Licentiate Thesis in Electrical Engineering

Adaptive Measurement Strategies for Network Optimization and Control

SIMON LINDSTÅHL

Stockholm, Sweden 2023
Adaptive Measurement Strategies for Network Optimization and Control

SIMON LINDSTÅHL

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Licentiate of Engineering on Friday the 6th October 2023, at 2:00 p.m. in Q2, Malvinas Väg 10, Stockholm.

Licentiate Thesis in Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden 2023
Abstract

The fifth generation networks is rapidly becoming the new network standard and its new technological capabilities are expected to enable a far wider variety of services compared to the fourth generation networks. To ensure that these services can co-exist and meet their standardized requirements, the network’s resources must be provisioned, managed and reconfigured in a far more complex manner than before. As such, it is no longer sufficient to select a simple, static scheme for gathering the necessary information to take decisions. Instead, it is necessary to adaptively, with regards to network system dynamics, trade-off the cost in terms of power, CPU and bandwidth consumption of the taken measurements to the value their information brings. Orchestration is a wide field, and the way to quantify the value of a given measurement heavily depends on the problem studied. As such, this thesis addresses adaptive measurement schemes for a number of well-defined network optimization problems. The thesis is presented as a compilation, where after an introduction detailing the background, purpose, problem formulation, methodology and contributions of our work, we present each problem separately through the papers submitted to several conferences.

First, we study the problem of optimal spectrum access for low priority services. We assume that the network manager has limited opportunities to measure the spectrum before assigning one (if any) resource block to the secondary service for transmission, and this measurement has a known cost attached to it. We study this framework through the lens of multi-armed bandits with multiple arm pulls per decision, a framework we call predictive bandits. We analyze such bandits and show a problem specific lower bound on their regret, as well as design an algorithm which meets this regret asymptotically, studying both the case where measurements are perfect and the case where the measurement has noise of known quantity. Studying a synthetic simulated problem, we find that it performs considerably better compared to a simple benchmark strategy.

Secondly, we study a variation of admission control where the controller must select one of multiple slices to enter a new service into. The agent does not know the resources available in the slices initially, and must instead measure these, subject to noise. Mimicking three commonly used admission control strategies, we study this as a best arm identification problem, where one or multiple arms is ”correct” (the arm chose by the strategy if it had full information). Through this framework, we analyze each strategy and devise sample complexity lower bounds, as well as algorithms that meet these lower bounds. In simulations with synthetic data, we show that our measurement algorithm can vastly reduce the number of required measurements compared to uniform sampling strategies.

Finally, we study a network monitoring system where the controller must detect sudden changes in system behavior such as batch traffic arrivals or handovers, in order to take future action. We study this through the lens of change point detection but argue that the classical framework is insufficient for capturing both physical time aspects such as delay as well as measurement costs independently, and present an alternative framework which
decouples these, requiring more sophisticated monitoring agents. We show, both through theory and through simulation with both synthetic data and data from a 5G testbed, that such adaptive schedules qualitatively and quantitatively improve upon classical change point detection schemes in terms of measurement frequency, without losing classical optimality guarantees such as the one on required measurements post change.

**Keywords:** Network optimization, network management, network orchestration, multi-armed bandits, admission control, change point detection, statistical learning
Sammanfattning


Slutligen studerar vi ett övervakningssystem där agenten måste upptäcka plötsliga förändringar i systemets beteende såsom förändringar i trafiken eller överräckningar mellan master, för att kunna agera därefter. Vi studerar
detta med ramverket förändringsdetektion, men argumenterar att det klassiska ramverket är otillräckligt för att bemöta aspekter berörande fysisk tid (som fördröjning) samtidigt som den bemöter mätningarnas kostnad. Vi presenterar därmed ett alternativt ramverk som frikopplar de två, vilket i sin tur kräver mer sostiferade övervakningssystem. Vi visar, genom både teori och simulering med både syntetisk och experimentell data, att sådana adaptiva mätscheman kan förbättra mätfrekvensen jämfört med klassiska periodiska mätscheman, både kvalitativt och kvantitativt, utan att förlora klassiska optimalitetsgarantier såsom det på antalet mätningar som behövs när förändringen har skett.
List of Papers

I Predictive bandits
Simon Lindstälh, Alexandre Proutiere, Andreas Johnsson
Published in Proc. of IEEE Conference on Decision and Control (2020)

II Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach
Simon Lindstälh, Alexandre Proutiere, Andreas Johnsson
Published in Proc. of IEEE Global Communications Conference (2022)

III Change point detection with adaptive measurement schedules for network performance verification
Simon Lindstälh, Alexandre Proutiere, Andreas Johnsson
Submitted

List of papers not included in thesis

I Reinforcement Learning with Imitation for Cavity Filter Tuning
Simon Lindstälh, Xiaoyu Lan
Published in Proc. of 2020 IEEE/ASME International conference on Advanced Intelligent Mechatronics (2020)
Acknowledgement

First and foremost, I want to extend my most sincere gratitude towards my main academic supervisor Alexandre Proutière and my industrial supervisor Andreas Johnsson for their continued support, inspiration, guidance and feedback. Navigating a field so deeply branched within both theory and industry has been challenging and would not have been possible without expertise from both academy and industry. Merging the two ways of thinking has not always been easy but it is thanks to the perseverance of my supervisors that we have gotten as far as we have.

I would also like to thank my colleagues at the Research Area for Artificial Intelligence at Ericsson, in particular my section members at Data Science and Automation, for continuing to support and include me even when I have been distant. I want to give special thanks to Carmen Lee, Maxime Bouton, Jalil Taghia, Ahmad Terra and Filippo Vanella for inspiring discussions and feedback on several aspects of my research. I also want to extend a special thank you to Hassam Riaz for providing a great deal of guidance regarding testbed experiments.

Similarly, I want to give thanks to my colleagues at KTH, particularly the members of the statistical learning and control group for continued inspiration and discussion, as well as my office mates Alexander Johansson, Ting Bai and Ines Lourenco for providing a great working environment.

Of course, I want to thank the Knut and Alice Wallenberg foundation, who have partially funded the work of this thesis through the Wallenberg Artificial Intelligence, Autonomous systems and Software Program (WASP).

Finally, I want to give a warm thanks to my friends and family for supporting me through the tougher parts of my work. A special thanks goes to Johan Nordstrand for his continued support both emotionally and intellectually, as well as to all my friends in Kårsbobet, without the motivation from which I surely would have given up on this project a long ago.
## Contents

<table>
<thead>
<tr>
<th>List of Papers</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgement</td>
<td>vii</td>
</tr>
<tr>
<td>Contents</td>
<td>1</td>
</tr>
</tbody>
</table>

### 1 Introduction
- 1.1 5G networks | 3
- 1.2 Sequential decision making | 6
- 1.3 Outline of remainder of thesis | 8
- 1.4 Aim of thesis: Analysis, design and evaluation of measurement schemes | 8
- 1.5 Studied problems in network optimization and control | 9
- 1.6 Methodology | 12
- 1.7 Contributions | 13
- 1.8 Conclusion and future work | 14

### 2 Paper I: Predictive Bandits
- 2.1 Introduction | 18
- 2.2 Related Work | 19
- 2.3 Models and Preliminaries | 20
- 2.4 Regret Lower Bounds | 22
- 2.5 Algorithms | 26
- 2.6 Numerical Experiments | 30
- 2.7 Conclusion | 31

### 3 Paper II: Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach
- 3.1 Introduction | 34
- 3.2 Related work | 35
- 3.3 Models: Dynamics, Admission Control, and Measurement Costs | 36
- 3.4 Best arm identification in admission control | 40
- 3.5 Numerical Results | 43
- 3.6 Conclusions | 45
## 4 Paper III: Change point detection with adaptive measurement schedules for network performance verification

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>48</td>
</tr>
<tr>
<td>4.2 Related work</td>
<td>50</td>
</tr>
<tr>
<td>4.3 Change point detection and adaptive measurement schedules</td>
<td>52</td>
</tr>
<tr>
<td>4.4 Analysis and algorithms</td>
<td>56</td>
</tr>
<tr>
<td>4.5 Numerical experiments</td>
<td>61</td>
</tr>
<tr>
<td>4.6 Conclusion</td>
<td>72</td>
</tr>
<tr>
<td>4.A Proofs</td>
<td>72</td>
</tr>
<tr>
<td>4.B Parameters for numerical evaluation</td>
<td>79</td>
</tr>
<tr>
<td>4.C Table of notations</td>
<td>80</td>
</tr>
</tbody>
</table>

## A Omitted proofs

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Proofs from Paper I</td>
<td>83</td>
</tr>
<tr>
<td>A.2 Proofs from Paper II</td>
<td>89</td>
</tr>
</tbody>
</table>

## References

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 5G networks

In order to meet increasing demands on connectivity, the fifth generation (5G) networks started deployment in 2019. While 5G is able to increase performance of typical 4G services (such as voice services and internet access) through enhanced mobile broadband (eMBB) its main motivator is the ability to provide a wider variety of services. The two key use case areas for such services are ultra reliable low latency communication (URLLC), useful for remote control of autonomous control of autonomous vehicles and process automation, as well as massive machine-type communication (mMTC), useful for Internet of Things applications and smart cities. 5G is enabled by a variety of new network technologies, the key enablers being mmWave new radio (NR) and softwarization of network components through software-defined networking (SDN) and network function virtualization (NFV) [1]. An overview of 5G infrastructure can be seen in Figure 1.1.

While 5G can operate on typical 4G frequencies, its main benefits are unlocked when using millimeter-length (mmWave) frequencies in the range of 24-29 GHz. These frequencies greatly increase transmission speed and reduce air latency to the range of 8-12 milliseconds. Furthermore, the reduced minimum size of antennas for high frequency waves unlocks the ability to shape and reduce the width of the transmission beam, a practice known as beamforming, allowing multiple beams of the same frequency to coexist within a network cell with minimal co-interference. This, in turn, allows for multiplexing done not only in time or frequency space, but also through positioning, known as spatial multiplexing. Using sufficiently many antennas, this allows the practice of massive multiple-input multiple output transmission, or massive MIMO for short. The shorter wavelength does, however, come with a reduction of range and penetration ability, requiring more antennas in the network. That said, this is not only a drawback as the shorter range allows for small, personalized cells suited for
different applications and services.  

Classically, backhaul networks have been employed using specialized routers and switches routing using distributed protocols. However, in the 2010s new approaches to network management emerged. By separating the data flowing through a network from the control messages used to manage it, it is possible to create visible and programmable networks. This practice is known as software-defined networking. Furthermore, due to increased computational capabilities, network functions such as routing or firewalls can be implemented as software functions on general purpose hardware, allowing for flexible redesign of networks as well as future-proofing by installing new functions on old hardware. This has come to be known as network function virtualization, as the softwareized network functions are referred to as virtual network functions (VNFs). Of course, there is no reason these functions have to be limited to classical network functions, and the use of general purpose hardware makes it convenient to run computational services, or edge computing, on this hardware in order to reduce total network delay. Importantly, a network can now be implemented using only part of the hardware and bandwidth resources, and as such, multiple logical communication networks could exist on the same physical resources. This is known as network slicing and is useful for creating networks which are optimized for certain types of services, as the widely varying demands are difficult to meet with the same logical network.

Network orchestration

With the new technologies, new services emerge. But these technologies create new challenges in network management. In particular, the virtualization of network functions create questions of how such functions can be mapped to physical resources, represented and configured. With the introduction of network slicing,
it is now relevant to divide such network functions or physical resources between slices in a way that is fair or optimal. The introduction of cloud-based services pose issues of whether computations should be performed within the cheap but high latency central cloud data centers, or if they should be performed within the low-latency but expensive edge compute resources. Finally, service providers must ensure that they meet established end-to-end performance requirements while saving cost and energy wherever possible. These questions comprise the field of network orchestration, and its importance only grows as networks increase in complexity [6].

Some examples of orchestration tasks are traffic engineering, accounting, load balancing, elephant flow detection and change detection. Traffic engineering is a wide field ranging from admission control of new flows to rerouting of existing flows [7]. Load balancing is the problem of ensuring that no single compute resource within a network is overloaded [8]. Elephant flow detection is the problem of detecting when one flow is using a large amount of network resources, in order to either optimize the network accordingly (if the flow is legitimate) or detect an attack (if not) [9]. Finally, change detection is the problem of detecting rapid network changes as quickly as possible without an excess of false alarms, so that fixing measures can be taken [10].

**Monitoring in network orchestration**

Shared between all network orchestration problems is the need of information about the network state to function - how can one hope to balance loads if the utilization of network paths is unknown, for instance? As such, it is necessary to monitor networks, which we define as gathering and maintaining information about network resources as well as the flows existing within the network. Such monitoring can be done either passively, by receiving information about traffic and resources as it arrives through other means or actively, by probing the network to evaluate its function and performance. Previously, monitoring has been done either by observing high-level statistics of the network [11] or by monitoring each flow individually [12]. However, high-level statistics are insufficient to solve orchestration problems when flows can be highly heterogeneous, and monitoring each flow individually becomes prohibitively expensive even in moderately large networks (such as a campus network) [13]. It is thus necessary to create new monitoring solutions to fit these new scenarios.

Measurements are, however, costly. This is especially true in modern networks, where the focus has shifted from maximizing network statistics to maximizing the end-to-end performance statistics for users. However, such statistics are challenging to calculate on a network level, although some progress has been made in recent years [14]. Instead, at least part of the network is probed from the user’s side, which could be expensive in terms of bandwidth and CPU power. Maintaining as few measurements as possible, while of course still making good orchestration performance, becomes both critical and challenging.
There have traditionally been two schools of thought in network monitoring, which are passive and active monitoring [15]. In passive monitoring, the network monitor logs user and network data and analyzes it either continuously or in hindsight. It is considered passive because no extra packets are being inserted into the network as measurement probes, instead it can use measurements that UEs and network components would have taken anyway. However, there is still an associated cost with pulling and logging measurement data from different UEs and network components into one centralized network controller. Furthermore, this has to be pulled relatively seldom in order to save any costs compared to active monitoring. As such, passive monitoring is well suited to keep track of performance for different users, but it can potentially be too slow to handle quick network failures, and often fails to monitor building network problems and predict future failures. Active measurements, by contrast, simulate user behavior by probing the network for different statistics, often RTT or throughput. These active measurements can be used to preemptively make changes to the network if a fault is detected or if a statistic is trending in the wrong direction, before these changes would have any impact on performance. However, because these probes use a different transport protocol compared to normal traffic, they do not necessarily show an accurate picture of the QoE of actual users in the network. Furthermore, these probes can compete with the real network traffic, imposing a potentially significant load on the network.

In 2014, Van Adrichem et al. introduced one of the first comprehensive SDN-based monitoring frameworks, called OpenNetMon [13]. It operates by polling individual flows for throughput, polling switches for packet loss, and doing path-based polling for delay. It can thus be seen as an active approach. Mardani et al. used network tomography to estimate traffic and anomaly maps of networks [11], taking a more passive approach while still reducing measurement costs. Shirali-Shareza and Ganjali created the flow collection framework FleXam which, based on wild-card rules, allowed for coarse-grained or fine-grained statistics collection. They did not, however, suggest any specific solutions to balance the tradeoff of information gathering to measurement cost [12]. In 2017, Kim et al. used clustering to passively monitor changes in SDN networks [16]. In 2018, Wang and Su created a more advanced flexible monitoring scheme FlexMonitor, taking advantage of the programmability of SDN, and evaluated their approach through DDoS detection [17]. In 2019, Xie et al. created a joint monitoring and service assurance framework, arguing that monitoring should not be done in a vacuum and should instead be performed with regards to some networking task, in this case service assurance [18].

1.2 Sequential decision making

Sequential decision making (SDM) is a wide term within artificial intelligence and control theory, referring to any system where a decision making agent is able
to delay all or parts of its decisions until it receives more information. Typical SDM problems are discrete-time in nature, as they require a new decision for each new batch of information. Some classical examples of SDM problems are optimal stopping problems, multi-armed bandits, sequential hypothesis testing and change point detection. Generally, all Markov Decision Processes, with or without a known transition model, can be considered SDM problems [19].

When designing a passive or active measurement scheme, one can design the scheme in advance and employ only the decision strategies sequentially. When designing an adaptive measurement schedule, however, it is important to take previous measurements into account before deciding on the near-future measurement schedule. As such, any decision problem with adaptive measurement schedules will be an SDM problem. In this thesis, we are particularly interested in two sub-branches of SDM problems: multi-armed bandit problems as well as change point detection problems.

In multi-armed bandit (MAB) problems, an SDM agent is presented with multiple random variables with different, unknown distributions; these random variables are denoted as arms. The arms have different expected outcomes and ideally, the agent wishes to choose as often as possible choose the arm with the greatest mean. However, since the distributions are initially unknown, they must be learned through sampling, which creates an inherent trade-off between exploration (learning the best arm by studying all arms) and exploitation (pulling the believed best arm as often as possible in order to avoid losing out on reward). This trade-off is typically modeled as minimizing the expected cumulative difference between the reward of the best arm and the reward of the studied agent, known as regret. That said, it has become increasingly common to study so called best arm identification (or pure exploration) problems, where the agent wishes to output the best arm as quickly as possible with a bounded probability of error, without regard to incurred loss during the learning process [20].

Multi-armed bandit problems are ideally analyzed in terms of lower bounds, imposed on any agent which performs sufficiently well on a wide class of problems. In classical MAB problems, theses lower bounds are on the regret, while for best arm identification they bound the sample complexity as a function of the probability of error. There exist interesting minimax (worst case) bounds for both problems, but the most illuminating lower bounds tend to be problem-specific, as they carry fundamental structural properties of the problem in question. The dependence on time (in the case of regret) or error probability (in the case of sample complexity) is well known for a wide class of problems through fundamental change-of-measure arguments, but there remains an interesting dependence on problem parameters, which typically needs to be solved as an optimization problem on a problem-to-problem basis. Once these lower bounds are known, they inform the agent of the proportions of measurements necessary for each arm (or sub-optimal arm, in the case of regret minimization), and can therefore be used to devise optimal algorithms [20].

In change detection, an SDM agent instead studies a single random variable,
which at some point changes in distribution. The agent would like to detect this change as quickly as possible while avoiding false alarms, which creates a trade-off when measurements are noisy. Classically, then, these change detection problems are examples of optimal stopping problems, which has a wide branch of theory attached. Generally, change point detection problem can be considered either from a Bayesian view or a frequentist view \[10\].

In Bayesian change detection, the change time is assumed to have a known distribution, typically exponential to create memoryless change. Then, the false alarm rate and expected delay can be traded off explicitly, as they are both well-defined quantities with regards to the change time distribution. This creates an interesting stopping time MDP, where the state is the belief that a change has occurred, updated with Bayes rule.

In frequentist change detection, the problem is instead considered minimax. The goal is to minimize the worst possible delay over all change times while maintaining a sufficiently great time to false alarm.

### 1.3 Outline of remainder of thesis

In section 1.4 we outline the purpose of this thesis work and present the research questions under consideration. These research questions are asked with respect to three different orchestration problems, presented in section 1.5. In section 1.6 we describe the methodology used to study and solve the presented network orchestration problems. In section 1.7 we describe the analysis, designed algorithms and results that contribute to the state of the art, and detail the exact contributions by the main author. In section 1.8 we conclude the introduction by answering the research questions posed in section 1.4 and present opportunities for future work, both with regards to each orchestration problems as well as more generally. Finally, in chapters 2, 3 and 4 we present the published and submitted papers included in this thesis, which go into more detail about the work done with these orchestration problems.

### 1.4 Aim of thesis: Analysis, design and evaluation of measurement schemes

In this thesis, we aim to design joint measurement and decision algorithms for a variety of network orchestration problems. We believe that there is no single monitoring algorithm which is measurement cost efficient while guaranteeing orchestration performance. Instead, the value of information of a given measurement heavily depends on how useful it is to determine the correct action in any given problem, inextricably linking measurement and decision algorithms to each other. Even if they would be implemented as separate agents, they would need to communicate often in order to ensure good joint performance. Preferably, the designed algorithms should be optimal in some way, exhausting the research in
its given problem and paving way for future research. We make no claims on covering all possible, or even all relevant orchestration problems within the scope of this thesis. Instead, we will focus on three problems, already studied as pure decision problems in literature, and show the impact on these problem from introducing the adaptive measurement aspect into them. These problems are specified in detail in section 1.5.

The research questions posed in this paper are thus as follows, for all studied orchestration problems:

1. What fundamental bounds exist on measurement costs, for an agent maintaining a sufficiently good orchestration performance?

2. Can we reach this bound by designing a joint measurement and decision algorithm? If so, under which circumstances? If not, what performance can we reach?

3. How much worse compared to the lower bound does a baseline measurement scheme, such as uniform sampling, perform? How much worse does such a baseline perform compared to the algorithm designed in question 2?

The first research question aims to elucidate the problem dependent nature of monitoring and information, by showing how different problems require information to be gathered in different ways. The second question aims to make our research practical, by designing algorithms that can be used in orchestration should the need arise. The third question aims to validate the need for the other two research problems, by investigating the gap in terms of measurement costs between agents with simple measurement schemes compared to ones with intelligent measurement schemes.

With respect to existing network structures, our aim is to help network orchestrators fulfill their purpose of controlling all layers of a network by gathering necessary but not superfluous measurements. The placement of our work compared to a conceptual view of 5G networks is shown in Figure 1.2.

1.5 Studied problems in network optimization and control

Network optimization and control is a vast area with a wide variety of subtopics to explore. Within the scope of this thesis, we have studied the following three problems.

**Opportunistic Spectrum Access: A Predictive Bandit Approach**

We begin by investigating the problem of spectrum sharing, specifically that of identifying viable secondary carriers [21]. Often is the case that some priority service occupies a large number of primary carriers, but that it is possible to utilize secondary carriers at uneven intervals. It is not, however, always known
to when these secondary carriers are available, or even what distribution each carrier follows. It is, however, often possible to probe a carrier to see whether it is available, typically subject to some measurement noise. These measurements are generally costly as they involve probing the spectrum and thus consuming bandwidth and/or user equipment power. Furthermore, decisions typically need to be taken fast and so, it may be prohibitive to take more than one such measurements. It is plausible to extend the problem to multiple sequential or batch measurements taken before the connection attempt, but it is not immediately clear what value this would bring.

Figure 1.2: A view of 5G networks as a three-layer model, complete with a network controller. The placement of our work in this model is highlighted in yellow.

We are thus interested in creating a learning agent that can adapt to stationary but unknown availability distributions on the secondary carriers. In doing so, the agent should learn whether it is more worthwhile to measure some available carrier before selecting the same - or another - to attempt to connect unto, or whether it is more worthwhile to immediately attempt a connection without utilizing measurements. Note that since the distributions are stationary, there will be a single policy, defined by either a carrier or a 2-tuple of carriers, that is optimal at all times. This is then an extension of classical multi-armed bandit (MAB) problems, and we similarly aim to minimize the regret incurred while learning. This regret is defined as the expected cumulative loss of objective compared to an agent which already knows the optimal policy. We aim to identify the form of an optimal policy and to create an optimal algorithm, preferably with finite time guarantees, by establishing a lower bound on the regret for any uniformly good (i.e. successful on several problems) algorithms and verifying that the regret of our created algorithm does not exceed that of the such derived lower bound.
Modelling admission control in sliced networks with best arm identification

While sliced networks are designed as to allow different services with different requirements to co-exist on the same infrastructure, it is often the case that the same service could function on multiple slices. In these cases, an admission control agent must decide which slice to admit an incoming flow to, if any. This then poses a problem of admission control with multiple, distinct servers and multiple heterogeneous service classes. Such problems are difficult to solve as they do not have the property of insensitivity [22]. As such, it is not immediately clear what the optimal strategy for admission would be, but several heuristic strategies are available. To utilize any strategy, however, the remaining available resources must be identified, with a precision dependent on the strategy employed. In several scenarios, such as those where many tenants share the network in a non-cooperative manner, the network state could change rapidly and information of the available resources could become stale between decisions. As such, measurements must be taken in order to sufficiently well identify the available resources, but these measurements can often be expensive if they (such as probes) demand information about the end-to-end performance.

