1(C))],\\ \mathbb{E}_\theta[T^*(\mu(C))] &\sim \mathbb{E}_\theta[T_\theta^{CUSUM}(\mu_1(C))] \end{aligned} \quad (47)$$

Since the total cost is determined uniquely by $E^{(\infty)}[T]$ and $E_\theta[T]$, $T^*(\mu(C))$ represents a defender's best response.

A.4 Proof of Theorem 2

Suppose that $\boldsymbol{\tau}_\theta \in \mathcal{F}_\Theta$ is fixed as the distribution chosen by the adversary. Let the T_1 be the optimal strategy for the defender, and that $\gamma = \mathbb{E}^{(\infty)}[T_1]$. Theorem 1 in (Lai 1998) implies that

$$\sup_\nu \mathbb{E}_\theta^{(\nu)}[T_1 - \nu + 1 | T_1 \geq \nu] \geq \frac{\log \gamma}{d(\theta)}, \text{ as } \gamma \to \infty. \quad (48)$$

On the other hand, for $T^*(\mu_\gamma)$ defined as in Section A.3, as $\gamma \to \infty$ we have $\mathbb{E}^{(\infty)}[T^*(\mu_\gamma)] \sim \gamma$, and

$$\begin{aligned}&\sup_\nu \mathbb{E}_\theta^{(\nu)}[T^*(\mu_\gamma) - \nu + 1 | T^*(\mu_\gamma) \geq \nu]\\ &\quad = \mathbb{E}_\theta[T^*(\mu_\gamma)] \sim \frac{\log \gamma}{d(\theta)}\end{aligned} \quad (49)$$

As mentioned in Section A.3, the total cost for the defender is a function of $\mathbb{E}^{(\infty)}[T]$ and $\mathbb{E}_\theta^{(\nu)}[T - \nu + 1 | T \geq \nu]$. Hence, (48) and (49) imply that $T^*(\mu_\gamma)$ is asymptotically the best response of the defender, as otherwise, the adversary can choose a ν that impose a higher cost. Moreover, as stated in Section A.3, $T^*(\mu_\gamma)$ as $\gamma \to \infty$ is equivalent to $T^*(\mu(C))$ as $C \to \infty$. Thus $T^*(\mu(C))$ is asymptotically a best response.

With $T^*(\mu(C))$ selected by the defender, Lemma 2 directly implies that the adversary can impose the highest possible cost by choosing θ_{min}. Moreover, $\nu = 1$ corresponds to the worst-case detection delay for the defender (see (49)).

B Additional Numerical Results

B.1 Tolerable Attack Impact

As mentioned in Section 5.1, in a practical scenario, the selection of ϵ and α is based on the player's tolerance for incurred costs and the acceptable false alarm rate. This decision is influenced by the maximum cost the adversary can impose within a known time horizon. We define the tolerable cost for the defender as the impact of an attack with θ_{min}, i.e., which represents the highest impact the adversary can achieve given the detector. Figure 1 shows the tolerable attack impact for two values of MTBFA as ϵ varies. As can be observed, a higher required MTBFA (or in other words, less probability of false alarms) necessitates accepting a slightly higher attack impact.

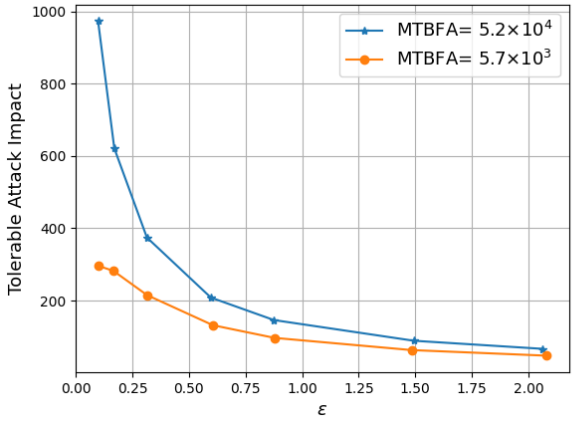


Figure 1: Tolerable attack impact as a function of ϵ for two fixed values of MTBFA in the Extended game of Chicken.

References

Lai, T. L. 1998. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information theory*, 44(7): 2917–2929.

Lorden, G. 1971. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 1897–1908.

Lorden, G. 1973. Open-ended tests for Koopman-Darmois families. *The Annals of Statistics*, 633–643.

Ritov, Y. 1990. Decision theoretic optimality of the CUSUM procedure. *The Annals of Statistics*, 1464–1469.

Wald, A. 1946. Differentiation under the expectation sign in the fundamental identity of sequential analysis. *The Annals of Mathematical Statistics*, 17(4): 493–497.