As such, we aim to create measurement agents which can identify a correct decision according to whatever strategy we are using, up to a probability of error $\delta$, with as few measurements as possible. The framework for these agents should be applicable to a vast number of strategies, but in the context of this paper, we focus on a number of commonly used examples. These are:

1. **Any-available slice**: identify any slice with sufficiently low load as to accommodate the incoming flow. This should be the strategy requiring the fewest measurements, but will not have any hard guarantees on flow-level performance.

2. **Packing slice**: identify the slice with the greatest, but still sufficiently low load. This serves the dual purpose of leaving space open for large incoming flows whenever possible as well as maximizing the time that servers can sleep (i.e. have a load of zero). However, since many services will stay in nearly overloaded slices, the performance of each service could be degraded.

3. **Least-loaded slice**: contrasting to the above, identify the slice with the smallest load (or none, if all slices are overloaded). This is the best effort attempt to maximize performance of admitted services, but lacks both of the mentioned benefits of packing slice.

This problem greatly resembles that of **best arm identification**, that of identifying the distribution out of many with the greatest mean with as few samples as possible, up to a margin of error. The differences to the classical framework is twofold: the correct output is not necessarily the arm with the greatest (or
lowest) mean, and in the case of any-available slice, there could be multiple correct outputs. As in that problem, we aim to minimize the sample complexity of the identification algorithm on a wide variety of problems, expected to be some problem-dependent constant multiplied by $d_B(\delta, 1 - \delta)$, the symmetric Bernoulli Kullback-Leibler divergence of distributions with means $\delta$ and $1 - \delta$.

Adapting measurement schedule to optimize monitoring with change point detection

While orchestrating network resources and managing the quality of end users, it is important to monitor performance in case of anomaly events. These events may take known forms, such as batch traffic arrivals or component failures. This is the case considered in this work, but extensions to cases where the anomaly events belong to some known family while being unknown themselves are possible. In the considered case, the anomaly event can be considered as a change in some monitored random variable (such as measured round-trip times, subject to noise). This can then be handled by using classical frequentist change point detection methods. However, these methods make no distinction between number of measurements taken and physical time, forcing us to infer that they take measurements periodically. As one can easily realize, this is not always an optimal strategy if there is a need to minimize measurement costs.

As such, we consider the case where the agent, other than raising alarms when it believes a change has occurred, can also increase or decrease its measurement frequency at will. We model this as letting the agent set a measurement period $\tau_n$ after measurement $n - 1$, having the ability to choose actions as any non-negative real number. This extends the problem beyond its classical status as an optimal stopping problem, having now to also consider a valued action at each time. Classically, agents must minimize their expected number of post-change measurements while maintaining a high average run length to false alarm. We also enforce this, adding extra dimensions by our introduction of $(\gamma, \beta)$-compliant agents. These agents must maintain an average run time to false alarm $\gamma$ and a worst case expected delay no greater than $\beta$, both measured in physical time. Doing so, we wish to minimize both the pre-change measurement frequency (or equivalently, maximize the expected measurement periods when no change occurs) as well as the number of measurements post change, while maintaining $(\gamma, \beta)$-compliance.

1.6 Methodology

For all three problems, we have divided the research into three sub-problems. First, we have designed a new framework, including the new aspects of adaptive measurements as compared to previous problems in the literature. We have also analyzed these frameworks and established critical aspects such as lower bounds
on the objective cost. Secondly, we have designed and analyzed algorithms to solve these problems with minimal costs, comparing them to the designed lower bounds and attempting to meet them asymptotically (as information and problem requirements both grow large) or even in finite time. Finally, we design simulators to evaluate the performance of these algorithms numerically, comparing them to naive benchmarks as would have been employed without analysis. In the case of change point detection, we also evaluate the designed algorithm on data generated from a 5G measurement testbed.

1.7 Contributions

**Paper I - Predictive bandits:** We have developed a new framework for single predictive arm measurements, showing that doing so introduces a new notion of regret and changes the problem form to a new, non-independent sequential analysis problem. We have shown a new regret lower bound for this problem under the framework of uniformly good algorithms. Based on similar literature, we have designed a new algorithm to exploit the framework and shown its asymptotic optimality. We have also evaluated it compared to a naive benchmark in simulation, using synthetic data.

**Paper II - Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach:** We have evaluated several common admission control strategies with applicability in 5G networks and designed them in the context of best arm identification, to deal with uncertainties in network resource knowledge. We have shown how best arm identification is designed for these strategies, by designing new best arm identification problems for each strategy and showing the problem specific lower bounds for each. We have utilized these lower bounds for the strategies by using the well-known but difficult to use algorithm Track-and-Stop and creating asymptotically optimal algorithms for each. We have further evaluated these strategies compared to a naive benchmark in simulation, using synthetic data.

**Paper III - Change point detection with adaptive measurement schedules for network performance verification:** We have reformulated the classical change point detection problem with a notion of physical time for the purpose of network monitoring, which changes the trade-off from a two-way trade-off between delay and false alarm rate to a four way trade-off between delay, false alarm rate, number of post-change measurements and pre-change measurement frequency. We have shown that classical algorithm have undesirable properties when measurement frequency is a concern, and propose a simple alternative that do not have these properties. We have explicitly designed new algorithm with performance guarantees. We have evaluated these strategies compared to a naive benchmark in simulation, using both synthetic data as well as data gathered in a 5G testbed specifically for this experiment. We showed that our guaranteed
algorithm outperforms even the best possible classical algorithm in terms of pre-change measurement frequency, with or without guarantees.

**Contribution of main author**

**Paper I - Predictive bandits:** I (the main author) have developed the new framework under supervision. I have independently derived all theoretical analysis found in this paper, and proven all necessary statements. Based on earlier work within the area, I have designed the algorithm used in theory and evaluation. I have designed and implemented the simulator used for evaluation, and used it to run all evaluation simulations, gathering and presenting all necessary data. I have acted as main author and written the majority of the text within the paper. I have compiled reviewer comments and refined the paper based on them, as well as virtually presenting the paper at CDC 2020.

**Paper II - Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach:** I (the main author) have developed the new framework under supervision. I have independently derived all theoretical analysis found in this paper, and proven all necessary statements. Based on earlier work within the area, I have designed the algorithm used in theory and evaluation. I have designed and implemented the simulator used for evaluation, and used it to run all evaluation simulations, gathering and presenting all necessary data. I have acted as main author and written the majority of the text within the paper. I have compiled reviewer comments and refined the paper based on them, as well as virtually presenting the paper at GLOBECOM 2022.

**Paper III - Change point detection with adaptive measurement schedules for network performance verification:** I (the main author) have independently developed the new measurement framework. I have independently derived all theoretical analysis found in this paper, and proven all necessary statements. I have independently designed the intelligent measurement algorithm used in theory and evaluation. I have designed the scenario used to gather test bed data and implemented it in the in-house test bed. I have collected this RTT data together with a co-worker. I have designed and implemented the simulator used for evaluation, and used it to run all evaluation simulations, gathering and presenting all necessary data. I have acted as main author and written the majority of the text within the paper.

### 1.8 Conclusion and future work

Returning to our research questions, we have answered them and obtained the answers specified below.
We have described fundamental performance bounds, in terms of measurement cost compared to network orchestration performance, for all three problems to some extent and in full for the first two problems.

For the first two problems, we have designed optimal joint measurement and decision algorithms (or near-optimal, in the case of admission control). For the problem of changepoint detection, we have designed a potentially suboptimal algorithm which still greatly outperforms benchmark schemes.

For the bandit-based problems, there is a large optimality gap between the performance lower bound and the benchmark schemes, which is made smaller or zero by our algorithms. For the changepoint detection problem, the benchmark schemes do not perform order-optimally, while our algorithm does so.

**Future work**

Other than extending the discussed problems, a vast variety of other orchestration problems could be considered through the lens of adaptive measurements, and could potentially benefit from them.

Admission control, in particular, can be viewed in several different ways other than the one considered in our work. In particular, if one has control over the entire network (such as in private networks) the information will no longer decay to the extent where it has to be relearned at each decision point. Furthermore, measurements could be taken even before a new unit arrives. It could therefore be possible to design agents which monitor the network separately from the one which controls admission, although the two will need to maintain a shared knowledge base.

More network specific problem can and should also be studied. For instance, network slices should be dynamically reconfigured and rescaled to adapt to changing traffic patterns. This, of course, requires monitoring of both traffic patterns and available slice resources, the latter typically taking the form of some graph. It would be interesting to study how to monitor in an adaptive way, in particular minimizing the total control plane load of network statistics given some (logically) centralized controller.

Load balancing is another such network problem to be studied. While tackled to some extent through the work in Chapter 3, it would be more interesting to handle load balancing of servers and routes to servers within any given slice, again minimizing the measurement costs.
Chapter 2

Paper I: Predictive Bandits

Foreword

Starting with this research work, we were initially surprised by how open the field was. Indeed, the idea of adaptive measurement approaches was and is still not particularly well-researched either in theory or in applications. As such, a natural first step in the research was to attempt to understand how measurement costs could be tied to the value of information in a simple sequential decision problem. For this purpose, opportunistic spectrum access was chosen, as it can be tied to well-understood bandit problems and there is little need for dynamic system modeling. The ideas behind this paper would later be tied, through several steps of iteration, to those of Paper II, where instead of using classical bandit frameworks we would apply best arm identification in network management.

Obviously, the model of using only a single predictive measurement before making a decision is rather simplistic. Before settling on this model, we attempted to analyze models where multiple measurements, taken either in batches or sequentially. Indeed, in the case with sequential noiseless measurements, we were able to derive partial results, which we decided not to include in the paper as they were not full results and did not allow for a meaningful extension to the existing algorithms. Noisy measurements, as well as measurements taken in batches, turned out to be rather difficult to analyze.

Because of page limits in the publication into the proceedings of CDC 2020, we were forced to omit some of the proofs within this paper. These proofs can be found in Appendix A.

Paper I is reprinted, with permission, from [23]. The text has been reformatted, but no text has been changed. ©IEEE 2020.
Predictive Bandits
Simon Lindståhl, Alexandre Proutiere, Andreas Johnsson

Abstract

We introduce and study a new class of stochastic bandit problems, referred to as predictive bandits. In each round, the decision maker first decides whether to gather information about the rewards of particular arms (so that their rewards in this round can be predicted). These measurements are costly, and may be corrupted by noise. The decision maker then selects an arm to be actually played in the round. Predictive bandits find applications in many areas; e.g. they can be applied to channel selection problems in radio communication systems, or in recommendation systems. In this paper, we provide the first theoretical results about predictive bandits, and focus on scenarios where the decision maker is allowed to measure at most one arm per round and the rewards are Bernoulli random variables. We derive asymptotic instance-specific regret lower bounds for these problems, and develop algorithms whose regret match these fundamental limits. We illustrate the performance of our algorithms through numerical experiments. In particular, we highlight the gains that can be achieved by using reward predictions, and investigate the impact of the noise in the corresponding measurements.

2.1 Introduction

In this paper, we introduce and study a new class of stochastic bandit problems, referred to as predictive bandits. In the classical stochastic Multi-Armed Bandit (MAB) problem [24], the decision maker selects an arm in each round, and observes a realization of its random reward. The average rewards of the arms are initially unknown, and the objective of the decision maker is to devise a learning algorithm maximizing its reward accumulated over time. In predictive bandits, in each round, the decision maker may before actually playing an arm gather information about the rewards of particular arms in this round. By measuring an arm, she can predict to some extent its outcome. Measurements however come with a (fixed and known) cost, and may be corrupted by noise. As in classical stochastic MAB problems, the average rewards of the various arms are initially unknown, which forces the decision maker to explore sub-optimal arms. With predictive bandits, she has the additional difficulty of learning whether measuring arms yield better accumulated rewards, and in that case, which arms should be measured.

Predictive bandits bear similarities with contextual bandits [25], where the decision maker observe feature vectors associated with each arm before playing an arm. Contextual bandits were motivated by the design of personalized recommender systems (the context may include information about both the items to be recommended and the user currently requesting a recommendation), and have been applied to the design of various web-based services. Contextual bandits
differ from predictive bandits since in the latter, the observation of the context is not free, and the decision maker needs to decide which part of the context (which arm) she wishes to observe if any. Beyond web-based services and recommender systems, predictive bandits can be also applied to numerous resource allocation problems in communication networks. For example, in the channel selection problems in radio communication systems (see e.g. [26, 27] and references therein), the transmitter needs to choose from several radio channels, with randomly varying conditions and unknown means. One may measure the state of a channel (using probe packets) before choosing a transmission channel, but acquiring this information is consuming time and power (i.e. it has a cost).

In this paper, we provide the first theoretical results on predictive bandits. We consider problems where the decision maker is allowed to measure at most one arm per round. In the aforementioned radio channel selection problem, such a scenario is motivated by the fact the transmitter may not have time to measure several channels without breaking the required delay guarantees of the underlying application. For predictive bandits with at most one measurement per round, our contributions are as follows.

(a) We derive asymptotic instance-specific regret lower bounds. These bounds constitute fundamental performance limits that no learning algorithm can beat, but they also provide insights into the design of efficient algorithms. Indeed, the lower bounds specify the optimal exploration process, i.e., the rates at which an optimal algorithm should explore sub-optimal actions. These rates depend on the average rewards of the arms and on the measurement cost. Hence, an algorithm following these exploration rates would truly and optimally adapt to the actual problem parameters.

(b) We present simple algorithms that rapidly learn the optimal action, and that in fact, match our regret lower bounds. These algorithms leverage KL-UCB indices [28] to explore sub-optimal actions, and critically rely on an aggressive exploitation strategy (in each round, the best empirical action is played with a strictly positive probability). Our main technical contribution is to establish that such an aggressive exploitation behavior is indeed asymptotically optimal. We believe that this result is general, and could be extended to many bandit problems.

(c) We illustrate the performance of our algorithms through numerical experiments. In particular, we highlight the gains that can be achieved by using reward predictions, and investigate the impact of the noise in the corresponding measurements.

2.2 Related Work

Stochastic bandit problems have been extensively studied. In their seminal paper [24], Lai and Robbins derived asymptotic regret lower bounds and proposed
algorithms achieving these fundamental limits. In [29], the authors proposed UCB, a very simple and popular algorithm approaching regret lower bounds and for which a finite-time regret analysis is possible. Another attractive algorithm, KL-UCB, inherits the simplicity of UCB, and has first been shown to be asymptotically optimal in [28]. Later, [30] proposed a finite-time analysis of the algorithm, and derived many interesting properties.

The aforementioned papers deal with standard bandit problems, where the reward of an arm is observed only if it is played. Other types of feedback to the decision maker have been considered in the literature. In expert problems [31], the rewards of all arms are observed in each round. Hybrid feedback, between the standard bandit and the expert feedback, has been analyzed in [32]. None of these work addresses the problem considered in this paper, where the reward of an arm (or a noisy version of it) can be observed, before actually playing an arm.

In a recent work [33], the authors study a bandit problem with Bernoulli rewards where in each round, the decision maker proposes an ordered list of the $K$ arms, and plays the first arm with observed reward equal to 1. This problem is similar to that investigated in [34]. The authors devise in this setting an algorithm with regret scaling as $K^2 \log(T)$. However, for this problem, it is easy to show that a constant regret (not scaling with $T$) is achievable. The problem differs from ours, since we assume that the decision maker may observe a single arm only before playing one. In addition, we consider the case of noisy measurements, and we do not restrict our attention to algorithms forced to select an arm, should its measurement returns 1 (this can be sub-optimal in the case of noisy measurements).

Finally, it is worth mentioning contextual bandit problems [25], where arm features are observed as a side information to help the arm selection process. One may think that our problem falls into the class of contextual bandits – features could be the actual arm rewards. However, here, we consider scenarios where the decision maker actively selects parts of context to be observed. Such a scenario in contextual bandits is considered in [35], but without any theoretical analysis.

### 2.3 Models and Preliminaries

We consider the classical stochastic bandit problem, with a set $[K] = \{1, \ldots, K\}$ of arms. The reward generated by arm $k$ in round $t \geq 1$ is denoted by $X_k(t)$. We assume that $(X_k(t))_{t \geq 1}$ is a sequence of i.i.d. random variables with Bernoulli distribution of mean $\theta_k$. Rewards are independent across arms. We denote $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$, and assume w.l.o.g. that $\theta_1 > \theta_2 > \ldots > \theta_K$.

**Measurements.** At the beginning of each round, before playing an arm, the decision maker may decide to measure an arm $k$ at a known cost $c > 0$. When she decides to measure arm $k$ in round $t$, she observes the realization $Z_k(t)$ of a binary random variable, correlated with $X_k(t)$. More precisely, the observation is assumed to correspond to the output of a noisy binary channel with input $X_k(t)$,
and the distribution of $Z_k(t)$ given $X_k(t)$ is: almost surely,

$$P[Z_k(t) = X_k(t)|X_k(t)] = 1 - \varepsilon, \forall k. \quad (2.1)$$

The noise level $\varepsilon$ defines the accuracy of the measurement, and is known to the decision maker. In this paper, we consider two scenarios depending on the measurement accuracy:

(i) Perfect measurements: $\varepsilon = 0$;

(ii) Imperfect measurements: $\varepsilon \in (0, 1/2)$.

**Static policies.** A static policy (also called action in the introduction) $u$ represents the sequence of decisions made in a single round. We distinguish two types of policies. (i) Those directly playing an arm: we denote by $u = (k)$ the policy consisting in playing arm $k$. (ii) Those measuring an arm before actually playing one: such a policy $u$ is described by a triplet $(k, \ell, m)$, where $k$ is the measured arm, and $\ell$ (resp. $m$) is the arm played if the outcome of the measurement is 1 (resp. 0). We denote by $\mathcal{U}$ the set of static policies, and by $\mu(u)$ the average reward of policy $u$. For simplicity, we also use the notation $(k, \ell)$ to denote the policy $(k, k, \ell)$. The objective is to design an algorithm learning the optimal static policy. It is straightforward to check that the optimal policy is either (1) (play the best arm without measuring) or (1, 2) defined as the policy consisting in measuring arm 1, in playing arm 1 if the outcome of the measurement is 1, and in playing 2 if this outcome is 0. One may also easily check that (1, 2) and (2, 1) have the same average reward (i.e., $\mu(1, 2) = \mu(2, 1)$). We have:

$$\left\{ \begin{array}{ll} 
\mu(1) = \theta_1, \\
\mu(1, 2) = -c + (1 - \varepsilon)\theta_1 + ((1 - \varepsilon)(1 - \theta_1) + \varepsilon\theta_1)\theta_2.
\end{array} \right.$$ 

In the second equation, the first term is the measurement cost, the second term is the probability of the measurement and arm 1 both returning 1, while the third is the probability of the measurement returning 0 while arm 2 returns 1. Throughout the paper, we assume that $\mu(1) \neq \mu(1, 2)$. Hence the optimal static policy, denoted by $u^*(\theta)$, is unique (when (1, 2) is optimal, the only other optimal policy is (2, 1)). For simplicity, we denote $\mu^* = \mu(u^*(\theta))$.

**Online learning algorithms and their regret.** An online learning algorithm $\pi$ starts with no knowledge of $\theta$, and aims at gathering data in an active manner to learn $u^*(\theta)$ as quickly as possible. Formally, we represent the observations gathered under $\pi$ up to the beginning of round $t$ by the $\sigma$-algebra $\mathcal{F}^\pi_t$. In round $t$, $\pi$ selects a policy $u^\pi_t$ to be applied in round $t$; $u^\pi_t$ is a $\mathcal{F}^\pi_t$-measurable random variable. The set of all possible online learning algorithms is denoted $\Pi$. The performance of an algorithm $\pi \in \Pi$ is captured through its regret defined, up to round $T$, as

$$R^\pi_\theta(T) = T\mu^* - \sum_{t=1}^T \mathbb{E}[\mu(u^\pi_t)].$$
The regret compares the cumulative reward collected under the learning algorithm \( \pi \) to that one would collect applying the best static policy in each round; it hence quantifies the price to pay to learn \( u^*(\theta) \). We aim at devising an online algorithm with minimal regret.

### 2.4 Regret Lower Bounds

In this section, we derive regret lower bounds satisfied by any online learning algorithms. These bounds constitute an insightful performance benchmark for learning algorithms, but also provide guidelines into their design. We distinguish the perfect and imperfect measurement scenarios.

#### Perfect measurements

To derive lower bounds, we use classical change-of-measure arguments (refer to [24], and to [36] for a general framework). These bounds will concern so-called uniformly good algorithms: \( \pi \in \Pi \) is uniformly good if its regret satisfies for any \( \theta \), \( R_\theta^\pi(T) = o(T^\alpha) \) for all \( \alpha > 0 \). Observe that such algorithms exist, since UCB applied to a bandit problem with set of 'arms' \( \mathcal{U} \) would exhibit a regret scaling logarithmically with \( T \). In the following, we denote by \( I(\lambda, \lambda') \) the KL divergence between two Bernoulli distributions with respective means \( \lambda \) and \( \lambda' \). More generally, we denote by \( KL(\nu_1 || \nu_2) \) the KL-divergence between two distributions \( \nu_1 \) and \( \nu_2 \) (when it is well-defined).

**Lemma 1.** The regret of any uniformly good algorithm \( \pi \in \Pi \) satisfies: for all \( \theta \),

\[
\lim_{T \to \infty} \inf \frac{R_\theta^\pi(T)}{\log(T)} \geq C(\theta),
\]

where \( C(\theta) \) is the value of the following optimization problem:

\[
\min_{\eta_u \geq 0, \forall u \in \mathcal{U}} \sum_{u \in \mathcal{U}} \eta_u (\mu^* - \mu(u))
\]

subject to

\[
\sum_{u \in \mathcal{U}} \eta_u (\mathbb{1}(u = (k)) + \mathbb{1}(u = (k, 1))) \geq \frac{1}{I(\theta_k, \bar{\theta})}
\]

\[
\forall k \notin u^*(\theta),
\]

and

\[
\bar{\theta} = \begin{cases} 
\theta_2, & \text{if } u^*(\theta) = (1, 2) \\
\min(\theta_1, \frac{c}{1 - \theta_1}), & \text{if } u^*(\theta) = 1.
\end{cases}
\]
2.4. REGRET LOWER BOUNDS

Theorem 1. The solution of the optimization problem in Lemma 1 is

Case 1: \( \eta^*_u = \sum_{k \notin u^*(\theta)} \mathbb{I}(u = (k,1))/I(\theta_k, \theta) \), \( \forall u \) when \( u^*(\theta) = (1,2) \) (i.e., when \( c < \theta_2(1 - \theta_1) \)). Hence, in this case

\[
C(\theta) = \sum_{k=3}^{K} \frac{(1 - \theta_1)(\theta_2 - \theta_k)}{I(\theta_k, \theta_2)}.
\]

Case 2: for any \( k \notin u^*(\theta) \), and any \( u \) such that \( k \in u \),

\[
\eta^*_u = \frac{1}{I(\theta_k, \theta)} \times \left\{ \begin{array}{ll}
\mathbb{I}(u = (k)) & \text{if } c < \theta_1(1 - \theta_k), \\
\mathbb{I}(u = (k, 1)) & \text{otherwise.}
\end{array} \right.
\]

when \( u^*(\theta) = (1) \) (i.e., when \( c > \theta_2(1 - \theta_1) \)). Hence, in this case, \( C(\theta) = \sum_{k=2}^{K} H_k(\theta) \) with

\[
H_k(\theta) = \frac{1}{I(\theta_k, \theta)} \times \left\{ \begin{array}{ll}
c - (1 - \theta_1)\theta_k & \text{if } c < \theta_1(1 - \theta_k) \\
\theta_1 - \theta_k & \text{otherwise.}
\end{array} \right.
\]

In the above theorem, the solution \( \eta^* \) to the optimization problem leading to \( C(\theta) \) may be interpreted as follows: \( \eta^*_u \log(T) \) is the expected number of rounds the policy \( u \) should be selected by a learning algorithm minimizing regret. Such an optimal algorithm would explore only very specific policies. Indeed, for any \( k \notin u^*(\theta) \), one and only one of the policies \( u = (k) \) or \( u = (k,1) \) should be explored a number of rounds of the order \( \log(T) \); all other policies have to be explored \( o(\log(T)) \) times.

Also observe that \( \bar{\theta} \) may be interpreted as the value to which the parameter \( \theta_k \) should be changed to make a policy using arm \( k \) (i.e. \( (k) \) or \( (k,1) \)) optimal. Now let \( \nu_\theta(u) \) denote the distribution of the observation made in a given round under the policy \( u \). As this will be come clear in the proof of the theorem, the quantity \( I(\theta_k, \bar{\theta}) \) is actually equal to \( KL(\nu_\theta(u)||\nu_\theta(u)) \) for \( u = (k) \) or \( u = (k, 1) \), where \( \theta' \) is such that \( \theta'_j = \theta_j \), for all \( j \neq k \), and \( \theta'_k = \bar{\theta} \). It can be interpreted as the amount of information brought by policy \( u \) in a single round to decide whether \( k \) is part of the optimal policy. It can be verified that the policy \( u \) including \( k \) that should be explored is the one minimizing the ratio of its regret \( \mu^* - \mu(u) \) to the amount of information brought to decide whether \( k \) is part of the optimal policy. This principle is general, and will also hold in the case of imperfect measurements.

Proof of Lemma 1 We use change-of-measure arguments. Let \( \pi \) be a uniformly good algorithm. Denote by \( \Lambda(\theta) \) the set of confusing problem parameters, i.e., those leading to a different optimal policy, and that cannot be distinguished from the true parameters if the optimal policy is always played. In other words:

\[
\Lambda(\theta) = \{ \lambda : KL(\nu_\theta(u^*(\theta))||\nu_\lambda(u^*(\theta))) = 0, u^*(\theta) \neq u^*(\lambda) \}.
\]
Note that if \( u^*(\theta) = (1) \), then
\[
KL(\nu_\theta(u^*(\theta))||\nu_\lambda(u^*(\theta))) = 0 \iff \lambda_1 = \theta_1,
\]
and if \( u^*(\theta) = (1, 2) \), then
\[
KL(\nu_\theta(u^*(\theta))||\nu_\lambda(u^*(\theta))) = 0 \iff (\lambda_1 = \theta_1, \lambda_2 = \theta_2).
\]
If \( \mathbb{E}^\pi[N_u(T)] \) is the expected number of rounds where \( \pi \) applies policy \( u \) up to time \( T \), we can show as in [36] that: for all \( \lambda \in \Lambda(\theta) \),
\[
\sum_u \mathbb{E}^\pi[N_u(T)]KL(\nu_\theta(u)||\nu_\lambda(u)) \geq \log(T)(1 + o(1)). \tag{2.3}
\]

Since \( R^\theta_u(T) = \sum_u \mathbb{E}^\pi[N_u(T)](\mu^* - \mu(u)) \), this implies that an asymptotic lower bound for the regret is \( C(\theta) \log(T) \), where \( C(\theta) \) is the value of the solution of the following optimization problem.
\[
\min_{\eta_u \geq 0} \sum_{u \in U} \eta_u(\mu^* - \mu(u)) \tag{2.4}
\]
\[
\text{s.t.} \sum_{u \in U} \eta_u KL(\nu_\theta(u)||\nu_\lambda(u)) \geq 1, \forall \lambda \in \Lambda(\theta). \tag{2.5}
\]

**Pruning constraints:** We argue that we can restrict the set of constraints in the above problem, by restricting the attention to \( \lambda \in \Lambda(\theta) \) such that only one coordinate of \( \lambda \) differs from those of \( \theta \). We distinguish two cases. First, if \( (1, 2) \) is the optimal policy under \( \theta \), then we have \( \lambda_1 = \theta_1 \) and \( \lambda_2 = \theta_2 \). If under \( \lambda \), \( (k) \) is optimal (for \( k \geq 3 \)), then it is easy to see that \( \lambda_k \) should be set just above \( \theta_2 \), and we do not need to change any other component of \( \theta \). Similarly, if under \( \lambda \), \( (k, 1) \) is optimal, then changing only \( \lambda_k \) is required. Now assume that under \( \lambda \), \( (k, \ell) \) is optimal for \( k, \ell \notin \{1, 2\} \). We must have: \( \lambda_k + \lambda_\ell - \lambda_k \lambda_\ell > \theta_1 + \theta_2 - \theta_1 \theta_2 \), from which we deduce that either \( \lambda_k \) or \( \lambda_\ell \) is greater than \( \theta_2 \). Hence, the constraint generated by this \( \lambda \) is not active. We can do the same reasoning to show that if \( (1) \) is optimal under \( \theta \), then the active constraints are those corresponding to \( \lambda \)'s that differ from \( \theta \) by one coordinate only. In this case, however, it suffices that \( \lambda_k > \frac{c}{1 - \theta_1} \), as this will imply \( \lambda_1 + (1 - \lambda_1)\lambda_k - c > \theta_1 \).

**Proof of Theorem 7** By studying the average rewards \( \mu(u) \) and the KL-divergence \( KL(\nu_\theta(u)||\nu_\lambda(u)) \) of the various policies, we can show that the solution \( \eta^*_u \) of (2.4)-(2.5) is such that for most policies \( u \), \( \eta^*_u = 0 \). We do so by showing that for such \( u \) and for any feasible solution \( \eta \), \( \eta_u > 0 \iff \eta \neq \eta^* \). Assume first that \( u^*(\theta) = (1, 2) \). Let \( k \geq 3 \). Then the set of constraints for \( \lambda \in \Lambda(\theta) \) such that \( \lambda_\ell = \theta_\ell \) for all \( \ell \neq k \) reduces to the single constraint
\[
\sum_{u : k \in u} \eta_u KL(\nu_\theta(u)||\nu_\lambda(u)) \geq 1,
\]
where \( \lambda_k = \theta_2 \). The KL divergences involved in this constraint are: for \( \ell, \ell_1, \ell_2 \neq k \),

\[
KL(\nu_\theta(u)||\nu_\lambda(u)) = \begin{cases} 
I(\theta_k, \theta_2) & \text{case I}, \\
(1 - \theta_\ell)I(\theta_k, \theta_2) & \text{case II}, \\
\theta_\ell, I(\theta_k, \theta_2) & \text{case III},
\end{cases}
\]

where case I holds for \( u = (k), (k, \ell), (k, \ell_1, \ell_2) \), case II for \( u = (\ell, k) \), and case III for \( u = (\ell_1, k, \ell_2) \). Consider \( u = (k, \ell) \) with \( \ell \neq k \) and \( \ell > 1 \), take any feasible solution \( \eta \) such that \( \eta_u > 0 \), and consider another feasible solution \( \eta' \), identical to \( \eta \) except \( \eta_u = 0 \), \( \eta'_{(k, 1)} = \eta_{(k, 1)} + \eta_u \) and \( \eta'_{(\ell, 1)} = \eta_{(\ell, 1)} + (1 - \theta_k)\eta_u \). Then the difference in cost function between \( \eta \) and \( \eta' \) is

\[
\eta_u((\mu^* - \mu((k, \ell))) - (\mu^* - \mu(k, 1)) - (1 - \theta_k)(\mu^* - \mu(\ell, 1)))
\]

\[
= \eta_u(1 - \theta_k)(((1 - \theta_\ell) - (\theta_2 - \theta_\ell)) > 0.
\]

Therefore \( \eta^* \neq \eta \) and hence \( \eta_{(k, \ell)}^* = 0 \). Similar arguments lead to \( \eta_{(\ell, k)}^* = \eta_{(k, 1), \ell_2}^* = \eta_{(\ell_1, k), \ell_2}^* = 0 \) for any \( \ell, \ell_1, \ell_2 \neq k \), and, comparing \( (k) \) to \( (1, k) \),

\[
\eta_{(k)}^* = 0.
\]

By process of elimination, we deduce the results of Case 1 in Theorem 1.

Assuming now that \( u^*(\theta) = (1) \), we prove the results of Case 2 in Theorem 1 using the same arguments.

**Imperfect measurements**

The following theorem provides regret lower bounds in the case of imperfect measurements. For simplicity, we define \( p_0(\theta_k) := P[Z_k(t) = 0] = \varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k) \).

**Theorem 2.** The regret of any uniformly good algorithm \( \pi \in \Pi \) satisfies: for all \( \theta \),

\[
\lim \inf_{T \to \infty} \frac{R^\pi_\theta(T)}{\log(T)} \geq C_\varepsilon(\theta),
\]

where \( C_\varepsilon(\theta) \) is the value of the following optimization problem:

\[
\min_{\eta_u \geq 0} \sum_{u \in U} \eta_u(\mu^* - \mu(u)) \quad \text{s.t.} \quad \sum_{u \in U_1(k)} \eta_u KL(\nu_\theta(u)||\nu_{(\theta_{(-k)}, \bar{\theta})(\theta)}(u)) \geq 1, \forall k \neq u^*(\theta),
\]

where \( U_1(k) = \{(k), (k, 1), (k, 1, k), (k, 1, 1)\} \), and where \( (\theta_{(-k)}, \bar{\theta}) = \theta' \) corresponds to arm rewards such that \( \theta'_j = \theta_j \), for \( j \neq k \), and \( \theta'_k = \bar{\theta} \). The parameter \( \bar{\theta} \) depends on \( \theta \) as follows. When \( u^*(\theta) = (1, 2) \), we have \( \bar{\theta} = \theta_2 \); when \( u^*(\theta) = (1) \),

\[
\bar{\theta} = \min(\theta_1, \frac{c + \varepsilon \theta_1}{p_0(\theta_1)}).
\]
The solution $\eta^*$ of the above optimization problem is:

$$
\eta^*_u = \sum_{k \notin u^*(\theta)} \frac{1(u = u^*_k)}{KL(\nu_{\theta}(u)||\nu_{(\theta(-k),\bar{\theta})}(u))}
$$  \hspace{1cm} (2.6)

where $u^*_k = \arg\min_{u \in U_1(k)} h_k(u)$ with

$$
h_k(u) = \frac{\mu^* - \mu(u)}{KL(\nu_{\theta}(u)||\nu_{(\theta(-k),\bar{\theta})}(u))}.
$$

Thus, $C_\varepsilon(\theta) = \sum_{k \notin u^*(\theta)} h_k(u^*_k)$.

Theorem 2 and its interpretation are very similar to Theorem 1 and in fact Theorem 2 reduces to Theorem 1 when $\varepsilon \to 0$, with $H_k(\theta) = h_k(u^*_k)$. In particular, we still have that policies including more than one suboptimal arm will not be considered for exploration. To decide whether $k$ belongs to the optimal policy, an optimal algorithm should explore a single policy containing arm $k$ and possibly arm 1. However, in the case of imperfect measurements, this policy can be any of the 4 policies in $U_1(k)$, depending on the parameter $\theta$. Again this policy is the one minimizing the ratio $h_k(u)$ of its regret to the amount of information it brings. We omit the full proof due to space constraints, but it is similar to that of Theorem 1 and can be found in our technical report [37].

2.5 Algorithms

In this section, we exploit our regret lower bounds to devise algorithms, in both scenarios, with perfect and imperfect measurements. We also provide an analysis of the regret of the proposed algorithms. Our algorithms can be seen as an extension of the algorithms in [34]. Their novelty lies in exploiting the structure of predictive bandits through the results of Theorems 1 and 2.

Perfect measurements

We present Single Predictive Arm Measurements (SPAM), an algorithm whose regret matches the lower bound derived in Theorem 1. SPAM maintains a leading arm $j_1(t)$ defined as the best empirical arm up to round $t$, $j_1(t) \in \arg\max_k \hat{\theta}_k(t)$ (ties are broken arbitrarily), where $\hat{\theta}_k(t)$ denotes the empirical reward of arm $k$ averaged over the $(t - 1)$ first rounds. It also maintains $j_2(t)$, the second best empirical arm, as well as the best empirical policy $\mathcal{L}(t)$ (either $(j_1(t))$ or $(j_1(t), j_2(t))$). SPAM uses KL-UCB indices: for arm $k$, this index is defined as:

$$
b_k(t) := \max\{q : n_k(t)I(\hat{\theta}_k, q) \leq f(t)\},
$$  \hspace{1cm} (2.7)
where \( f(t) = \log(t) + 4 \log \log(t) \) and \( n_k(t) \) is the number of times arm \( k \) has been observed up to time \( t \). In each round, to decide whether SPAM should explore apparently sub-optimal policies, these indices are compared to an estimated threshold \( \hat{\theta}(t) \), equal to \( \hat{\theta}_{j_2}(t) \) if \( \hat{\theta}_{j_2}(t) \geq \frac{c}{1-\theta_{j_1}(t)} \) and \( \min(\hat{\theta}_{j_1}(t), \frac{c}{1-\theta_{j_1}(t)}) \) otherwise. SPAM only explores policies containing arms in the following set of uncertain arms:

\[
\mathcal{B}(t) := \{ k : b_k(t) \geq \hat{\theta} \}. \tag{2.8}
\]

SPAM exploits, i.e., select the leading policy \( \mathcal{L}(t) \), very regularly (with probability at least 1/2 in each round), so that the arms in the leading policy are very well estimated. SPAM explores apparently sub-optimal policies only if the set \( \mathcal{B}(t) \) is not empty. More precisely, it explores either \((k)\) or \((k, j_1(t))\) for \( k \in \mathcal{B}(t) \). All the design choices made in SPAM are aligned to the optimal exploration process suggested in our regret lower bound. The pseudo-code of SPAM is presented in Algorithm 1.

### Algorithm 1 SPAM

```plaintext
1: Initialize \( \hat{\theta}_k(1) = 1 \) and \( b_k(1) = 1 \) for all arms \( k \),
2: \( \mathcal{B}(1) = \emptyset \), and \( \mathcal{L}(1) \) arbitrarily.
3: for \( t = 1, 2, \ldots \) do
4:    if \( \mathcal{B}(t) = \emptyset \) then exploit: \( u \leftarrow \mathcal{L}(t) \),
5:        else
6:           w.p. 1/2, exploit: \( u \leftarrow \mathcal{L}(t) \),
7:           w.p. 1/2, explore: choose \( k \) uniformly at random
8:          from \( \mathcal{B}(t) \), then:
9:              \( u \leftarrow (k, j_1(t)) \) if \( (1 - \hat{\theta}_{j_1}(t))(\hat{\theta}_k(t) > c), \)
10:             \( u \leftarrow (k) \) otherwise.
11:        end if
12:    Play policy \( u \) and observe its outcomes.
13: Compute \( \hat{\theta}_k(t+1) \) and \( b_k(t+1) \) for all arms \( k \),
14: Compute \( \mathcal{B}(t+1), \mathcal{L}(t+1) \).
15: end for
```

Before we provide, in the theorem below, a finite-time analysis of the regret of SPAM, we introduce the following notations. For any \( \theta \), let \( \delta_0 \) be such that (i) \( \delta_0 \leq \min_{i<K}(\frac{1}{2}(\theta_i - \theta_{i+1})) \), (ii) if \( u^*(\theta) = (1, 2) \), \( \frac{c}{1-\theta_1-\delta_0} \leq \theta_2 - \delta_0 \), and (iii) if \( u^* = (1) \), \( \frac{c}{1-\theta_1+\delta_0} \geq \theta_2 + \delta_0 \). It can be easily checked that such a \( \delta_0 \) indeed exists. Let \( \beta = (1 - \theta_1)^{-1} \), and define for \( \delta > 0 \),

\[
g(\theta_1, \theta_2, \delta) := \begin{cases} 
\theta_1 - \delta, & \bar{\theta} = \theta_1 \\
\theta_2 - \delta, & \bar{\theta} = \theta_2 \\
\frac{c}{1-\theta_1+\delta}, & \bar{\theta} = \frac{c}{1-\theta_1}.
\end{cases}
\]
Finally, we introduce the functions $H_k$ so that the constant $C(\theta)$ involved in regret lower bound derived in Theorem 1 can be written as $C(\theta) = \sum_{k \not\in u^*} H_k(\theta)$ in all cases. Hence, if $u^* = (1, 2)$, we have $H_k(\theta) := (1 - \theta_1)(\theta_2 = \theta_k)/I(\theta_k, \theta_2)$ and if $u^* = (1)$, $H_k(\theta)$ is defined as in Theorem 1.

**Theorem 3.** There is a constant $d > 0$ such that for any $\theta$, any $\delta < \delta_0$ and any $\epsilon \in (0, 1/2)$, the regret of SPAM satisfies: for all $T \geq 1$,

$$R^{\text{SPAM}}_\theta(T) \leq \sum_{k \not\in u^*(\theta)} \frac{H_k(\theta)I(\theta_k, \bar{\theta})}{(1 - \epsilon)I(\theta_k, g(\theta_1, \theta_2, \delta))} f(T) + dK(K + \beta^2) + \epsilon^{-2} + \delta^{-2}(\beta + 1). \quad (2.9)$$

An immediate consequence of the above theorem, whose proof can be found in our technical report [37], is that SPAM is asymptotically optimal. Indeed, by letting first $T$ tend to $\infty$, and then $\epsilon, \delta$ to 0, we obtain:

$$\limsup_{T \to \infty} \frac{R^{\text{SPAM}}_\theta(T)}{\log(T)} \leq C(\theta).$$

**Imperfect measurements**

The design of our algorithm for the case of noisy measurements follows the same principles as that of SPAM, but is slightly complicated because: (i) According to our lower bounds, to determine whether arm $k$ belongs to the optimal policy, the 4 policies of $U_i(k)$ could be used in the exploration process. (ii) Due to the noisy measurements, the estimation of $\theta_k$ is slightly involved. Next, we propose Noisy Single Predictive Arm Measurements (NoSPAM), an extension of SPAM to the case of noisy measurements. The regret analysis of NoSPAM is complicated by the aforementioned facts. We conjecture that NoSPAM is asymptotically optimal, just as SPAM, but omit the analysis here. The main difference between SPAM and NoSPAM lies in the estimation of the parameters $\theta$, which we explain next.

**Estimating average arm rewards.** To derive $\hat{\theta}_k(t)$, an estimator of $\theta_k$, we use the following quantities. Let $n_{1,k}(t)$ be the number of rounds $s$ up to round $t$ where it has been observed that $X_k(s) = 1$; let $n_{2,k}(t)$ be the number of rounds $s$ where it has been $X_k(s) = 0$; let $n_{3,k}(t)$ be the number of rounds $s$ where $Z_k(s) = 1$ has been observed but $X_k(s)$ has not been observed, and finally let $n_{4,k}(t)$ be the number of rounds $s$ where $Z_k(s) = 0$ has been observed but $X_k(s)$ has not been observed. Define $n_k(t) = \sum_{i=1}^4 n_{i,k}(t)$, the number of rounds $s$ where either $Z_k(s)$ or $X_k(s)$ have been observed. It can be readily shown that the maximum-likelihood estimator $\hat{\theta}_k$ of $\theta_k$ is the root $X \in [0, 1]$ of a polynomial, and can therefore be efficiently found numerically.
Algorithm 2 NoSPAM

1: Initialize $\hat{\theta}_k(1) = 1$ and $b_k(1) = 1$ for all arms $k$,
2: $B(1) = \emptyset$, and $\mathcal{L}(1)$ arbitrarily.
3: for $t = 1, 2, \ldots$ do
4: if $B(t) = \emptyset$ then exploit: $u \leftarrow \mathcal{L}(t)$,
5: else
6: w.p. 1/2, exploit: $u \leftarrow \mathcal{L}(t)$,
7: w.p. 1/2, explore: choose $k$ uniformly at random
8: from $\mathcal{B}(t)$, then:
9: for $u \in \Gamma := \{(k, j_1(t)), (k), (k, j_1(t), k),
10: (k, j_1(t), j_1(t))\}$, calculate
11: $h_k(u) = \frac{\mu(L(t)) - \mu(u)}{KL(\nu_{\hat{\theta}(t)}(u)||\nu_{(\hat{\theta}^{(k)}, \hat{\nu}(t))}(u))}$
12: $u \leftarrow \arg \min_{u \in \Gamma} h_k(u)$
13: end if
14: Play policy $u$ and observe its outcomes.
15: Compute $\hat{\theta}_k(t + 1)$ and $b_k(t + 1)$ for all arms $k$,
16: Compute $B(t + 1), \mathcal{L}(t + 1)$.
17: end for

Now, defining $n_k^{\text{play}}(t)$ (resp. $n_u(t)$) as the number of rounds $s$ up to $t$ where $X_k(s)$ is observed but not $Z_k(s)$ (resp. $u$ is selected), we can define the KL-UCB index of arm $k$ as:

$$b_k(t) = \max\{q : \sum_{u \in \mathcal{U}_m(k)} n_u(t) KL(\nu_{\hat{\theta}(t)}(u)||\nu_{(\hat{\theta}^{(k)}, \hat{\nu}(t))}(u))$$

$$+ n_k^{\text{play}}(t) I(\hat{\theta}_k(t), q) \leq f(t),$$

where $f(t) = \log(t) + 4 \log(\log(t))$ and $\mathcal{U}_m(k)$ denotes the set of all policies where $k$ is measured. Recall that $\nu_{(\hat{\theta}^{(k)}, \hat{\nu}(t))}(u)$ is defined in Theorem $2$, $\mathcal{B}(t), j_1(t), j_2(t)$ and $\mathcal{L}(t)$ are defined as for SPAM, with $\hat{\theta}(t) = \hat{\theta}_{j_2(t)}(t)$ if $\hat{\theta}_{j_2(t)}(t) \geq \frac{c + \epsilon \hat{\theta}_{j_1(t)}(t)}{p_0(\hat{\theta}_{j_1(t)}(t))}$ and $\hat{\theta}(t) = \min(\hat{\theta}_{j_1(t)}, \frac{c + \epsilon \hat{\theta}_{j_1(t)}(t)}{p_0(\hat{\theta}_{j_1(t)}(t))})$ otherwise. The pseudo-code of NoSPAM is presented in Algorithm $2$. As discussed with Theorem $2$, $(k, j_1(t), k)$ or $(k, j_1(t), j_1(t))$ will never minimize $h_k(u)$ when $\epsilon = 0$, and so NoSPAM reduces to SPAM in this case. Further, when $\epsilon = \frac{1}{2}$, $h_k(u)$ will always be minimized by $u = (k)$, and so NoSPAM reduces to a more aggressive variant of KL-UCB in this case.
2.6 Numerical Experiments

In this section, we illustrate the performance of SPAM and NoSPAM. We compare their performance to that of KL-UCB, as applied to the set of static policies rather than to the arms. (Recall that a policy may include a measurement before an arm is played, so this allows the KL-UCB agent to utilize measurement information.) KL-UCB is known to be asymptotically optimal when the various arms have uncorrelated rewards. Here, however, the rewards of policies using the same arm are correlated, and this is precisely this structure that SPAM and NoSPAM optimally exploit.

**Implementation of KL-UCB.** KL-UCB selects the policy with the highest KL-UCB index. We know a priori that the optimal static policy is of the form $(k)$ or $(k, \ell)$, and so naturally, we restrict KL-UCB to these policies. To exploit all the observations made up to round $t$, we define the KL-UCB index of policy
2.7. Conclusion

In existing bandit and contextual bandit problems, the decision maker cannot
decide to observe the rewards of specific arms or their contexts before actually
playing an arm. Such an observation in a given round would help the decision

maker to predict the rewards in that round, but would typically come with a cost. In this paper, we move towards such predictive bandits and investigate problems where the agent can measure the reward of at most one arm before making playing an arm. These measurements are either perfect or have a known probability of being incorrect. We derive a regret lower bound for these problems, and devise algorithms in the endeavor of matching these bounds. This paper proposes the first analytical results on predictive bandits, and naturally suggests interesting research directions. We can for instance extend the analysis to problems where the agent may measure multiple arms, or where the agent needs to learn the uncertainty of arms. More generally, it would be also interesting to investigate contextual bandit problems where the agent must choose which parts of the context to observe. Further, we look forward to the application of predictive bandits to real world applications.
Chapter 3

Paper II: Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach

Foreword

After the initial analysis of Paper I, we were eager to apply the ideas within it to slightly more involved networking problems. Several approaches were considered, most prominently one where we would apply adaptive measurement schedules to hypothesis testing in order to minimize measurement costs - this would lay the ground-work for Paper III. Eventually, we settled on admission control, where the focus was put on one-shot learning in order to more clearly emphasize the effects of adaptive measurement schedules. Instead of attempting to device optimal admission control algorithms, which we knew would be difficult, we applied known heuristic algorithms and showed the effect these algorithms have on measurement cost in environments where resources are unknown. This also had a clearer connection to the ideas of Paper I.

Since the paper was published, the ideas of the paper have been attempted using Reinforcement Learning as part of a master thesis. However, due to the difference in approaches, no conclusive insights were gained from this.

As with Paper I, some proofs had to be omitted in the final publication to GLOBECOM 2022. These proofs are found in Appendix A.

Paper II is reprinted, with permission, from [38]. The text has been reformatted, but no text has been changed. ©IEEE 2022.
Measurement-based Admission Control in Sliced Networks: A Best Arm Identification Approach
Simon Lindstål, Alexandre Proutiere, Andreas Johnsson

Abstract

In sliced networks, the shared tenancy of slices requires adaptive admission control of data flows, based on measurements of network resources. In this paper, we investigate the design of measurement-based admission control schemes, deciding whether a new data flow can be admitted and in this case, on which slice. The objective is to devise a joint measurement and decision strategy that returns a correct decision (e.g., the least loaded slice) with a certain level of confidence while minimizing the measurement cost (the number of measurements made before committing to the decision). We study the design of such strategies for several natural admission criteria specifying what a correct decision is. For each of these criteria, using tools from best arm identification in bandits, we first derive an explicit information-theoretical lower bound on the cost of any algorithm returning the correct decision with fixed confidence. We then devise a joint measurement and decision strategy achieving this theoretical limit. We compare empirically the measurement costs of these strategies, and compare them both to the lower bounds as well as a naive measurement scheme. We find that our algorithm significantly outperforms the naive scheme (by a factor $2 - 8$).

3.1 Introduction

In next generation telecom networks, the network resources will be divided and allocated to multiple slices shared between several slice tenants. With limited to no knowledge of the behavior of other tenants, a slice tenant must, in order to uphold certain service guarantees, decide to accept or reject incoming data flows, while adapting to rapidly changing network occupancy levels. This is further complicated by an unclear dependency of the slice occupancy on the resources of individual slice components. The admission control agent must therefore measure the network resources and current utilization before an admission decision can be made, reintroducing a need for measurement-based admission control (MBAC), a popular method in the context of call admission control which has recently fallen out of favor. MBAC schemes have the advantage to adapt to uncertainties arising due to the difficulty of characterizing traffic sources or to that of estimating the available resources (in wireless networks, these evolve depending on e.g. user mobility, fading, interference). However, MBAC comes with an inherent cost since a non-negligible fraction of the resources is used for the measurements. This
cost can become substantial as the admission criteria grows in complexity \cite{43}, especially in system with inherently scarce resources such as wireless systems \cite{44}.

In this paper, we investigate the design of MBAC strategies in multi-slice networks, where the controller has to decide whether a new data flow can be admitted and if so, on which slice. The controller has no knowledge about the slice loads, but may gather this knowledge conducting noisy and costly measurements. To this aim, it can sequentially measure the traffic handled (over a fixed duration – a time slot) by a selected slice, and stop whenever it believes it has gathered enough information to come up with a \textit{correct} decision with some level of certainty. A correct decision should be to reject the flow if all slices are already fully loaded, or to select one of the slices that has enough available resources if any, given assumptions on the new network flow. The objective is to devise a joint measurement and decision strategy that returns a correct decision with a certain level of confidence while minimizing the measurement cost (the number of measurements made before committing to the decision). We study the design of such strategies for several natural admission criteria specifying what a correct decision is. These criteria can consist in selecting (i) \textit{any} of the slices with available resources, (ii) the most loaded slice with available resources (we refer to this slice as the \textit{packing} slice), or (iii) the \textit{least loaded} slice with available resources.

We address the design of joint measurement and decision strategies using the formalism of pure exploration in stochastic Multi-Armed Bandits (MAB). Online exploration algorithms for MAB specify an adaptive sequence of arms (for us, slices) to observe (here, the traffic handled by the selected slice in a slot), a stopping rule indicating when to output a decision, and a decision rule. For each of the aforementioned admission criteria, we first derive an explicit information-theoretical lower bound on the cost of any algorithm returning the correct decision with fixed confidence. We then devise a joint measurement and decision strategy achieving this theoretical limit. We compare empirically the costs of these strategies, and compare these cost both to the lower bound and to a naive sampling strategy. These results allow us to analyze the trade-off between measurement cost and complexity of the proposed admission criteria. Our main contribution, as such, is the adaptation of previous best arm identification methods into a format well suited for slice admission control and the measurement-efficient decision algorithms this allows.

### 3.2 Related work

\textbf{Stochastic bandit} problems have received plenty of attention since they were introduced by Thompson in the 30’s and formalized by Robbins in 1952. While bandit problems were initially motivated by clinical trials, they have recently found important applications in the design of protocols and algorithms in communication networks (mostly in cognitive radio systems, see \cite{45,46} and references therein, or rate adaptation in wireless systems \cite{47}). Most often in bandits, the
focus has been on the design of algorithms with low regret, defined as the cumulative loss of a learning algorithm compared to an optimal policy over time [24]. The problem of identifying the best arm using a minimal number of samples is more recent, see [48, 49] for early work. Algorithms to find the best arm with minimal sample complexity have been developed in [50]. Since then, researchers have tried to extend these algorithms to more general pure exploration problems [51], such as [52] where one searches for the arm with average reward the closest to a given threshold with as few total samples as possible. In this paper, we investigate three novel pure exploration problems, each corresponding to a different admission criterion, and we use the framework developed in [50] to derive sample complexity lower bounds and to devise optimal algorithms based on these lower bounds. Compared to [50], however, our methods can deal with more general problems than simply identifying the arm with greatest/lowest mean. While [51] constructed general lower bounds for such problems, they are often implicit and non-trivial to compute. Furthermore, their algorithm Sticky Track-and-Stop cannot typically be implemented without both an explicit form of these bounds and the assumption of Gaussian random variables. By contrast, we provide explicit bounds for our admission criteria as well as an algorithm applicable for a wide class of measurement distributions.

Admission control methods in the context of network slicing are summarized in [53]. These methods vary in slice elasticity, inter vs intra-slice admission control, single vs multi-tenant systems, and use both heuristic and optimal methods. None of these methods explicitly take measurement overhead into account. As far as we are aware, this paper proposes the first approach to actually optimize the measurement strategy in admission control. It is worth noting that our admission control problems may seem similar to the problem of dynamic channel assignment in wireless networks, see e.g., [54–56]. However, most existing work in this field concerns the design of Medium Access Control (MAC) protocols (a faster time scale than that of flow arrivals), and most often, channels may take two states only, busy or free.

### 3.3 Models: Dynamics, Admission Control, and Measurement Costs

This section presents our network model, and states our admission control problem. The network uses \( K \) slices of equal capacity to handle traffic flows or services generated by end users to provide a certain level of Quality of Service. That is, the resources are shared by many users. When a new flow is created, the slice tenant managing these users, or in other words the controller, has no knowledge about the current traffic conditions on the various slices, but wishes to select a slice so that the performance guarantees of existing flows in the slice remains as high as possible. In this case, this translates to ensuring the loads of all slices
remains below some *threshold*. To determine which slice should handle the flow or whether the flow should be rejected, the controller has to measure the traffic intensity on slices. This measurement procedure induces a cost, such as power or bandwidth consumption, that the controller wishes to minimize. We describe this cost minimization problem in detail below, and an outline of the system is found in Figure 3.1. In this figure, the slices are visualized as a chain of virtual network functions (VNFs) depicted as blue boxes, connected to radio over network links with a controller monitoring the network. The shaded areas correspond to the utilization level of the VNFs as consumed by a set of flows. In this case, the correct decision is for the controller to admit into Slice 1 as it is the only slice with resources across the whole slice.

**Packet-level dynamics and admission criteria**

Flows are assumed to generate packets according to a stationary process. When for a given slice, the set of accepted flows is fixed, we assume that the aggregate packet arrival process has statistics described, for simplicity, by a single parameter. Time is slotted, and this parameter is defined as the average number of packets arriving in one slot. In this paper, the processes is either Bernoulli (if the slot duration is very small) or Poisson (for the usual Poisson model at packet level in data networks). For slice $k \in [K]$, we denote by $\mu_k$ as the mean packet arrival rate per slot, fixed during a decision setting, and define $\mu = (\mu_1, \ldots, \mu_K)$. $\mathbb{P}_\mu$ (resp. $\mathbb{E}_\mu$) denotes the probability distribution (resp. the expectation) of observations when the packet arrival rates are parameterized by $\mu$.

In this paper, we assume that a user generates a new flow with known packet arrival rate $r$. Further, we assume that the current traffic in the network is described by $\mu$. We consider scenarios where accepting the flow should be ideally decided based on $r$ and $\mu$. This happens for example when we wish to guarantee that the average packet delay of accepted flows remains smaller than a given threshold. For Poisson packet arrivals, the threshold $\gamma$ is obtained simply by plugging $r$, $\mu$, the slice capacity, and packet size statistics in the M/G/1
Pollaczek–Khinchine formula. As a result, the flow should be ideally accepted in slice \( k \) only if \( \mu_k < \gamma \), in which case, we say that slice \( k \) is available. The flow should be rejected if none of the slices are available. In this paper, when a new flow is created, \( \mu \) is unknown and has to be learned through repeated measurements.

**Best slice identification problems**

The controller applies a joint measurement and decision strategy to decide whether a newly generated flow can be accepted and if so, on which slice. In each slot \( t \geq 1 \), we may measure for a selected slice, say \( k \), the number of packets \( X_k(t) \) handled by the slice in that slot. For example under the Poisson traffic assumption, the r.v. \( X_k(t) \) are i.i.d. with Poisson distribution of unknown mean \( \mu_k \). Now a joint measurement and decision strategy consists of three components:

(i) **A sampling strategy.** It specifies, in each slot, the slice to measure. Measurements are taken once per slice and consider end-to-end load for simplicity, rather than load in individual VNFs (as in Figure 3.1). For \( t \geq 1 \), denote by \( k_t \) and by \( X_{k_t}(t) \) the slice probe in slot \( t \) and the corresponding number of packets observed. Then \( k_t \) depends on past observations, i.e., \( k_t \) is \( F_{t-1} \)-measurable where \( F_t \) is the \( \sigma \)-algebra generated by \((k_1, X_{k_1}(1), \ldots, k_t, X_{k_t}(t))\).

(ii) **A stopping rule.** It controls the end of the data acquisition phase and is defined as a stopping time \( \tau \) with respect to the filtration \((F_t)_{t \geq 1}\) such that \( \mathbb{P}_\mu(\tau < \infty) = 1 \).

(iii) **A decision rule.** At the end of slot \( \tau \), the algorithm returns a decision \( \hat{k}(\tau) \in \{0, 1, \ldots, K\} \), where \( \hat{k}(\tau) = 0 \) means that the flow is rejected, and \( \hat{k}(\tau) = k \geq 1 \) is the selected available slice. \( \hat{k}(\tau) \) depends on all the observations made and is hence \( F_\tau \)-measurable.

A correct decision is obtained when \( \hat{k}(\tau) = 0 \) if there is no slice with load below the threshold, or when \( \hat{k}(\tau) \) is an available slice. There may be multiple available slices, and we can further specify the admission criteria by refining the definition of a correct decision. It is not immediately obvious that any criterion is strictly better than one another, but we will consider algorithms pertaining to each of the following three criteria and compare them to one another. In all scenarios, we denote by \( \mathcal{C}(\mu) \subset \{0, 1, \ldots, K\} \) the set of correct answers given the server loads \( \mu \). We also define \( \mu_* = \min_k \mu_k \) as the load of the least loaded slice and \( k_*(\mu) \in \arg \min_k \mu_k \) as the corresponding slice. Finally we let \( k^*(\mu) \in \arg \max_{k: \mu_k < \gamma} \mu_k \) be the most loaded available slice, defined only when \( \mu_* < \gamma \). We assume that \( \mu \) is such that \( k_*(\mu) \) and \( k^*(\mu) \) are unique, and denote by \( \mathcal{M} \) the set of \( \mu \) satisfying this assumption. We remark that for \( \mu \) such that e.g. \( k_*(\mu) \) is not unique, we would need to modify our results and algorithms as this is done in [51]. For simplicity and clarity of the paper, we restrict ourselves to the case where \( \mu \in \mathcal{M} \). Our three admission criteria are:
1. Any-available-slice. Under this criterion, we have $C(\mu) = \{k : \mu_k < \gamma\}$ if $\mu_\star < \gamma$ and $C(\mu) = \{0\}$ otherwise.

2. Packing-slice. Here, we wish to select the most loaded available slice, referred to as the packing slice. This choice allows us to get a minimum number of active slices, and in some scenarios where the service rates of incoming flows are heterogeneous, to reduce the blocking rate. Under this criterion, $C(\mu) = \{k_\star(\mu)\}$ if $\mu_\star < \gamma$ and $C(\mu) = \{0\}$ otherwise and we denote $\mu_\star = \mu_{k_\star(\mu)}$.

3. Least-loaded-slice. Selecting the least loaded available slice is also a natural admission criterion, since it will tend to homogenize the loads of the slices, and hence ensure fairness (packet of the various flows experience similar delays) and low packet delay. Here, $C(\mu) = \{k_\star(\mu)\}$ if $\mu_\star < \gamma$ and $C(\mu) = \{0\}$ otherwise. While superficially similar to the problem considered by $[52]$, this criterion differs in the requirement that the slice be available which creates a discontinuity for loads near $\gamma$ and thereby disqualifies the methods considered in that paper.

Given one of the aforementioned admission criteria, we wish to design algorithms returning a correct answer with a fixed level of certainty. Note that since $\mu$ is unknown and measurements are inherently noisy, it is impossible to surely get a correct answer. We fix $\delta > 0$, and target $\delta$-PC ($\delta$-Probably Correct) algorithms, that is, algorithms which are guaranteed to return the correct answer with at least probability $1 - \delta$:

**Definition 1 ($\delta$-PC algorithms).** A joint measurement and decision algorithm is $\delta$-PC if and only if for any $\mu \in M$, $\Pr_{\mu}[\tau < \infty] = 1$ and $\Pr_{\mu}[\hat{k}(\tau) \notin C(\mu)] \leq \delta$.

The objective is to devise a $\delta$-PC algorithm with minimal expected measurement cost or sample complexity $\mathbb{E}_{\mu}[\tau]$ for the various envisioned admission criteria.

**Induced flow-level dynamics**

While this paper mainly focuses on devising efficient measurement schemes, it is worth mentioning the impact of the chosen admission criteria on the flow-level performance, i.e., on the flow blocking probabilities. To simplify the discussion below, we assume that the admission decisions are always correct, so that we can focus on the impact of the chosen admission criteria. The deviations caused by the fact that our algorithms may sometimes fail to output a correct decision are assessed numerically in the companion technical report $[57]$.

When the flows have different rates, then the selected admission criterion impacts the flow-level dynamics and blocking probabilities. It has been shown that with heterogeneous flows, the steady-state distribution of the population of
flows is sensitive to flow size distribution, arrival process and time scale \[22\], and we cannot analytically characterize the blocking rates. As a consequence, it is difficult to predict the behavior of any given admission controller. We will not investigate the trade-off achieved under different admission criteria, but we expect Packing-slice to have the best performance in terms of fairness and blocking probability, and Any-available-slice to have the worst, and we evaluate this in \[57\].

### 3.4 Best arm identification in admission control

To devise $\delta$-PC algorithms with minimal measurement cost for each admission criterion, we first derive lower bounds on this cost. For a given criterion, we show that the lower bound is the value of an optimization problem, whose solution specifies the optimal measurement process (it characterizes the numbers of times an algorithm with minimal cost should measure each slice before stopping). We then develop algorithms whose sampling and stopping rules perform this optimal measurement process.

#### Lower bounds

**Notations.** To state the lower bounds, we introduce the following notations. Let $\Lambda$ be the $(K - 1)$-dimensional simplex $\Lambda = \{w \in [0, 1]^K : \sum_k w_k = 1\}$. We denote by $d(a, b)$ the Kullback-Leibler divergence (KL-divergence) between two distributions of the same one-parameter exponential family, parameterized by means $a$ and $b$, respectively. $d_B(a, b)$ denotes this KL-divergence in the case of Bernoulli distributions. In the sequel, to avoid pathological cases where one cannot identify an available slice even with an infinite number of measurements, we assume that $\mu_\star \neq \gamma$. Furthermore, we introduce the information deviation function as $g_{j,k}(x) = d(\mu_k, (\mu_k + x\mu_j)/(1 + x)) + xd(\mu_j, (\mu_k + x\mu_j)/(1 + x))$ and its inverse $x_{j,k}(y) = g^{-1}_{j,k}(y)$. We use this to introduce the equilibrium function for a set of candidate arms $S$ and a target arm $k$

$$F_k(y; S) = \sum_{j \in S} \frac{d(\mu_k, (\mu_k + x_{j,k}(y)\mu_j)/(1 + x_{j,k}(y)))}{d(\mu_j, (\mu_k + x_{j,k}(y)\mu_j)/(1 + x_{j,k}(y)))}.$$ 

#### Lower bounds and the optimal measurement process.** Following the approach developed in \[50\], we identify the cost lower bounds, as well as the corresponding optimal fractions of time each slice should be measured. These fractions, denoted by $w^*(\mu) \in \Lambda$, depend on whether there is an available slice and on the admission criterion. The following propositions are established in \[57\] by examining the structure of each problem in detail and applying classical best arm identification theory.

**Proposition 1.** Assume that there is no available slice ($C(\mu) = \{0\}$). Then under any of the three admission criteria, any $\delta$-PC algorithm fulfills $E_{\mu}[\tau] \geq$
3.4. BEST ARM IDENTIFICATION IN ADMISSION CONTROL

\[ T_0(\mu)dB(\delta, 1 - \delta) \] where \( T_0(\mu) = \sum_{k=1}^{K} d(\mu_k, \gamma)^{-1} \). The optimal measurement process is given by, for all \( k \), \( w_k^*(\mu) = \frac{d(\mu_k, \gamma)^{-1}}{T_0(\mu)} \).

**Proposition 2.** [Any-available-slice] For the any-available-slice problem with \( C(\mu) \neq \{0\} \), any \( \delta \)-PC algorithm fulfills \( \mathbb{E}_\mu[\tau] \geq T_1(\mu)dB(\delta, 1 - \delta) \) with \( T_1(\mu) = d(\mu^*, \gamma)^{-1} \).

The optimal measurement process is given by, for all \( k \), \( w_k^*(\mu) = \mathbb{1}_{(k=k^*(\mu))} \).

**Proposition 3.** [Packing-slice] For the packing-slice problem with \( C(\mu) \neq \{0\} \), define \( S_P = \{ j : \mu_j < \mu^* \} \) and \( z^* = \min(d(\mu^*, \gamma), y^*) \) where \( y^* \) is the unique solution to the equation \( F_{k^*(\mu)}(y; S_P) = 1 \) \((y^* \text{ and } z^* \text{ are well defined})\). Any \( \delta \)-PC algorithm fulfills \( \mathbb{E}_\mu[\tau] \geq T_2(\mu)dB(\delta, 1 - \delta) \) where

\[ T_2(\mu) = \sum_{k: \mu_k > \mu^*} d(\mu_k, \gamma)^{-1} + \frac{1}{z^*} \sum_{k: \mu_k < \gamma} x_{k,k^*(\mu)}(z^*). \quad (3.1) \]

The optimal measurement process is given by, for all \( k \),

\[ w_k^*(\mu) = \frac{1}{T_2(\mu)} \left( \frac{x_{k,k^*(\mu)}(z^*) \mathbb{1}_{(\mu_k < \gamma)}}{z^*} + \frac{\mathbb{1}_{(\mu_k > \gamma)}}{d(\mu_k, \gamma)} \right). \quad (3.2) \]

**Proposition 4.** [Least-loaded-slice] For the least-loaded-slice problem with \( C(\mu) \neq \{0\} \), define \( S_{LL} = \{ j : \mu_j > \mu_* \} \) and \( z_* = \min(d(\mu^*, \gamma), y_*) \) where \( y_* \) is the unique solution to the equation \( F_{k^*(\mu)}(y; S_{LL}) = 1 \) \((y_* \text{ and } z_* \text{ are well defined})\). Any \( \delta \)-PC algorithm fulfills \( \mathbb{E}_\mu[\tau] \geq T_3(\mu)dB(\delta, 1 - \delta) \) where

\[ T_3(\mu) = \frac{1}{z_*} \sum_{k=1}^{K} x_{k,k^*(\mu)}(z_*). \quad (3.3) \]

The optimal measurement process is given by, for all \( k \),

\[ w_k^*(\mu) = \frac{x_{k,k^*(\mu)}(z_*)}{z_*T_3(\mu)}. \quad (3.4) \]

**Track-and-Stop algorithm**

Next, we describe the Track-and-Stop (TaS) algorithm, a generic algorithm that will be instantiated for the three admission criteria, and establish its asymptotic optimality (when \( \delta \) approaches 0).

**Sampling rule.** The measurement cost lower bounds and the corresponding optimal measurement process provide the design principle of the sampling rule. We follow the Track-and-Stop framework developed in [50]: the sampling rule is designed so as to track the optimal fractions \( w^*(\mu) \) of time each slice should be
measured. Here $\mu$ is unknown, and hence, for the $t$-th measurement, we track $\hat{w}^*(t) := w^*(\hat{\mu}(t - 1))$ instead, where $\hat{\mu}(t - 1)$ are the estimated slice loads from the $(t - 1)$-th first measurements. The algorithm will work as long as we can make sure that $\hat{\mu}(t)$ converges to $\mu$ almost surely. To this aim, the sampling rule includes a forced exploration phase: after $t$ measurements, slices that have not been measured more than $\sqrt{t}$ times are measured. If the algorithm is not in a forced exploration phase, it tracks the allocation $\hat{w}^*(t)$, i.e., it measures the slice $k_t \in \arg \max_k t\hat{\mu}_k(t) - n_k(t - 1)$, where $n_k(t - 1)$ is the number of times $k$ has been measured so far. Finally note that the sampling rule depends on the functions $\mu \mapsto \mu^*(\mu)$ specified by Propositions 1-4 for the various admission criteria.

Stopping rule. We maintain, depending on the admission criterion, a target slice $\hat{k}(t)$ that would be the output of the algorithm should it stop. If $C(\hat{\mu}(t - 1)) = \{0\}$, $\hat{k}(t) = 0$. Otherwise, $\hat{k}(t) = k^*(\hat{\mu}(t - 1))$ for the any-available-slice and least-loaded-slice criteria, and $\hat{k}(t) = k^*(\hat{\mu}(t - 1))$ for the packing-slice criterion. The target slice for any-available-slice coincides with the least-loaded-slice, but this is coincidence caused by the least loaded slice being the easiest discernable available slice. The stopping rule we use relies on a similar stopping criterion as in all previously devised pure exploration algorithms. Specifically, it is based on the Generalized Likelihood Ratio (GLR) statistics $Z_{k,k'}(t)$ evaluating the probabilities that given the observations, the targeted correct answer is $k = \hat{k}(t)$ or $k'$, see details in [50] and [51]. We stop when these GLR are large enough. The resulting statistical test can be summarized by comparing $Q(t) := \inf_{\lambda, \hat{k}(t) \notin C(\lambda)} \sum_{k=1}^K n_k(t - 1)d(\hat{\mu}_k(t - 1), \lambda_k)$ to an exploration threshold $\delta(t)$ appropriately chosen.

Decision rule. When the algorithm stops measuring after $\tau$ measurements, it returns the slice $\hat{k}(\tau)$.

The pseudo-code of the algorithm is presented in Algorithm 3. There, $n(t)$ denotes the vector counting the number of times each slice has been measured up to time $t$. The following theorem establishes the asymptotic (as $\delta$ goes to 0) optimality of TAS for any of our admission criteria, up to a factor 2.

**Theorem 4.** Let TAS be instantiated for any of our admission criteria with input functions $\mu \mapsto \mu^*(\mu)$ given in Propositions 2, 3, and 4, respectively. Select the exploration threshold equal to $f_\delta(t) = \log(Ct^2/\delta)$ with $C$ such that $C \geq e\sum_{i=1}^{\infty}(1/k^i)K(\log(Ct^2/\delta)^2/t^2)$. Then, TAS is $\delta$-PC, and its sample complexity satisfies for any $\mu \in \mathcal{M}$:

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau]}{d_B(\delta, 1 - \delta)} \leq 2T_1(\mu),$$  \hspace{1cm} (3.5)

where if $C(\mu) = \{0\}$, $i = 0$, and if $C(\mu) \neq \{0\}$, $i = 1$ (resp. $i = 2$ and $i = 3$) for the any-available-slice (resp. packing-slice and least-loaded-slice) criterion.
3.5 Numerical Results

We have run simulations to illustrate the performance of our joint measurement and admission strategies under the various admission criteria. We fixed the slice loads, and compared the measurement costs of our algorithms and those obtained by using a simple round robin sampling rule. Experiments with flow-level dynamics (accounting for flow arrivals and departures) are reported in [57].

Set-up

The system consists of $K = 8$ slices of equal capacities. We focus on flows in service of Conversational Voice, which has a packet frequency of about 50 packets/second, each packet with a size of 200 bytes [58]. We assume that each slice handle 25 of such flows with the required QoS. Measurement slots last 20 ms, so that the admission threshold $\gamma$ was fixed equal to $\gamma = (25 - 1) \times 50 \times 0.02 = 24$ packets/slot (this corresponding to a bit rate of 16 Mbps for the entire system).

The traffic loads $\mu$ on the different slices are generated as follows. We first fix the total load handled by the 8 slices, and consider three scenarios: (i) Low load with a total load of $8 \times 17$ packets/slot, (ii) Medium load with $8 \times 23$ packets/slot and (iii) High load with $8 \times 30$ packets/slot. We then assign the total load of the various slices using a multinomial distribution: each unit (1 packet/slot) of load is assigned to a slice chosen uniformly at random.
### CHAPTER 3. PAPER II: MEASUREMENT-BASED ADMISSION CONTROL IN SLICED NETWORKS: A BEST ARM IDENTIFICATION APPROACH

#### Table 3.1: Measurement costs in slots (averaged over 100 runs – confidence intervals are not shown due to space constraints but are typically small, with radius of the order of 5% of mean).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>TAS</th>
<th>Uniform</th>
<th>Lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Any-available-slice</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low load</td>
<td>12.2</td>
<td>27.7</td>
<td>0.843</td>
</tr>
<tr>
<td>Medium load</td>
<td>28.4</td>
<td>105</td>
<td>2.88</td>
</tr>
<tr>
<td>High load</td>
<td>571</td>
<td>3522</td>
<td>92.6</td>
</tr>
<tr>
<td><strong>Packing-slice</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low load</td>
<td>808</td>
<td>2800</td>
<td>245</td>
</tr>
<tr>
<td>Medium load</td>
<td>1790</td>
<td>5480</td>
<td>571</td>
</tr>
<tr>
<td>High load</td>
<td>897</td>
<td>3640</td>
<td>261</td>
</tr>
<tr>
<td><strong>Least-loaded-slice</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low load</td>
<td>390</td>
<td>1020</td>
<td>139</td>
</tr>
<tr>
<td>Medium load</td>
<td>565</td>
<td>1350</td>
<td>180</td>
</tr>
<tr>
<td>High load</td>
<td>821</td>
<td>2310</td>
<td>290</td>
</tr>
</tbody>
</table>

A new flow arrives and the algorithm needs to output a decision. In Scenarios (i) and (ii), there is obviously always an available slice. Hence the sample complexity or measurement cost should depend on the admission criterion. In Scenario (iii), it is possible that none of the slices is available, and in this case, the measurement cost does not depending on the chosen admission criterion.

We implemented Tas in Python 3.7 under the three admission criteria, and compared their performance to that of a naive algorithm using the same stopping and decision rules but with a sampling rule picking slices in a round-robin manner. This benchmark is useful as it allows us to see the impact of only our intelligent sampling rule, removing the impact of confidence levels and other performance guarantees. The level of confidence for the stopping rule was fixed to \( \delta = 0.01 \).

Each algorithm was tested on 100 independent runs, and in each run the traffic intensity \( \mu \) was regenerated.

#### Results

The results are shown in Table 3.1 including the averaged lower bounds from Section 3.4 for comparison. They are also visualized in Figure 3.2 normalized around the uniform sampling sample complexity for visibility.

We observe that the problems with the three admission criteria have different difficulties. As expected, the any-available-slice criterion leads to a much lower measurement cost, except for the high load scenario where there is often no or a single available slice. In all scenarios, Tas significantly outperforms a naive
algorithm using uniform sampling: the improvements in the measurement cost are by a factor 2 to 8 on all problems. TAS measurement costs are not so close to the lower bound; this is due to the moderate confidence level of $\delta = 0.01$. For higher confidence levels (smaller $\delta$) the gap between TAS performance and the lower bound becomes much smaller [57].

Another important result, not presented in these figures, is that during our experiments, there was no occasion during which an unavailable slice was chosen. This suggests that TAS is more conservative compared to what the targeted confidence level $\delta = 0.01$ imposes.

3.6 Conclusions

In this paper, we investigated the problem of admission control in network slicing where before admitting or rejecting a flow, the slice utilization in terms of network traffic of several slices needs to be measured. Inspired by Best Arm Iden-
tification methods, we designed a framework to allow for robust admission control with confidence guarantees. We applied this framework to devise optimal joint measurement and admission schemes realizing three different admission criteria. We verified, using simulations, the efficiency of our algorithms and showed their advantage over a baseline measurement method. In this paper, we assumed that the unknown parameters, driving the admission decisions and implicitly learned by our algorithms, just dictate the loads of the slices. In real sliced networks, there might be other types of uncertainty (e.g. unknown flow rate or unknown slice capacity), and we plan to extend our methods and results to deal with these additional uncertainties.

Acknowledgment

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.
Chapter 4

Paper III: Change point detection with adaptive measurement schedules for network performance verification

Foreword

Ever since the beginning of the research project, we have known that one main motivation for adaptive measurement schedules is the ability to not only choose which measurement to take, but how often. For a long time, it was difficult to find a suitable use case for this idea, but an opportunity arose in a collaboration with the multi-institute research project IMMINENCE, who wanted to analyze some interesting behavior in Non-Standalone 5G networks, which would sometimes suddenly change behavior drastically. This gave us the idea to research adaptive measurement schedules in changepoint detection. As such, the theory within Paper III is disconnected from that of Papers I and II, the connection being the desire to intelligently minimize the measurement costs within orchestration problems.

We analyzed the problem both from a frequentist and a Bayesian perspective, where the frequentist approach is the one presented in Paper III. The Bayesian approach showed promise but could not yield conclusive results, and the modeling assumptions required were considered too strict to be practical. That said, it remains an interesting avenue for future work.

The paper has been submitted for review.
Change point detection with adaptive measurement schedules for network performance verification
Simon Lindstål, Alexandre Proutiere, Andreas Johnsson

Abstract

When verifying that a communications network fulfills its specified performance, it is critical to note sudden shifts in network behavior as quickly as possible. Change point detection methods can be useful in this endeavor, but classical methods rely on measuring with a fixed measurement period, which can often be suboptimal in terms of measurement costs. In this paper, we extend the existing framework of change point detection with a notion of physical time. Instead of merely deciding when to stop, agents must now also decide at which future time to take the next measurement. Agents must now minimize the necessary number of measurements pre- and post-change, while maintaining a trade-off between post-change delay and false alarm rate. We establish, through this framework, the suboptimality of typical periodic measurements and propose a simple alternative, called crisis mode agents. We show analytically that crisis mode agents significantly outperform periodic measurements schemes. We further verify this in numerical evaluation, both on an array of synthetic change point detection problems as well as on the problem of detecting traffic load changes in a 5G test bed through end-to-end RTT measurements.

4.1 Introduction

A wide spectrum of applications with varying network requirements have emerged with the promises of 5G, ranging from cyber-physical control applications and robotic telesurgery, to augmented reality and video streaming. To maintain trust and user satisfaction, network operators must ensure that their networks perform well and fulfill the requirements from a user perspective. If the performance drops, the operator should take a corrective action such as reallocating resources, provisioning more, or blocking low-priority applications. To take such actions, however, the operator must first detect the performance drops. As such, the operator would monitor the network by measuring aspects of its performance, for instance by measuring the round-trip time (RTT) of certain classes of packets, either end-to-end in cooperation with the user, or within certain partitions of the network. In Figure 4.1, we illustrate a scenario where the RTT is monitored between an agent in the radio cell and a reflector in the cloud.

One critical aspect of such monitoring, also depicted in Figure 4.1, is detecting sudden shifts in the network, such as major changes in traffic volume or handovers. In such cases, the network behavior may shift in a significant but
Figure 4.1: An agent in a radio cell probes the RTT performance statistics to a reflector in the cloud, with the aim of detecting and adapting to sudden changes triggered by traffic variations (induced by e.g. handovers).

often well-behaved way. Then, detecting these shifts become an instance of the classical problem of change point detection, which is the task of detecting when a random variable sampled many times (such as a monitored network statistic) changes distribution. The classical objective is to minimize the required number of measurements post-change, while ensuring that the agent does not raise false alarms too often. In this classical framework, measurements are implicitly assumed to be taken periodically, which means that the number of measurements post-change exactly translates to the delay – the time it takes for the agent to detect the change.

In the classical change detection framework, one does not really account for the cost of measurements\footnote{There, of course, the post-change measurement cost is considered through the number of measurements taken after the change. But compared to the measurements taken before the change, this cost is of less importance as one can often afford extra measurements in times of crisis.}. In practice however, pre- and post-change measurements increase the network utilization and the energy consumption, and could potentially cause a drop in performance of existing services. To represent the measurement cost, we introduce a new change point detection framework where the agent can schedule over time and in an adaptive manner her measurements. Intuitively, the agent would sometimes have greater or smaller confidence that a change has not yet occurred, based on its measurement history. We argue that by increasing measurement frequency whenever a change is suspected to have occurred, the total measurement frequency pre-change can be reduced while maintaining the same or better detection performance in terms of post-change delay, false alarm rate and number of post-change measurements.

**Contributions and paper outline.** In the next section, we summarize the existing literature related to change point detection and network monitoring. In Section 4.3 we start by recalling the classical frequentist change point detection framework and the main results therein. We propose a new framework to account for the measurement costs, and discuss the performance trade-offs in terms of
false alarm rate, delay, and measurement costs. To the best of our knowledge, this paper is surprisingly the first to study change point detection problems accounting for measurement costs in this way. Most of our results are presented in Section 4.4. We provide a lower bound on the number of measurements post-change satisfied by agents with given false alarm rate and delay guarantees. We then discuss the pre-change average measurement frequency and show that non-adaptive schedules will need to measure more frequently if the false alarm rate guarantee becomes more demanding, even if the delay requirement remains unchanged. In contrast, we introduce simple adaptive measurement agents called crisis mode agents which can maintain a fixed average measurement frequency even when the false alarm rate constraint becomes stricter. Finally, in Section 4.5 we evaluate our agents compared to a non-adaptive benchmark on both synthetic data as well as data from a real 5G testbed.

4.2 Related work

Change point detection

The classic formulation of change point detection from a frequentist perspective was crafted by Lorden [63]. In his paper, change point detection is considered from a minimax perspective: his objective was the maximal delay from the change point to the alarm, taken over all possible finite change times. For this problem he showed the importance of tracking the maximal cumulative sum of log-likelihood ratios, referred to as CUSUM, by showing a lower bound on the maximum delay and showing how agents which consider only the CUSUM statistic meet this bound asymptotically as the bound on false alarm rate grows to infinity. The test was earlier introduced by Page [64] but was only introduced with heuristic arguments, whereas Lorden showed its asymptotic optimality. Moustakides [65] later showed that this stopping rule is in fact exactly optimal, even in finite time.

Another important milestone came from Lai [66], who considered a similar problem where the pre- and post-change measurements were not i.i.d. but had some history dependence, such as markovian measurements. He defined a new CUSUM statistic and introduced a heuristic test, limiting the stopping search to a certain window. He proved a lower bound for the delay within this framework and showed that his heuristic rule is nearly optimal.

After this, a variety of extensions to frequentist change point detection have been proposed and studied. Lai and Xing studied the case where both the pre- and post-change distributions are unknown and showed that this creates new bounds which are heavily dependent on the change time [67]. Recently, Maillard [68] used Laplace concentration results to show finite time delay lower bounds, and Berrett and Yu studied the problem in the context of local differential privacy [69]. The methods have also been applied in diverse fields such as security [70], public health [71] and network management [72, 73].
As can be seen from the above, frequentist change point detection has been well studied from a vast number of perspectives. However, these perspectives all have one thing in common: they consider only the number of measurements taken before stopping, pre-change and post-change. This implicitly means that measurements are taken periodically, and that their cost cannot be modelled or optimized. Our work is novel in the sense that measurements can be scheduled over time. Our framework then allows us to study agents trying to minimize their measurement cost while maintaining good performance in both false alarm rate and delay.

A well-studied alternative to frequentist change point detection is Bayesian change point detection, where the change point is modelled to occur with some probability distribution over the set of positive integers. This framework was initially described by Girschick and Rubin [74], maximizing the expected gain per observation. Now the most commonly used framework either minimize a trade-off between the observation cost post change and the false alarm rate probability, or minimize the expected cost post-change subject to a constraint to the false alarm rates. The former was extensively studied by Shiryaev [75] and the latter by Roberts [76], both independently coming up with the same detection procedure, based on a sequential probability ratio test. This procedure is thus known as the Shiryaev-Roberts procedure, and was applied to the frequentist framework by Pollak, who showed that a variation of this scheme is asymptotically optimal [77]. In this work we do not study the Bayesian framework as it imposes greater constraints on the model (a known distribution of change points) as well as introduces several complications when combined with adaptive measurement schedules, but doing so is still an interesting direction for future work.

In recent years, a number of interesting extensions to the classical change point detection problems, including dealing with non-i.i.d. time series, applying change-of-measure to approximate the average run time to false alarm, dealing with multiple changes, dealing with high-dimensional data and dealing with adversarial post-change distributions have been proposed. They are not immediately related to our extension of adaptive measurement schedules and so, we have not included them, but the interested reader is referred to the survey by Xie et al. [78].

Adaptive measurement schedules in network monitoring

Performance monitoring and assessment of network characteristics are important applications of change-point detection. Network operators are accustomed to use active measurement protocols such as such as ICMP [79] and TWAMP [80] with the objective of assessing performance metrics including one-way delay (OWD), round-trip time (RTT), and throughput as defined in e.g. ITU-T Recommendation Y.1540. While these protocols, metrics, and tools are heavily deployed in industry, they remain hot research topics [81].

The literature is rich on proposals for measurement frameworks and platforms that enable adaptive measurement schedules. For example, Wang and Su [17] and
CHAPTER 4. PAPER III: CHANGE POINT DETECTION WITH ADAPTIVE MEASUREMENT SCHEDULES

Figure 4.2: In (a), the classical change point detection framework when adapted naively requires the agent to measure periodically, before raising an alarm at some point after $\nu$. In (b), the agent is instead allowed to measure with adaptive schedules, potentially reducing both delay and the number of required measurements.

Adrichem et al [13] targets the challenges with measurement overhead in software defined networking (SDN) by proposing mechanisms and methods for adaptive polling and sampling. Both frameworks have the capability of modifying the measurement frequency for reduced overhead while maintaining the observability to a certain degree. Further, Xie et al addresses adaptive monitoring for analysis and orchestration in 5G systems [18]. More specifically, they propose a framework for service assurance where monitoring can be adapted to the analytical results with the objective of achieving a balance between monitoring cost and network performance. Steinert and Gillblad proposed a heuristic method for adapting the measurement frequency for assessment of RTT over a network path [82]. The methods rely on computing KL-divergence between empirically computed distributions of RTT. Further, Quan et al describes an adaptive probing system aimed at detecting outages in edge networks [83]. The network is probed at regular intervals to guarantee freshness whereas the frequency is increased to resolve uncertainty when necessary.

By comparison, our work provides a complementary and more generic approach for adaptive measurement schedules with performance guarantees. Further, our approach can be incorporated into existing architectures, frameworks, and software packages.

4.3 Change point detection and adaptive measurement schedules

In this section, we first present the classical framework for change point detection, where measurements are taken periodically. We then extend this framework to account for measurement costs and to include adaptive measurement schedules.
4.3. CHANGE POINT DETECTION AND ADAPTIVE MEASUREMENT SCHEDULES

Change point detection: the classical framework

In change point detection, an agent is able to take a sequence of measurements \( \{X_k\}_{k \in \mathbb{Z}^+} \). These measurements are independent random variables. Before the change occurring at the \( \nu \)-th measurement for some unknown \( \nu \in \mathbb{N} \), the distribution of the measurements is \( f_0 \), while after the \( \nu \)-th measurement, it becomes \( f_1 \). We assume that \( f_0 \) and \( f_1 \) are known; but this assumption can be relaxed as done in [67]. We denote by \( \mathbb{P}_\nu \) (resp. \( \mathbb{E}_\nu \)) the probability measure (resp. the expectation) of the measurements when the distribution change occurs at the \( \nu \)-th measurement, and by \( \mathbb{P}_\infty \) (resp. \( \mathbb{E}_\infty \)) the corresponding quantities when no change occurs.

The agent is tasked with raising an alarm, here taken to be a stopping time \( N \) (defined w.r.t. the natural filtration), as soon as possible after the change has occurred but no sooner. In particular, the agent should avoid stopping when there is no change, to keep down the cost induced by false alarms. The agent’s objective is then to balance the delay, i.e., the number of measurements taken after the change or delay and before stopping, and the false alarm rate.

False alarm rate. It is tempting to model the false alarm rate as \( \mathbb{P}_\infty (N < \infty) \), and hence to aim at minimizing the number of post-change measurements while maintaining this probability below a specified threshold. Lorden [63] showed that this design objective was overly restrictive, and proposed instead to capture the false alarm rate using the “average run length” \( \mathbb{E}_\infty [N] \).

Worst-case delay. Regarding the number of post-change measurements or delay, the performance metric used in the classical framework is the worst-case delay, defined as\(^2\)

\[
\sup_{\nu \in \mathbb{N}} \left( \text{ess sup} \mathbb{E}_\nu \left[ (N - \nu)^+ | \mathcal{F}_\nu \right] \right),
\]

where \( \mathcal{F}_\nu \) is the \( \sigma \)-algebra generated by the measurements prior to \( \nu \).

Trading false alarm rate and delay. The trade-off between the two performance metrics has been thoroughly investigated in the classical framework, first by Lorden [63] and then by Moustakides [65]. The latter established that Page’s stopping rule based on the CUSUM achieves an optimal trade-off characterized by the following result. Page’s stopping rule minimizes the worst-case delay subject to an average length constraint of the type \( \mathbb{E}_\infty [N] \geq \gamma \). More precisely, under the latter constraint, the following asymptotic lower bound for the worst-case delay holds and is achieved by Page’s rule:

\[
\sup_{\nu \in \mathbb{N}} \left( \text{ess sup} \mathbb{E}_\nu \left[ (N - \nu)^+ | \mathcal{F}_\nu \right] \right) \geq (D(f_1||f_0)^{-1} + o(1)) \log(\gamma),
\]

\(^2\)The essential supremum “ess sup” of a random variable \( X \) is defined as the smallest value \( a \) such that the probability of \( X \) being greater than \( a \) is 0. Note that in our case, the underlying probability measure is \( \mathbb{P}_\nu \).
where $D(f_1||f_0)$ denotes the Kullback-Leibler divergence between the distributions $f_1$ to $f_0$ and $o(1)$ vanishes as $\gamma \to \infty$. In other words, there will always exist some $\nu \in \mathbb{N}$ such that there is a (non-zero under measure $\mathbb{P}_\nu$) subset of $\mathcal{F}_\nu$ where the delay is greater than the right hand side expression of equation (4.2).

Change point detection with adaptive measurement schedules

In the classical framework for change point detection, measurements are taken periodically, with predefined period (say 1 – so that there is no distinction between the number of measurements taken and physical time). Using this framework, we are unable to represent scenarios where the agent may actually schedule her measurements in real-time with the aim at balancing the delay expressed in physical time and the number (or equivalently the cost) of measurements. These scenarios require a new framework in which after each measurement, the agent either decides to raise an alarm or schedule the next measurement (i.e., decides the time at which the next measurement will be taken). Formally, the decisions made by the agent are as follows:

(a) **Measurement schedules.** For any $k \geq 1$, after the $(k - 1)$-th measurement whose outcome is $X_{k-1}$, the agent schedules the next measurement at interval $\tau_k$ after the previous one. $\tau_k$ is $\mathcal{G}_{k-1}$-measurable, where $\mathcal{G}_{k-1}$ is the $\sigma$-algebra generated by $(\tau_1, X_1, ..., \tau_{k-1}, X_{k-1})$. For convenience, $\mathcal{G}_0$ is defined as the trivial $\sigma$-algebra.

(b) **Stopping rule.** After any measurement, the agent may decide to raise an alarm. This decision is dictated by a stopping time $N$ adapted to the filtration $(\mathcal{G}_k)_{k\geq0}$. Here the event $\{N = n\}$ means that the agent raises an alarm after the $n$-th, and it happens at time $t_n = \sum_{k=1}^{n} \tau_k$.

The time at which the distribution shift occurs is now a real number $\nu \in \mathbb{R}_+$. As before, we assume that the random variables $(X_k)_{k\geq0}$ are independent, and that the distribution of $X_k$ is $f_0$ if $t_k < \nu$ and $f_1$ if $t_k \geq \nu$. For a given agent, we denote by $n_\nu$ the number of measurements taken before or at $\nu$: $n_\nu = \sup\{n \leq N : t_n \leq \nu\}$. Note that $n_\nu$ is a random variable but is a stopping time adapted to the filtration $(\mathcal{G}_k)_{k\geq0}$. This is because while it depends on the next time of measurement $t_{n_{\nu} + 1}$ this time is assumed to be known as it is a controlled variable. In this new framework, the agent needs to balance the false alarm rate, the delay, but also the number of measurements taken before and after the distribution change-point. We formalize these performance metrics below.

**False alarm rate.** As in the classical framework, we capture this rate using the average run length, defined here as $\mathbb{E}_\infty[t_N]$.

**Worst-case delay.** The delay is now measured in physical time, which means that the agent is interested in minimizing $\mathbb{E}[t_N] := \sup_{0 \leq \nu < \infty} (\text{ess sup} \mathbb{E}_\nu[(t_N - \nu)^+|\mathcal{G}_{n_\nu}])$.

\[3\text{The notation } \mathbb{E}[t_N] \text{ is similar as that used in Lai's seminal paper [66].}\]
4.3. CHANGE POINT DETECTION AND ADAPTIVE MEASUREMENT SCHEDULES

Pre- and post-change measurement costs. The key motivation for introducing adaptive measurement schedules in our new framework is the ability to reduce the day-to-day operations cost. If distribution changes occur rarely, this cost is mainly related to the frequency at which measurements are taken before the change-point. The number of measurements post-change may seem less important, as one can often afford extra measurements in times of crisis. However, the number of measurements is still interesting, partly because they are required in order to have a complete picture of measurement cost and partly because the number of measurements post-change is intimately involved in the trade-off between false alarm rate and delay, as we will demonstrate in the next section.

Pre-change measurement cost. To describe the cost related to the measurements taken before the change, we actually look at the frequency of these measurements should the algorithm never stop. Let $(\tau_k)_{k \geq 1}$ denote the lengths of the inter-measurement intervals. We define $E_\infty[\hat{\tau}] = \lim \inf_{n \to \infty} E_\infty[\frac{1}{n} \sum_{k=1}^{n} \tau_k]$ as the metric used to assess the cost of pre-change measurements. $E_\infty[\hat{\tau}]$ corresponds to the long-term average inter-measurement times if the measurement schedules form an ergodic process. We will further justify the definition of $E_\infty[\hat{\tau}]$ in the next section for the class of algorithms defining the stopping rule and the measurement schedules based on the CUSUM statistics. For any algorithm in this class, we show that the actual average inter-measurement times before the algorithm stops is upper bounded by $E_\infty[\hat{\tau}]$ (See Lemma 2).

Post-change measurement cost. To quantify the cost of measurements taken after the change, we use a similar metric as that used to describe the delay. More precisely, we use:

$$E[N] := \sup_{0 \leq \nu < \infty} \left( \text{ess sup} \ E_{\nu} [(N - n_{\nu})^{+} | G_{n_{\nu}}] \right).$$

Trade-offs and objectives. As in the classical change-point detection problem, the objective is to guarantee a false alarm rate, say ensuring that $E_\infty[t_N] \geq \gamma$, while keeping the delay $E[t_N]$ and the measurement costs $1/E_\infty[\hat{\tau}]$ and $E[N]$ as low as possible. Observe that here, the agent is allowed to measure arbitrarily often, and hence the trade-off between the false alarm rate $E_\infty[t_N] \geq \gamma$ and the worst case delay $E[t_N]$ becomes meaningless. Indeed, it is always possible to guarantee sufficiently low false alarm rates and delays, by just increasing the frequency at which measurements are taken. It makes then sense to impose constraints on both the false alarm rate and the worst case delay, and to attempt to minimize the measurement costs subject to these requirements. To encode the latter, we introduce, for any $\gamma > 0, \beta > 0$, the notion of $(\gamma, \beta)$-compliant algorithms:

**Definition 2.** An agent is $(\gamma, \beta)$-compliant if it fulfills $E_\infty[t_N] \geq \gamma$ and $E[t_N] \leq \beta$.

Our objective is then to devise a $(\gamma, \beta)$-compliant agent with minimum measurement costs $1/E_\infty[\hat{\tau}]$ and $E[N]$. While our framework allows any $\gamma$ and $\beta$, we
observe that the detection problem is trivial unless $\gamma > \beta$, otherwise the agent which stops immediately and sets $\tau_1 = \beta$ is always optimal.

**Change point detection in network performance verification:** When verifying a network’s performance, it is important to as quickly as possible detect if an event occurs (say at time $\nu$) that drastically changes network performance (we call these crisis events), so that an appropriate root cause analysis can be done and corrective actions can be taken. Thus, the physical time delay $E[t_N]$ is an important constraint, often playing a larger role than $E[N]$. Typically, the performance requirements of a service will be violated if some statistic such as RTT or throughput leaves a range of acceptable values. Often, the distribution $f_0$ of normal values is either well-known through gathered data or defined as part of a standard where the 3GPP specification on “Service requirements for cyber-physical control applications in vertical domains” serves as one example. When crisis events occur, it may not be known what the exact distributions $f_1$ of the statistic or measurements are during these events, but they are often generated by a similar standard or generate a composite hypothesis.

### 4.4 Analysis and algorithms

In this section, we first provide, for our new framework, a fundamental lower bound on the post-change measurement cost $E[N]$, satisfied by any $(\gamma, \beta)$-compliant agent. We then consider agents whose decisions rely on the CUSUM statistics and show that these agents are well-behaved in several senses. Next, we analyze the performance of agents with periodic schedules in our new framework, and explain their limits and issues. We finally introduce and analyze CUSUM-based crisis mode agents and show that they outperform any periodic agents.

**Minimal post-change measurement cost**

Our first result is a lower bound of the post-change measurement cost $E[N]$, satisfied by any $(\gamma, \beta)$-compliant agent. Assume that $f_0$ is absolutely continuous with respect to $f_1$ and let $D(f_1||f_0)$ denote the Kullback-Leibler divergence between distributions $f_1$ and $f_0$.

**Theorem 5.** Assume that $D(f_1||f_0) > 0$. For any $(\gamma, \beta)$-compliant agent, we have: for any fixed $\beta > 0$ and as $\gamma \to \infty$,

$$E[N] \geq (D(f_1||f_0)^{-1} + o(1)) \log \left( \frac{\gamma}{\beta} \right). \quad (4.3)$$

The above result can be seen as the equivalent of the lower bound (4.2) in the classical change detection framework. In what follows, we will construct $(\gamma, \beta)$-compliant agents whose post-change measurement cost is asymptotically upper bounded by $D(f_1||f_0)^{-1} \log \left( \frac{\gamma}{\beta} \right)$, and hence our lower bound is tight.
Cumulative sum statistics

In general, the decisions taken after each measurement could depend on all previous observations. The resulting measurement scheduling policy and stopping rule would be even difficult to describe. It makes then sense to restrict our attention to agents whose decisions are taken solely based on the CUSUM statistics. We know that in the classical change point detection framework, Page’s stopping rule is optimal while indeed it uses the CUSUM only. We will also establish that agents using the CUSUM statistics can be efficient.

The CUSUM is defined, after \( n \) measurements, as
\[
S_n = \max_{1 \leq i \leq n} \sum_{k=1}^{n} \log \left( \frac{f_1(X_i)}{f_0(X_i)} \right).
\]
We consider agents with scheduled measurements characterized by, for any \( n \in \mathbb{N} \),
\[
\tau_{n+1} = g(S_n)
\]
for some function \( g \), and with stopping rule defined by \( N = \min\{n \geq 1 : S_n \geq S^{(0)}\} \) for some threshold \( S^{(0)} > 0 \). The following lemma provides useful properties of such agents.

**Lemma 2.** Consider an agent as defined above: \( \forall n \geq 0, \tau_{n+1} = g(S_n) \) for some non-increasing function \( g \) and \( N = \min\{n \geq 1 : S_n \geq S^{(0)}\} \). Then we have:

(i) \( \mathbb{E}_\nu[(N - n_\nu)^+ | G_{n_\nu}] \) is independent of the post-change measurement intervals \( \{\tau_n\}_{n=n_\nu+1}^N \).

(ii) For any \( n \geq 1 \), \( \mathbb{E}_\infty[\tau_n | N > n - 1] \geq \mathbb{E}_\infty[\tau_n] \geq \mathbb{E}_\infty[\hat{\tau}] \).

In the above lemma, (i) the measurement schedules after the change do not affect the post-change measurement cost; whereas (ii) justifies the choice of \( \mathbb{E}_\infty[\hat{\tau}] \) as a metric to capture the pre-change measurement cost.

**Periodic schedules**

A simple way to schedule measurements is to take them periodically. This means that for all \( n \geq 1 \), \( \tau_n = \tau \). We investigate how \( \tau \) and the stopping rule should be chosen so that the corresponding agent is \((\gamma, \beta)\)-compliant.

We start the analysis by quantifying the *price* to pay in terms of pre-change measurement cost (or equivalently here, in terms of measurement frequency) for an agent to be \((\gamma, \beta)\)-compliant. We establish the following result valid for any \((\gamma, \beta)\)-compliant agent with periodic schedules.

**Proposition 5.** (i) Fix the average run length guarantee \( \gamma > 0 \). For any \( \tau > 0 \), there exists \( \beta(\tau) \) such that for all \( \beta < \beta(\tau) \), there exists no periodic agent with measurement period \( \tau \) that is \((\gamma, \beta)\)-compliant.

(ii) Fix the delay guarantee \( \beta > 0 \). For any period \( \tau > 0 \), there exists \( \bar{\gamma}(\tau) \) such that for any false alarm rate guarantee \( \gamma > \bar{\gamma}(\tau) \), there exists no periodic agent with measurement period \( \tau \) that is \((\gamma, \beta)\)-compliant.

The first statement of the proposition is rather intuitive. When we fix the false alarm rate guarantee, we can only improve the delay guarantee by increasing the measurement frequency. The second statement is less intuitive, and also indicates
that for any fixed delay guarantee, we can only improve the false alarm rate by increasing the measurement frequency. Later, we show that we can devise agents with non-periodic schedules under which the pre-change measurement cost does not depend on the false alarm rate guarantee. Hence, agents with periodic schedules seem to pay a price in terms of measurements to become \((\gamma, \beta)\)-compliant higher than necessary.

We now study periodic agents whose decisions are taken solely based on the CUSUM statistics. The stopping rule is defined through the threshold \(S^{(0)}\):

\[
N = \min\{n \geq 1 : S_n \geq S^{(0)}\}.
\]

Using such a stopping rule, we construct a \((\gamma, \beta)\)-compliant and periodic agent whose measurement periods depend solely on the CUSUM statistics, and that exhibits minimal post-change measurement cost (it matches the lower bound \((4.3)\)). To this aim, we introduce \(\xi = \max_{r \geq 0} \mathbb{E}_{X \sim f_1} [Z - r | Z \geq r]\), where \(Z = \log f_1(X)/f_0(X)\), and we denote \(I = D(f_1||f_0)\) for brevity.

**Proposition 6.** Select \((\tau, S^{(0)})\) satisfying:

\[
\tau = \frac{\beta I}{S^{(0)}} + \xi \quad \text{and} \quad S^{(0)} = \log \left( \frac{\gamma}{\tau} \right).
\] (4.4)

When \(\gamma > \beta\), the above system of equations has two solutions, and we pick that with the greatest inter-measurement time. The agent defined by \(\tau\) and \(S^{(0)}\) is \((\gamma, \beta)\)-compliant and fulfills as \(\gamma \to \infty\),

\[
\mathbb{E}[N] \leq (D(f_1||f_0)^{-1} + o(1)) \log \left( \frac{\gamma}{\beta} \right).
\]

**Crisis mode schedules**

As mentioned above, agents with periodic schedules have the issue that if you wish to strengthen the guarantee on the false alarm rate (or the delay), then the only solution consists in increasing the measurement frequency. In other words, pushing the false alarm rate down comes at the expense of a higher pre-change measurement cost. We show that we can devise simple but non-periodic agents that do not suffer from this issue. Specifically, we present agents who can operate in two distinct modes: when in calm mode, she schedules her measurement periodically, whereas in crisis mode, she measures as quickly as possible. The idea behind such a behavior is to accept greater measurement costs whenever a change is likely to have occurred in order to save costs when there is no reason to raise an alarm.

The proposed non-periodic agent, referred to as crisis mode agent, only uses the CUSUM statistics to take decisions, and is defined through two thresholds \(S^{(0)}\) and \(S^{(1)}\), and a measurement period \(\tau\) for the calm mode. After the \(n\)-th measurement, the CUSUM \(S_n\) is first compared to \(S^{(0)}\), and if it exceeds \(S^{(0)}\), the agent raises an alarm; otherwise, \(S_n\) is compared to \(S^{(1)}\). If \(S_n \leq S^{(1)}\), the agent...
Algorithm 4 Change point detection with crisis mode

**Input:** Crisis threshold $S^{(1)}$, stopping threshold $S^{(0)}$, calm period $\tau$

**Initialization:** $S_0 = 0$

for $n = 1,...$ do

  if $S_{n-1} > S^{(0)}$ then
    Stop and raise alarm.
  else if $S_{n-1} > S^{(1)}$ then
    $\tau_n := 0$
  else
    $\tau_n := \tau$
  end if

  Wait for time period $\tau_n$.

  Obtain measurement outcome $X_n$.

  Update $S_n := \log \frac{f_1(X_n)}{f_0(X_n)} + \max(0, S_{n-1})$

end for

remains calm and schedules the next measurement $\tau_n = \tau$ seconds later, else the agent enters the crisis mode and immediately takes the $(n + 1)$-th measurement, i.e., $\tau_n = 0$. The pseudo-code of our agent is presented in Algorithm 4.

Our main result is stated in the following theorem: crisis mode agents can be tuned so as to be $(\gamma, \beta)$-compliant and so that the pre-change measurement cost does not depend on $\gamma$. In particular, using crisis mode agents, we can push the false alarm rate (by increasing $\gamma$) without paying any price in terms of pre-change measurement cost.

**Theorem 6.** Fix the delay requirement $\beta$. Assume that $f_1$ and $f_0$ are such that $Z$ is sub-gaussian under $f_1$ with parameter $\sigma_Z$, that is, there exists $\sigma_Z > 0$ such that for any $z$ it holds that $\Pr_{X \sim f_1}(|Z| \geq z) \leq 2 \exp\left(-\frac{z^2}{\sigma_Z^2}\right)$. Then, for any $\gamma > 0$, there is a $(\gamma, \beta)$-compliant crisis mode agent with pre-change measurement cost satisfying $\mathbb{E}_\infty[\hat{\tau}] \geq \tau_c$, for some constant $0 < \tau_c < \infty$ that does not depend on $\gamma$.

In the sequel, when we refer to $\sigma_Z$, we refer to the smallest value this constant can take. Observe that by Lemma 2(ii), the constant $\tau_c$ introduced in Theorem 6 provides an upper of the average pre-change inter-measurement times. Crisis mode agents are in stark contrast with periodic agents, for which it was shown to be impossible to maintain the same (average) measurement frequency while decreasing the false alarm rates. In practice, whenever the false alarm rate is a concern, crisis mode agents will significantly outperform periodic agents. This will be confirmed by our numerical experiments presented in the next section.

**Proof sketch.** The proof is constructive and proceeds in two main steps.

**Step 1. Bounding the worst-case delay.** In the first step, we derive an upper bound
on the worst-case delay of an agent parametrized by \((\tau, S^{(0)}, S^{(1)})\). This is achieved using concentration-of-measure results to show that:

\[
\sum_{n=n_{\nu}+1}^{\infty} \mathbb{P}_\nu(S_n \leq S^{(1)}|N > n_{\nu}) \leq \frac{\exp(S^{(1)} I/\sigma_Z^2)}{1 - \exp(-I^2/2\sigma_Z^2)}
\]

This implies that the worst-case delay is upper bounded as follows:

\[
\mathbb{E}[t_N] \leq \tau \frac{\exp(S^{(1)} I/\sigma_Z^2)}{1 - \exp(-I^2/2\sigma_Z^2)}.
\]

From this upper bound, we can guarantee the delay lower than \(\beta\) by just selecting \(\tau\) and \(S^{(1)}\) such that the r.h.s. of the previous inequality is smaller than \(\beta\). Such a selection is independent of \(\gamma\). Note that by definition, the pre-change measurement cost \(\mathbb{E}_\infty[\hat{\tau}]\) only depends on \(\tau\) and \(S^{(1)}\). Hence, if we manage to tune \(S^{(0)}\) to achieve the desired guarantee on the average run length (it should be larger than \(\gamma\)), then we would have designed an agent with pre-change measurement cost \(\mathbb{E}_\infty[\hat{\tau}]\) bounded by a constant independent of \(\gamma\).

**Step 2. Controlling the average run length.** Next, we establish a connection between the average run length \(\mathbb{E}_\infty[t_N]\) and the pre-change measurement cost \(\mathbb{E}_\infty[\hat{\tau}]\). Specifically, using Lemma 2, we prove that \(\mathbb{E}_\infty[t_N] \geq \exp(S^{(0)})\mathbb{E}_\infty[\hat{\tau}]\). We further leverage martingale arguments to derive a lower bound on \(\mathbb{E}_\infty[\hat{\tau}]\), and establish that it is greater than \(\tau(1 - \exp(-S^{(1)}))\). We deduce that:

\[
\mathbb{E}_\infty[t_N] \geq \tau(1 - \exp(-S^{(1)})) \exp(S^{(0)}).
\]

Finally selecting \(S^{(0)} = \log(\gamma/(1 - \exp(-S^{(1)})))\), the agent meets the false alarm rate requirement. In conclusion, we have built a \((\gamma, \beta)\)-compliant agent in such a way that \(\tau\) and \(S^{(1)}\) and hence the pre-change measurement cost do not depend on \(\gamma\). The full proof is found in Appendix 4.A.

The proof of Theorem 6 is constructive and identifies explicit values of \(S^{(0)}, S^{(1)}\), and \(\tau\), leading to a \((\gamma, \beta)\)-compliant agent. These values are summarized in the next theorem, where we also establish that the resulting agent has minimal post-change measurement cost.

**Theorem 7.** Select \((\tau, S^{(0)}, S^{(1)})\) satisfying:

\[
\tau = \beta \frac{(1 - \exp(-I^2/2\sigma_Z^2))}{\exp(S^{(1)} I/\sigma_Z^2)} \quad \text{and} \quad S^{(0)} = \log \left( \frac{\gamma}{\tau(1 - \exp(-S^{(1)}))} \right) \tag{4.5}
\]

The crisis mode agent defined by \((\tau, S^{(0)}, S^{(1)})\) is \((\gamma, \beta)\)-compliant, and in addition, as \(\gamma \to \infty\),

\[
\mathbb{E}[N] \leq (D(f_1||f_0)^{-1} + o(1)) \log \left( \frac{\gamma}{\beta} \right).
\]
4.5 Numerical experiments

Methodology

Periodic and crisis mode agents. We evaluate both periodic and crisis mode agents, as defined by (4.4) and (4.5) respectively. Note that for crisis mode agents, many choices of pairs \((\tau, S^{(1)})\) are possible – we will precise the choice made later.

Tuning the agents’ parameters according to (4.4) and (4.5) ensures their \((\gamma, \beta)\)-compliance. However, we do not know whether this leads to the best possible \((\gamma, \beta)\)-compliant agents in terms of pre-change measurement cost. For completeness, we will also empirically search for the \((\gamma, \beta)\)-compliant periodic and crisis mode agents with minimal pre-change measurement cost – refer below for a description of this search procedure. These best agents will provide pre-change measurement cost lower bounds.

Performance metrics. For each agent and setting, we evaluate the pre-change measurement frequency. To this aim, we let the agents run in a scenario without distribution change until they raise a false alarm and estimate the average measurement frequency by \(E_\infty[N/t_N] \approx 1/E_\infty[\hat{\tau}]\). We also, simultaneously, estimate the average run time to false alarm \(E_\infty[t_N]\). Furthermore, for each setting, we use a grid of change points \(\nu\) and for each change point, evaluate the delay from the change point to the alarm several times. Taking the maximum of these estimates generates an estimate of \(E[t_N]\). However, for agents driven only by the CUSUM statistics initialized with \(S_0 = 0\), the above supremum is achieved when \(\nu = 0\), since the agent will never be further away from stopping than when \(S_n = 0\) and since the agent decides on the measurement period \(\tau_1\) using non-increasing functions of \(S^{(0)}\). Using this change point as part of our grid allows us to effectively obtain an unbiased estimate. Finally, we vary the change point \(\nu\) and, averaging over multiple runs, evaluate the total measurement volume. This allows us to evaluate the total cost savings of the crisis mode agent compared to the periodic agent, as well as to investigate for which change points the post-change measurement costs dominate the pre-change measurement costs and vice versa.

To illustrate the differences between the periodic and crisis mode agents, we also plot the outcome and physical time of each measurement for sample episodes. We couple these figures with the evolution of the CUSUM statistic in the same episodes, to show how its evolution affects or doesn’t affect the measurement density.

Empirical pre-change measurement cost lower bounds. As mentioned above we try to search for the best compliant periodic and crisis mode agents. This search leads to empirical lower bounds on the pre-change measurement cost. The lower bound for the periodic agent is extracted as follows. For each studied \((\gamma, \beta)\) pair, we study a grid of values for \((\tau, S^{(1)})\). On this grid, we evaluate both delay and false alarm rate for each grid point. On this grid, we find the smallest measurement period \(\tau\) where in 95% of cases both the delay and the false alarm
rate violate the threshold - only by observing both violations can we be sure that a \((\gamma, \beta)\)-compliant agent cannot be found with this period. Alternatively, for low values of \(\gamma/\beta\) we find values for \(\tau\) where the delay guarantee is violated for all values of \(S^{(0)}\).

Obtaining a lower bound for the crisis mode agent is significantly more involved. First, we create a grid of measurement thresholds \(S^{(1)}\). Then, for each value of \(S^{(1)}\), we obtain the greatest frequency at which a lower bound can be established by using the process described for the periodic agent. Finally, we (assuming convexity), choose the smallest such frequency to be such a lower bound - any frequency greater than this smallest frequency cannot, in general, be confidently considered a lower bound for the pre-change measurement frequency.

**Synthetic data sets - Normal distributions**

To evaluate the performance of agents using synthetic data, we chose two sub-gaussian distributions that are close enough so that detecting the change point is not trivial: \(f_0 = \mathcal{N}(0, 1)\) and \(f_1 = \mathcal{N}(0.5, 1)\). We used Gaussian distributions for they simplify the computations of the KL-divergence, the log-likelihood values for different measurement outcomes, as well as the statistical properties of \(Z\). Indeed, we have \(\sigma_Z = (\mu_1 - \mu_0)^2\) for Gaussian distributions with means \(\mu_0\) and \(\mu_1\), and variance 1. Introducing \(\tau_{\text{min}}\) as the smallest allowed period between two measurements, we set \(\tau_{\text{min}} = 0\) to match the framework of Section 4.3. Many values for \((S^{(1)}, \tau)\) are possible to allow \((\gamma, \beta)\)-compliance, we chose \(S^{(1)} = 1\) heuristically by comparing the pre-change frequency for several values of \(S^{(1)}\) and noting that it could not be significantly improved beyond that of \(S^{(1)} = 1\). We set \(\tau\) and \(S^{(0)}\) according to equations (4.4) and (4.5). The full list of setting and agent parameters is given in Table 4.1 in Appendix 4.B. Since this setting is merely aimed at illustrating the nature of the problem and how periodic as well as crisis mode agents behave in it, we use natural units for each value. As such, we set \(\beta = 1\) without loss of generality.

The performance metrics, pre-change measurement frequencies, delays and average run times to false alarm, are shown in Figure 4.3. For the compliant periodic and crisis mode agents, we include (small) 95% confidence intervals. The lower bounds of each agent type do not include such intervals, as they themselves are lower bounds with 95% confidence. In Figure 4.4 we present the average total number (combined pre-change and post-change) measurements for each episode as a function of the change point, also including 95% confidence intervals. However, these confidence intervals are small enough to be nearly invisible. We examine these volumes both at a large scale in Figure 4.4a and at a small scale in Figure 4.4b. Finally, we show the behavior of both agents on a sample episode for each in Figure 4.5. The measurement times and outcomes are plotted along with the evolution of the CUSUM statistic moving towards the thresholds. For the crisis mode agents, the measurements are separated according to whether the agent is
4.5. NUMERICAL EXPERIMENTS

Figure 4.3: Properties of crisis mode agents and periodic agents with analytical guarantees, compared with lower frequency bounds. Properties of frequency bounds included to show tightness or looseness of bounds.

in calm mode or crisis mode once the measurement has been taken. The change point for this evaluation is $\nu = 5/\beta = 5$ in both cases.

**Sensitivity analysis.** To observe to which extent the results above depend on the exact distributions and to which extent they are robust, we perform the following sensitivity analysis. Maintaining a similar setting as above, we fixed $\gamma = 500$ and instead changed the post-change distribution, maintaining Gaussian distributions with variance 1 but varying the mean in the interval $[0.2, 2.0]$. We then repeated the experiments to obtain pre-change frequency, maximum average delay and average run time to false alarm as in Figure 4.3. The results of this analysis is presented in Figure 4.6. Note that in the pre-change frequency results, we now set a logarithmic scale on the y-axis in order to more clearly see the differences between the two curves for all values of the post-change mean.

**Robustness analysis.** Finally, we examined whether the results depend on the assumption of i.i.d. measurements or if this assumption can be relaxed under certain conditions. To do so without changing too many details of the original
setting, we construct an experiment with similar but unknown and non-stationary pre-change distributions. Maintaining the same setting and agents as above, we let the pre-change distribution vary, switching between different gaussian distributions (with exponentially distributed switch times) with means taken uniformly at random from the interval $[-1, 0]$. As such, the agent does not have access to the true mean of the pre-change distribution, but knows its shape and the upper limit of its mean. We varied $\gamma$ as above and repeated the experiments for the guaranteed periodic and crisis mode experiments, obtaining pre-change measurement frequency, maximum average delay and average run time to false alarm. The results of this analysis is presented in Figure 4.7.

5G measurements

Next, we evaluate the performance of the crisis mode agents using data traces collected from an in-house 5G-mmWave testbed.

**Testbed description:** The hardware of the testbed corresponds to a 5G NSA system, where the control-plane is served through a 4G LTE eNB, and the user-plane is served through a 5G gNB as illustrated in Figure 4.8. The eNB operates on band B3, 1800 MHz, with 5 MHz bandwidth whereas the gNB operates on band n257, 28 GHz, with 100 MHz bandwidth. The spectrum is time-shared between downlink and uplink using a 4:1 TDD pattern [84], and the antenna operates at 2W.

In the testbed, there are two user equipment (UEs), one for end-to-end performance measurements and one for generating traffic load scenarios and thus change points. The UEs correspond to Ubuntu 20.04 laptops connected via USB-C to 5G modems. Both UEs are positioned in line of sight approximately 5m from the 5G antenna.
4.5. NUMERICAL EXPERIMENTS

Figure 4.5: Behavior of both agents, with change point $\nu = 5\beta$.

Data-plane network performance in this testbed is measured in terms of RTT between the measurement controller and reflector every 100ms using 1400 byte ICMP packets [79].

**Measurement scenarios.** Change points were created by introducing uplink traffic into the network using the second laptop instrumented with Iperf [85]. By introducing higher traffic load at different points, we were able to capture change point behavior for multiple known values of $\nu$, which is necessary to evaluate maximum average delay.

First, we continuously measured (with 100 ms intervals) the RTT values of the ICMP ping for 600 s with an induced load of 10 Mbps, which corresponds to the pre-change distribution $f_0$. Then, we produced 5 experiments where the load was changed from 10 Mbps to 50 Mbps at different change points. By concatenating all of the high load data, we are able to estimate $f_1$. We also tested if there is a transition period when introducing the load and, if so, if this interferes with the agents’ performance.

**Data processing and evaluation.** We estimated both $f_0$ and $f_1$ by using Gaussian kernel density estimates (KDE). A histogram view of both data distri-
Figure 4.6: Properties of crisis mode agents and periodic agents with analytical guarantees for a number of different post-change distributions. The smaller the difference in means, the more difficult it is to identify a change in distribution while maintaining a constant false alarm rate.

\[\text{Average run time to false alarm}\]

\[\text{Maximum average delays}\]

\[\text{Pre-change measurement frequencies}\]
4.5. NUMERICAL EXPERIMENTS 67

(a) Pre-change measurement frequencies
(b) Maximum average delays

(c) Average run time to false alarm

Figure 4.7: A repetition of the experiments of Figure 4.3, using unknown and non-i.i.d. but limited pre-change distributions. Only the guaranteed periodic and crisis mode agents have been considered.

Figure 4.8: An illustration of the 5G testbed setup used for this work. Measurements of RTT is conducted between the measurement controller and the measurement reflector over the 5G wireless channel. Traffic is induced by the Iperf source towards the Loader sink.
to see changes in frequency properties, particularly when it came to the lower bounds, as measurement periods can now only take discrete values. As such, we set a relatively large value for $\beta$, namely $\beta = 70s$. To draw parallals to the initial evaluation, we let the ratio $\gamma/\beta$ take the same values as in that evaluation. We use $S^{(1)} = 0.2$ as it seems to give the smallest pre-change frequency in this case, and set $\tau$ and $S^{(0)}$ according to equations (4.4) and (4.5). The full list of parameters is given in Table 4.2 in Appendix 4.B.

Since the agent may fail to detect a change before the end of the data, we whenever this is the case bootstrap new pre-change or post-change data whenever necessary. When doing so, we exclude the first 2 $s$ of data directly after each change point as this was observed to be the window when transient behavior, if any, would be apparent.

Armed with distribution estimates and the necessary parameters to use for the periodic and crisis mode agents, we evaluate all agents as in Section 4.5, and the obtained properties are shown in Figure 4.10. In Figure 4.11 we present the average total number of measurements for each episode as a function of the change point. Finally, we show the behavior of both agents on a sample episode for each in Figure 4.12. The change point here is set to $\nu = 5\beta = 350$ $s$ in both cases.

Discussion

In Figures 4.3 and 4.10, we can see that the average pre-change frequency of the guaranteed crisis mode agent remains more or less constant as the ratio $\gamma/\beta$ increases, albeit with some higher variance in the case with 5G measurements, which is consistent with the claim of Theorem 6. By contrast, the frequency of both the guaranteed periodic agent and the periodic agent’s lower bound increases by an order of $\log \left( \frac{\gamma}{\beta} \right)$, also consistent with Proposition 5 and its proof. The lower
4.5. NUMERICAL EXPERIMENTS

Figure 4.10: Properties of crisis mode agents and periodic agents with analytical guarantees, compared with lower frequency bounds. Properties of frequency bounds included to show tightness or looseness of bounds.

The average pre-change frequency and maximum average delay for crisis mode agents and periodic agents are compared with their lower frequency bounds. The guaranteed crisis mode agent is shown to be in no way tight or optimal, as evidenced by the large frequency gap between it and its lower bound, as well as the large maximum average delay gap between it and the delay upper bound $\beta$.

An interesting detail is that the “guaranteed” periodic agent performs worse as the strictness increases in the case with 5G testbed measurements. Indeed, despite supposedly being guaranteed to be $(\gamma, \beta)$-compliant it fails the false alarm condition.
We examine cost both at a large scale and at a small scale. As expected, the measurement cost as detailed in Figures 4.4 and 4.11 is essentially an affine function of the change point \( \nu \) in all cases, since we expect the pre-change measurement cost to be linear in \( \nu \) and the post-measurement cost to be a constant value. It is especially interesting that even at very low values of \( \gamma \), the post-change measurement cost is completely dominated by the pre-change measurement cost, as shown by Figures 4.4b and 4.11b. We thus propose that when considering network performance verification, one should focus on pre-change measurements rather than post-change measurements. The improvement in measurement cost is larger in the case with Gaussian distributions rather than 5G measurements, which we attribute to the Gaussian distributions having a more well-behaved log-probability ratio \( Z \).

The sensitivity analysis depicted in Figure 4.6 shows that the relative improvement in pre-change measurement frequency of using the crisis mode agent over periodic agents stays more or less the same regardless of the post-change distribution, given a fixed value of \( \gamma \). Furthermore, the false alarm rate is maintained at essentially the same value regardless. It is, however, curious to note that the maximum average delay increases as the difference between the two distributions increases, initially counter-intuitive but clearly a result of how the crisis mode agent is defined. This implies that when the difference between distributions is large, our definition of the crisis mode agent could be near-optimal in the subspace of all possible crisis mode agents. Since it shows no major changes compared to Figure 4.3, the robustness analysis depicted in Figure 4.7 implies that both the periodic and the crisis mode agents are robust to minor modeling errors, as long as both the shape of the pre-change distribution as well as the...
worst case distribution are known. Indeed, the crisis mode agent seems able to take advantage of the pre-change distribution being further from the post-change distribution than in the nominal case and measure less frequently.

Figures 4.5 and 4.12 show the behavior of the agents on the two environments, and they behave like we would expect. Compared to the periodic agent, the crisis mode agent sometimes has very high measurement density (even before the change) but as a payoff, it is able to very quickly identify the change despite a lower overall measurement frequency.

While there are some complications (i.e. difficulty in estimating post-change distributions and relative parameters), the implications of these experiments are highly interesting. By using relatively simple probes it is possible to detect changes in network dynamics, and the crisis mode agents are able to cut down on measurement costs as much as 25% in a realistic scenario. Our agent admits arbitrary constraints on maximum average delay and false alarm rate, and it has been shown to be far from tight. We thus believe that intelligent measurement strategies such as this one could significantly reduce network operation costs when developed further. The proposed crisis mode agents already lead to

Figure 4.12: Behavior of both agents, with change point $\nu = 5\beta$. 
4.6 Conclusion

In this work, we have presented a new framework for change point detection to account for measurement costs and offering the opportunity to the agent to schedule her measurements over time as she wishes and in an adaptive manner. This framework was motivated in the context of network performance verification. We have provided a first set of important results for our framework. First we derived lower bound on the post-change measurement cost subject to some constraints on the delay and the average run time to false alarm. Then we have developed simple agents with adaptive schedules that provide a provably better trade-off between pre-change measurement cost, delay and the average run time to false alarm than classical agents using periodic measurements. Finally we have assessed numerically the performance of these agents using both synthetic data and data from a 5G testbed. These experiments illustrate the cost reductions one can achieve in practice by using change point detection with adaptive schedules.

There are many interesting open questions related to our framework of change point detection with optimized cost. Indeed we have not been able so far to really characterize $(\gamma, \beta)$-compliant agents with minimal pre-change measurement cost. Another interesting research direction is to extend our results to scenarios where the post-change distribution is initially unknown: we could analyze in this case how to tune the adaptive measurement scheduling rule as time passes (as more and more data is available to the agent).

4.A Proofs

Proof of Theorem 5

Let $I = D(f_1 || f_0)$ for short, and introduce for measurement $X_k$ the log-likelihood ratio $Z_k = \log \left( \frac{f_1(x_k)}{f_0(x_k)} \right)$. The proof will use the following steps:

1. First, since $N$ and $n_\nu$ are both stopping times we have that $\{ N > n_\nu \} \in \mathcal{G}_{n_\nu}$. As such, for any $\nu$ such that $P_\nu(N > n_\nu) > 0$, it follows that

$$\text{ess sup}_\nu E_\nu[(N - n_\nu)^+ | \mathcal{G}_{n_\nu}] \geq E_\nu[N - n_\nu | N > n_\nu].$$

In other words: we only need to consider events such that $N > n_\nu$ for our bound.

2. For any $\delta \in (0, 1)$, we define our main concentration event $\mathcal{E}_{\delta, \phi}(n) = \{ \sum_{k=n_\nu+1}^{n_\nu+n} Z_k > \phi(1 + \delta)I \}$, depending on the number of measurements $n$. By classical results,

$$\lim_{\phi \to \infty} P_\infty(\bigcup_{n=1}^{\lfloor \phi \rfloor} \mathcal{E}_{\delta, \phi}(n)) = 0.$$
3. Next, we introduce a similar event $\mathcal{E}_{\delta, \phi}(N - n_\nu)$ for the stopping time $N$ and show that the above implies that the probability of both $\mathcal{E}_{\delta, \phi}(N - n_\nu)$ and $\{N < n_\nu + \phi\}$ occurring simultaneously approaches 0 as $\phi \to \infty$.

4. Next, we will use Markov’s inequality show that to fulfill the requirement $\mathbb{E}[t_N] \leq \beta$ any agent must fulfill $\mathbb{P}_\nu(t_N > \nu + \alpha | N > n_\nu) \leq \frac{\beta}{\alpha}$ for all values of $\nu$ with $\mathbb{P}_\infty(N > n_\nu) > 0$. In other words: To stop soon enough on average, the probability of stopping late must be sufficiently small.

5. Then, we will choose $\alpha$ and $\nu$ cleverly and show that as $\gamma \to \infty$, no agent can fulfill all three of $\mathbb{E}[t_N] \geq \gamma$, $t_N < \nu + \alpha$ and $\mathcal{E}_{\delta, \phi}(N - n_\nu)$ with non-zero probability. In other words: If you stop very soon with a low degree of confidence in a change occurring, you will get a high false alarm rate and therefore will not be compliant.

6. Finally, we show that with a union bound that $\mathbb{P}_\nu(N - n_\nu > \phi | N > n_\nu) \to 1$ as $\gamma \to \infty$ keeping $\beta$ fixed, which in turn lower bounds the expectation $\mathbb{E}_\nu[N - n_\nu | N > n_\nu]$ and gives us our desired bound.

**Proof of step 1:** $N$ is a stopping time by definition, and because $\tau_{n+1}$ is $\mathcal{G}_n$-measurable, it follows that $\{t_{n+1} > \nu\} \in \mathcal{G}_n$, and subsequently $\{n_\nu \leq n\} \in \mathcal{G}_n$. Therefore $n_\nu$ is a stopping time adapted to the filtration $(\mathcal{G}_k)_{k \geq 0}$. It follows, then that $\{N > n_\nu\} \in \mathcal{G}_{n_\nu}$ and therefore, if $\mathbb{P}_\nu(N > n_\nu) > 0$ we obtain $\text{ess sup}_\nu \mathbb{E}_\nu[(N - n_\nu)^+ | \mathcal{G}_{n_\nu}] \geq \mathbb{E}_\nu[N - n_\nu | N > n_\nu]$. We will therefore lower bound the latter expression in the sequel, which will lower bound the former - indeed, we will show that there exists $\nu$ with $\mathbb{P}_\infty(N > n_\nu) > 0$ such that the bound is realized.

**Proof of step 2:** This result is classical and has been shown at several points before, as referenced in [66].

**Proof of step 3:** We introduce the event $\mathcal{E}_{\delta, \phi}(N - n_\nu) = \left\{ \sum_{k=n+1}^{N} Z_k > I(1+\delta)\phi \right\}$ and note that $\mathbb{P}_\nu(N < n_\nu + \phi, \mathcal{E}_{\delta, \phi}(N - n_\nu)) \leq \mathbb{P}_{\nu} \left( \bigcup_{n=1}^{\phi} \mathcal{E}_{\delta, \phi}(n) \right) \to 0$ as $\phi \to \infty$, by step 2. As such, as $\phi$ grows, the probability of maintaining both $N < n_\nu + \phi$ and $\mathcal{E}_{\delta, \phi}(N - n_\nu)$ shrinks to 0.

**Proof of step 4:** By Markov’s inequality, it must be true for any $\alpha > 0$ that for any $\nu$ with $\mathbb{P}_\nu(N > n_\nu) > 0$,

$$
\mathbb{P}_\nu(t_N \geq \nu + \alpha | N > n_\nu) = \mathbb{P}_\nu((t_N - \nu)^+ \geq \alpha | N > n_\nu) \\
\leq \frac{\mathbb{E}_\nu[(t_N - \nu)^+ | N > n_\nu]}{\alpha} \\
\leq \frac{\mathbb{E}[t_N]}{\alpha} \\
\leq \frac{\beta}{\alpha}
$$
if the agent is $(\gamma, \beta)$-compliant, which completes this step.

**Proof of step 5:** First, we can show as in the proof of Theorem 1 in [66] that if an agent is $(\gamma, \beta)$-compliant, for any $m > 0$ there exists $\nu \geq 0$ such that $\mathbb{P}_\nu(N > n_\nu) > 0$ and $\mathbb{P}_\infty(t_N < \nu + m|N > n_\nu) \leq m/\gamma$, otherwise the agent cannot fulfill the condition $\mathbb{E}_\infty[t_N] \geq \gamma$. Thus, we select $m = (\log \gamma/\beta)^2$. Define $C_\delta = \{0 \leq t_N - \nu < \beta \phi\} \cap \neg \mathcal{E}_{\delta,\phi}(N\!-\!n_\nu)$ for $\phi = (1 - \delta) \log(\gamma/\beta)I^{-1}$. Importantly, $\phi$ is selected as a function of $\gamma$ so that $\lim_{\gamma \to \infty} \phi = \infty$ when keeping $\beta$ fixed.

On $C_\delta$ it holds that $\sum_{k=n_\nu+1}^{N} Z_k \leq \phi(1 + \delta)I = (1 - \delta^2) \log(\gamma/\beta)$, so we can see that

$$
\mathbb{P}_\nu(C_\delta) = \int_{C_\delta} \exp \left( \sum_{k=n_\nu+1}^{N} Z_k \right) d\mathbb{P}_\infty 
\leq \exp((1 - \delta^2) \log(\gamma/\beta))\mathbb{P}_\infty(C_\delta).
$$

Since $\mathbb{P}_\nu(N > n_\nu) = \mathbb{P}_\infty(N > n_\nu)$ it follows that we can select $\nu$ such that for all sufficiently large $\gamma$, keeping $\beta$ fixed,

$$
\mathbb{P}_\nu(C_\delta|N > n_\nu) \leq \left( \frac{\gamma}{\beta} \right)^{1-\delta^2} \mathbb{P}_\infty(t_N - \nu < \beta \phi|N > n_\nu) 
\leq \left( \frac{\gamma}{\beta} \right)^{1-\delta^2} \mathbb{P}_\infty(t_N < \nu + (\log(\gamma/\beta))^2|N > n_\nu) 
\leq \left( \frac{\gamma}{\beta} \right)^{-\delta^2} (\log(\gamma/\beta))^2 \to 0
$$

as $\gamma \to \infty$. Ergo, we can choose $\nu$ (depending on $\gamma$) such that $\mathbb{P}_\infty(t_N > \nu) > 0$ and $\mathbb{P}_\nu(C_\delta|N > n_\nu) = o(1)$, and we will use this value of $\nu$ in the sequel. This concludes the proof of step 5.

**Proof of step 6:** Using $\phi$ and $\nu$ of step 5, we set $\alpha = \beta \phi$ in step 4. By union bound, it follows that

$$
\mathbb{P}_\nu(\neg \mathcal{E}_{\delta,\phi}(N - n_\nu)) \leq \mathbb{P}_\nu(C_\delta|N > n_\nu) + \mathbb{P}_\nu(t_N \geq \nu + \beta \phi|N > \nu) 
\leq o(1) + \frac{\beta}{\beta \phi} = o(1) + \frac{1}{\phi},
$$

which tends to 0 as $\phi \to \infty$. Another union bound gives us that

$$
\mathbb{P}_\nu(N < n_\nu + \phi|N > n_\nu) \leq \mathbb{P}_\nu(N < n_\nu + \phi, \mathcal{E}_{\delta,\phi}(N - n_\nu)|N > n_\nu) + \mathbb{P}_\nu(\neg \mathcal{E}_{\delta,\phi}(N - n_\nu)) = o(1)
$$

Finally, we can then see that since $\mathbb{P}_\nu(N - n_\nu \geq \phi|N > n_\nu) \to 1$ as $\gamma \to \infty$, we deduce that $\mathbb{E}_\nu[N - n_\nu|N > n_\nu] \geq \phi(1 + o(1)) = ((1 - \delta)I^{-1} + o(1)) \log(\gamma/\beta)$. Since this is true for any $\delta > 0$, Theorem 5 follows. \(\square\)
4.A. PROOFS

Proof of Proposition 5

Statement (i) follows trivially when noting that for $\beta < \tau$ it will always be true that either $\mathbb{E}[t_N] > \beta$ or $\mathbb{E}_\infty[t_N] < \beta$. For statement (ii) we note, for periodic agents with period $\tau$, by Lorden \[63\] that

$$\mathbb{E}_\infty[t_N] \geq \gamma \implies \mathbb{E}[N] \geq (I^{-1} + o(1)) \log \left( \frac{\gamma}{\tau} \right)$$

when $\gamma \to \infty$. As such, if for $\tau$ it holds that $\theta(\gamma) := I^{-1} \log \left( \frac{\gamma}{\tau} \right) > \beta$ then there exists some $\bar{\gamma}(\tau)$ such that for all $\gamma' > \bar{\gamma}(\tau)$ it holds that if $\mathbb{E}_\infty[t_N] > \gamma'$ then $\mathbb{E}[t_N] > \beta$. Thus, this $\bar{\gamma}(\tau)$ would therefore fulfill the condition of the first statement of Proposition 5.

As such, for any $(\gamma, \beta)$-compliant periodic agent, we require

$$\lim_{\gamma \to \infty} \theta(\gamma) \leq \beta.$$ 

However, it is simple to see that $\theta(\gamma)$ is a strictly increasing function of $\gamma$ and that $\lim_{\gamma \to \infty} \theta(\gamma) = \infty$, regardless of $\tau$ and $f_0$, $f_1$ (with $f_0 \neq f_1$). It immediately follows that no periodic agent with period $\tau$ can be $(\gamma, \beta)$-compliant for such $\gamma$ and $\beta$, which proves the statement (ii).

Proof of Proposition 6

We begin by showing that the equation system has exactly two solutions when $\gamma > \beta$. Note that the existence of a solution to the system is equivalent to the existence of a root to the continuous function $f(\tau) := \tau (\log(\gamma) + \xi) - \beta I$. We note that $f(0) < 0$, $\lim_{\tau \to \infty} f(\tau) = -\infty$ and that the function has the only extremal point $\tau_{\text{max}} = \gamma \exp(\xi - 1)$ where it takes the value $f(\tau_{\text{max}}) = \gamma \exp(\xi - 1) - \beta I$. Note that $\xi = \sup_{\tau \geq 0} \mathbb{E}_{X \sim f_1}[Z - r | z \geq r] \geq \mathbb{E}_{X \sim f_1}[Z | Z \geq 0] \geq \mathbb{E}_{X \sim f_1}[Z] = I$, and note further that by elementary calculus $\exp(\xi - 1) \geq \xi \geq I$ uniformly in $\xi$. Therefore, since $\gamma > \beta$ it follows that $f(\tau_{\text{max}}) > 0$ and since this is the only extremal point, it is a maximal point and the equation system has exactly two solutions.

Next, note that once the change has occurred, the nature of the problem changes to that of a one-sided sequential hypothesis test where the hypothesis $H_1$ to be accepted is that $X \sim f_1^4$. Then, we see that the periodic agent will stop as least as fast as a sequential probability ratio test with $A = \exp(S^{(0)})$ and $B = -\infty$. It is then well known, e.g. by Wald \[86\] that the expected number of measurements for this test is upper bounded by $I^{-1}(\log(A) + \xi)$ or in other words $I^{-1}(S^{(0)} + \xi)$, given that $H_1$ is true. Now, we set $S^{(0)} = \log \left( \frac{\gamma}{\tau} \right)$ and recall

---

4In actuality, the agent will often stop faster than the case of one-sided hypothesis test, as the CUSUM statistic is reset to 0 whenever it would go negative, but this does not impact our proof.
that \( \tau = \frac{\beta I}{S(0)+\xi} \). Then, we find that, for any \( \nu \)

\[
\text{ess sup } \mathbb{E}_\nu[(N - n_\nu)^+|\mathcal{G}_{n_\nu}] 
\leq I^{-1} \log \left( \frac{\gamma(S(0) + \xi)}{\beta I} \right) + I^{-1} \xi 
= I^{-1} \log \left( \frac{\gamma}{\beta I} \right) + I^{-1} o \left( \log \left( \frac{\gamma}{\beta} \right) \right) + I^{-1} \xi 
= (I^{-1} + o(1)) \log(\gamma/\beta)
\]

as \( \gamma \to \infty \) when \( \beta \) remains fixed since neither \( I \) nor \( \xi \) depend on \( \gamma \) or \( \beta \). The first equality comes from the fact that \( \lim_{\gamma \to \infty} \frac{S(\gamma)}{\gamma} = 0 \), which in turn follows from the fact that \( \tau \) does not decrease faster than \( 1/\log(\gamma) \).

Next, let us prove that the agent is \((\gamma, \beta)\)-compliant. Similarly to the proof of Theorem 6, we can deduce that \( \mathbb{E}_\infty[N] \geq \exp(S(0)) \). But by definition, \( S(0) = \log(\frac{\gamma}{\tau}) \). Thus, we get

\[
\mathbb{E}_\infty[t_N] = \tau \mathbb{E}_\infty[N] \geq \tau \exp(S(0)) = \tau \frac{\gamma}{\tau} = \gamma
\]

and thus, the agent fulfills the false alarm rate condition. Next, as we saw above, \( \text{ess sup}_\nu [(N - n_\nu)^+|\mathcal{G}_{n_\nu}] \leq I^{-1}(S(0) + \xi) \) for any \( \nu \), so we obtain

\[
\text{ess sup}_\nu [(t_N - \nu)^+|\mathcal{G}_{n_\nu}] \leq \tau \left( \text{ess sup}_\nu [(N - n_\nu)^+|\mathcal{G}_{n_\nu}] \right)
\leq \tau I^{-1}(S(0) + \xi)
= \frac{\beta I}{S(0) + \xi} I^{-1}(S(0) + \xi)
= \beta
\]

where the first equality follows by the definition of \( \tau \) and \( S(0) \). Thus, the agent also fulfills the delay condition. This concludes the proof. \( \square \)

**Proof of Lemma 2**

Statement (i) follows immediately by definition of \( N \) and \( n_\nu \). It remains to prove statement (ii).

Since \( N = \min\{n \geq 1 : S_n \geq S(0)\} \), we find that \( \mathbb{P}_\infty(\tau_n|N > n - 1) = \mathbb{P}_\infty(g(S_{n-1})|N > n - 1) = \mathbb{P}_\infty(g(S_{n-1})|S_{n-1} < S(0)) \). But it is easy to see that for any \( a > 0 \), \( \mathbb{P}_\infty(S_n > a|S_n \leq S(0)) \leq \mathbb{P}_\infty(S_n > a) \) and since \( g \) is a non-increasing function it follows that \( \mathbb{E}_\infty[\tau_n|N > n - 1] = \mathbb{E}_\infty[g(S_{n-1})|S_{n-1} \leq S(0)] \geq \mathbb{E}_\infty[g(S_{n-1})] = \mathbb{E}_\infty[\tau_n] \) by properties of stochastic dominance. It remains to prove that \( \mathbb{E}_\infty[\tau_n] \geq \mathbb{E}_\infty[\hat{\tau}] \). We will do this by showing that \( \mathbb{E}_\infty[\tau_i] \) is a non-increasing sequence in \( i \), because then it follows that

\[
\mathbb{E}_\infty[\hat{\tau}] = \lim_{m \to \infty} \mathbb{E}_\infty \left[ \sum_{k=1}^{m} \frac{T_k}{m} \right] = \lim_{m \to \infty} \mathbb{E}_\infty \left[ \sum_{k=1}^{m} \frac{T_k}{m} \right] = \lim_{m \to \infty} \mathbb{E}_\infty[\tau_m] \leq \mathbb{E}_\infty[\tau_n]
\]
Denote by \( R_{i+1} = \max_{2 \leq j \leq i+1} \sum_{k=j}^{i+1} Z_k \). Then \( R_{i+1} \) is a random variable, and because \( \{Z_k\}_{k=1}^{\infty} \) are all i.i.d. under \( f_0 \) it follows that \( R_{i+1} \) is identically distributed to \( S_i \) under \( \mathbb{P}_\infty \). Note now that \( S_{i+1} = \max(R_{i+1}, \sum_{k=1}^{i+1} Z_k) \) which in turn implies \( S_{i+1} \geq R_{i+1} \). As such, \( S_{i+1} \) stochastically dominates \( S_i \) under \( \mathbb{P}_\infty \), and since \( g \) is a non-increasing function, it follows that \( \mathbb{E}_\infty[\tau_i] = \mathbb{E}_\infty[g(S_{i-1})] \geq \mathbb{E}_\infty[g(S_i)] = \mathbb{E}_\infty[\tau_{i+1}] \). Ergo, \( \mathbb{E}_\infty[\tau_i] \) is a non-increasing sequence in \( i \), which concludes the proof.

**Proof of Theorem [6]**

We will proceed with the proof by construction. First, we will prove that under \( f_1 \),

\[
\sum_{n=1}^{\infty} \mathbb{P}_{X \sim f_1}(S_n \leq S^{(1)}) \leq \frac{\exp(S^{(1)} I / \sigma_Z^2)}{1 - \exp(-I^2 / 2 \sigma_Z^2)}
\]

where \( \sigma_Z \) is the subgaussianity property of \( Z \) under \( f_1 \), existing by assumption. It follows that for any \( \nu \geq 0 \), \( \sum_{n=n_\nu+1}^{\infty} \mathbb{P}_{\nu}(S_n \leq S^{(1)} | N > n_\nu) \) is upper bounded by the same expression and therefore, so is \( \mathbb{E}[t_N] / \tau \). Using a Chernoff bound, we find that

\[
\mathbb{P}_{X \sim f_1}(S_n \leq S^{(1)}) = \mathbb{P}_{X \sim f_1} \left( \max_{1 \leq i \leq n} \sum_{j=i}^{n} Z_j \leq S^{(1)} \right)
\]

\[
\leq \mathbb{P}_{X \sim f_1} \left( \sum_{j=1}^{n} Z_j \leq S^{(1)} \right)
\]

\[
\leq \exp(-(S^{(1)} - nI)^2 / 2n \sigma_Z^2)
\]

\[
= \exp(-S^{(1)}^2 / 2n) \exp(S^{(1)} I / \sigma_Z^2) \exp(-nI^2 / 2 \sigma_Z^2)
\]

\[
\leq \exp(S^{(1)} I / \sigma_Z^2) \exp(-nI^2 / 2 \sigma_Z^2)
\]

Summing this from 1 to \( \infty \) we get the above inequality by a geometric sum.

Now

\[
\mathbb{E}_\nu[t_N - \nu | t_N \geq \nu] = \sum_{n=n_\nu+1}^{N} \mathbb{E}_\nu[\tau_n | N > n_\nu]
\]

\[
\leq \tau \sum_{n=n_\nu+1}^{\infty} \mathbb{P}_{\nu}(S_n \leq S^{(1)} | N > n_\nu),
\]

so the above result bounds \( \mathbb{E}_\nu[t_N - \nu | N > n_\nu] \) independent on \( S^{(0)} \), depending only on \( S^{(1)} \) and \( \tau \).

The next step, then, is to prove that with such choices of \( S^{(1)} \) and \( \tau \), it is possible to also fulfill the false alarm rate condition. Define \( \sigma_0 = 0 \), \( \sigma_1 = \)
min\{n \geq 1 : \sum_{k=1}^{n} Z_k \leq 0\}, and \(\sigma_{m+1} = \min\{n > \sigma_m : \sum_{k=\sigma_m+1}^{n} Z_k < 0\}\) as the points where the sum of likelihood ratios crosses the x-axis. Then, as in [66], defining \(M = \inf\{m : \exists \sigma_m < n < \sigma_{m+1} : \sum_{k=\sigma_m+1}^{n} Z_k > S^{(0)}\}\), we can show that \(E_{\infty}[M] \geq \exp(S^{(0)})\). Next, recall that \(E_{\infty}[\hat{\tau}] = \liminf_{n \to \infty} E_{\infty}\left[\frac{1}{n} \sum_{k=1}^{n} \tau_k\right]\). By Lemma 2, for each \(n < N\), \(E_{\infty}[\tau_n N > n - 1] \geq E_{\infty}[\hat{\tau}]\). We therefore see that \(E_{\infty}[\tau_n] \geq E_{\infty}[M] E_{\infty}[\hat{\tau}] \geq \exp(S^{(0)}) E_{\infty}[\hat{\tau}]\). To fulfill the false alarm rate guarantee, it remains to lower bound \(E_{\infty}[\tau]\).

Now,

\[
E_{\infty}\left[\frac{1}{n} \sum_{k=1}^{n} \tau_k\right] = \frac{1}{n} \sum_{k=1}^{n} E_{\infty}[\tau_k] = \tau \sum_{k=1}^{n} P_{\infty}(S_k \leq S^{(1)}).
\]

Define \(S'_k = \max_{k \leq n \leq \infty} \sum_{j=k}^{n} Z_j\). Note that \(S'_k \geq \max_{k \leq n \leq 2k} \sum_{j=k}^{n} Z_j\) a.s. Therefore, because \(Z_k\) are all i.i.d under \(P_{\infty}\) it follows that for all \(k\), \(S_k\) is stochastically dominated by \(S'_k\). We can therefore bound \(P_{\infty}(S_k \leq S^{(1)}) \geq P_{\infty}(S'_k \leq S^{(1)})\). Furthermore, \(S'_k\) are all identically distributed (not necessarily independent), so \(P_{\infty}(S'_k \leq S^{(1)}) = P_{\infty}(S'_1 \leq S^{(1)})\). Note also that \(R'_n := \exp(\sum_{i=1}^{n} Z_n)\) is a non-negative martingale when \(X \sim f_0\) and that \(S'_1 \leq S^{(1)} \iff \max_{1 \leq n \leq \infty} R'_n \leq \exp(S^{(1)})\). Therefore,

\[
E_{\infty}\left[\frac{1}{n} \sum_{k=1}^{n} \tau_k\right] \geq \tau P_{\infty}(S'_1 \leq S^{(1)}) = \tau P_{\infty}\left(\max_{1 \leq n \leq \infty} R'_n \leq \exp(S^{(1)})\right) \geq \tau(1 - \exp(-S^{(1)}))
\]

by Doob’s submartingale inequality, uniformly in \(n\). But it follows then that \(E_{\infty}[\hat{\tau}] \geq \tau(1 - \exp(-S^{(1)}))\) and thereby \(E_{\infty}[t_N] \geq \tau(1 - \exp(-S^{(1)})) \exp(S^{(0)})\).

Therefore, choose

\[
S^{(0)} = \log\left(\frac{\gamma}{\tau(1 - \exp(-S^{(1)}))}\right)
\]

and \(E_{\infty}[t_N] \geq \gamma\), i.e. the false alarm rate guarantee is met. Now, \(E_{\infty}[\hat{\tau}]\) is by definition independent of the stopping threshold \(S^{(0)}\) when both \(S^{(1)}\) and \(\tau\) are independent of \(S^{(0)}\). As such, fixing \(S^{(1)}\) and \(\tau\) such that the delay guarantee is met, which can be done by choosing any \(S^{(1)} > 0\) and choosing

\[
\tau = \beta \frac{(1 - \exp(-I^2/2\sigma^2_Z))}{\exp(S^{(1)}I/\sigma^2_Z)},
\]

we see that any agent with these values will have the same value of \(E_{\infty}[\hat{\tau}] =: \tau_c\). Then, the Theorem follows by simply choosing \(S^{(0)}\) as above. \(\square\)
Proof of Theorem 7

The $(\gamma, \beta)$-compliance follows immediately from the proof of Theorem 6. It remains to show that the agent’s post-change measurements are appropriately upper bounded.

The proof logic very closely matches that of the proof of Proposition 6, as by Lemma 2 we do not need to consider the measurement strategy when evaluating the measurement volume. Recall that $S(0) = \log(\frac{\gamma}{\tau(1-\exp(-S(1)))})$ and $
abla = \beta \left(1-\exp(-\frac{I^2}{2\sigma_Z^2})\right)$, and since by the aforementioned proof

$$\text{ess sup} \mathbb{E}_\nu[(N-n_\nu)^+|\mathcal{G}_{n_\nu}] \leq I^{-1}(S(0) + \xi),$$

we obtain

$$\text{ess sup} \mathbb{E}_\nu[(N-n_\nu)^+|\mathcal{G}_{n_\nu}]$$

$$\leq I^{-1} \log \left(\frac{\gamma \exp(S(1)I/\sigma_Z^2)}{\beta(1-\exp(-I^2/2\sigma_Z^2))(1-\exp(-S(1)))}\right) + I^{-1} \xi$$

$$= (I^{-1} + o(1)) \log(\gamma/\beta)$$

as $\gamma \to \infty$ when $\beta$ remains fixed, since neither $I$, $\sigma_Z$, $\xi$ nor $S(1)$ depends on $\gamma$. Since this is true for any $\nu \geq 0$, this concludes the proof.

4.B Parameters for numerical evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_0$</td>
<td>$\mathcal{N}(0,1)$</td>
</tr>
<tr>
<td>$f_1$</td>
<td>$\mathcal{N}(0.5,1)$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>${1, 2, 5, 10, 20, 50, 100, 200, 500, 1000}$</td>
</tr>
<tr>
<td>$\tau_{\min}$</td>
<td>0</td>
</tr>
<tr>
<td>$S(1)$</td>
<td>1</td>
</tr>
<tr>
<td>$\nu$ (for delay evaluation)</td>
<td>$\gamma \ast [0, 0.01, 0.1, 0.2, 0.5, 1]$</td>
</tr>
</tbody>
</table>

Table 4.1: Parameters used in the initial evaluation in Section 4.5.
Table 4.2: Parameters used in the evaluation on 5G testbed data in Section 4.5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_0 )</td>
<td>Gaussian KDE of pre-change data</td>
</tr>
<tr>
<td>( f_1 )</td>
<td>Gaussian KDE of post-change data</td>
</tr>
<tr>
<td>( \beta )</td>
<td>70 s</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>( \beta \in {1, 2, 5, 10, 20, 50, 100, 200, 500, 1000} )</td>
</tr>
<tr>
<td>( \tau_{\text{min}} )</td>
<td>0.1 s</td>
</tr>
<tr>
<td>( S^{(1)} )</td>
<td>0.2</td>
</tr>
<tr>
<td>( \nu ) (for delay evaluation)</td>
<td>[0s, 10s, 20s, 25s, 30s]</td>
</tr>
</tbody>
</table>

4.C Table of notations
### Table 4.3: Table of Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Ordinal number of measurement.</td>
</tr>
<tr>
<td>$X_n$</td>
<td>Measurement number $n$.</td>
</tr>
<tr>
<td>$f_0$</td>
<td>Known distribution of $X_n$ in nominal scenario.</td>
</tr>
<tr>
<td>$f_1$</td>
<td>Known distribution of $X_n$ in crisis event.</td>
</tr>
<tr>
<td>$Z$</td>
<td>Log-likelihood ratio of $f_0$ and $f_1$, $\log \frac{f_1(X)}{f_0(X)}$.</td>
</tr>
<tr>
<td>$D$</td>
<td>Kullback-Leibler divergence between two distributions.</td>
</tr>
<tr>
<td>$I$</td>
<td>Short-hand for $D(f_1</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Change time. Unknown by agent.</td>
</tr>
<tr>
<td>$n_\nu$</td>
<td>Number of measurements taken prior to $\nu$.</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>Intermittent time between measurements $X_{n-1}$ and $X_n$, chosen by agent.</td>
</tr>
<tr>
<td>$t_n$</td>
<td>Physical time where measurement $n$ occurs, $t_n = \sum_{k=1}^{n} \tau_k$.</td>
</tr>
<tr>
<td>$\mathcal{F}_n$</td>
<td>Sigma-field $\sigma(X_1, ..., X_n)$.</td>
</tr>
<tr>
<td>$\mathcal{G}_n$</td>
<td>Sigma-field $\sigma(\tau_1, X_1, ..., \tau_n, X_n)$.</td>
</tr>
<tr>
<td>$\mathbb{P}<em>\nu$ ($\mathbb{E}</em>\nu$)</td>
<td>Probability measure (expectation) when change time is $\nu$.</td>
</tr>
<tr>
<td>$\mathbb{P}<em>\infty$ ($\mathbb{E}</em>\infty$)</td>
<td>Probability measure (expectation) when no change occurs.</td>
</tr>
<tr>
<td>$S_n$</td>
<td>CUSUM statistic to time $n$, $\max_{1 \leq i \leq n} \sum_{k=i}^{n} Z_k$.</td>
</tr>
<tr>
<td>$N$</td>
<td>Ordinal number of measurements when alarm is raised.</td>
</tr>
<tr>
<td>$E[t_N]$</td>
<td>Physical time when alarm is raised.</td>
</tr>
<tr>
<td>$E[N]$</td>
<td>Average number of post-change measurements in the worst case.</td>
</tr>
<tr>
<td>$E_\infty[\hat{\tau}]$</td>
<td>Average stationary measurement period of adaptive measurement schedule.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Average run length guarantee. Compliant agents fulfill $E_\infty[t_N] \geq \gamma$.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Delay guarantee. Compliant agents fulfill $E[t_N] \leq \beta$.</td>
</tr>
<tr>
<td>$S^{(0)}$</td>
<td>Stopping threshold for considered agents - they stop iff $S_n &gt; S^{(0)}$.</td>
</tr>
<tr>
<td>$S^{(1)}$</td>
<td>Crisis mode threshold.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Measurement period of considered agents.</td>
</tr>
<tr>
<td>$\xi$</td>
<td>Maximal expected overshoot of $Z$, $\sup_{r \geq 0} \mathbb{E}_{X \sim f_1}[Z - r</td>
</tr>
<tr>
<td>$\sigma_Z$</td>
<td>Subgaussianity property of $Z$.</td>
</tr>
</tbody>
</table>


Appendix A

Omitted proofs

In this chapter, we present proofs of statements in papers I and II that was not included in the official publication, but which are nevertheless important to have a complete picture of the theory behind the problem formulations and solutions.

A.1 Proofs from Paper I

Proof of Theorem 2

Proof. We use a similar argument as in the proof of Theorem 1. Recall that \( \nu_\theta(u) \) denotes the distribution of the random observation when under policy \( u \). It is easy to see that if \( u = (k) \), the mapping \( \theta_k \to \nu_\theta(u) \) is one to one, and if \( u = (k, \ell, m) \) the mapping \( (\theta_k, \theta_\ell, \theta_m) \to \nu_\theta(u) \) is one-to-one. Denote by \( \Lambda(\theta) \) the set of confusing parameters, defined in equation (2.2). Since we have one-to-one mappings, it is again true that if \( u^*(\theta) = (1) \) we have \( \lambda_1 = \theta_1 \) and if \( u^*(\theta) = (1, 2) \) we have \( \lambda_1 = \theta_1, \lambda_2 = \theta_2 \). We furthermore have that if \( E^\pi[N_u(T)] \) is the expected number of rounds where \( \pi \) applies policy \( u \) up to time \( T \), for all \( \lambda \in \Lambda(\theta) \) and for large \( T \)

\[
\sum_u E^\pi[N_u(T)]KL(\nu_\theta(u)||\nu_\lambda(u)) \geq \log(T)(1 + o(1)). \tag{A.1}
\]

This implies that an asymptotic lower bound for the regret is \( C(\theta) \log(T) \), where \( C(\theta) \) is the value of the solution of the following optimization problem.

\[
\min_{\eta \geq 0} \sum_{u \in U} \eta_u(\mu^* - \mu(u)) \tag{A.2}
\]

s.t. \( \sum_{u \in U} \eta_uKL(\nu_\theta(u)||\nu_\lambda(u)) \geq 1, \forall \lambda \in \Lambda(\theta). \tag{A.3} \)

We can show, by precisely the same reasoning as in the proof of Theorem 1, that it is enough to consider \( \lambda \) which differ from \( \theta \) in only one component \( \lambda_k \) (with \( k \notin u^* \)) and with \( \lambda_k > \bar{\theta} \). It thus remains to solve the optimization problem \((A.2)\).
We will show that there exists an optimal solution $\eta$ such that, for most cases, $\eta_u = 0$. We will treat the case $u^*(\theta) = (1,2)$, the case $u^*(\theta) = (1)$ will be analogous. First, take $u = (k, \ell, m)$ with $m \neq k \neq \ell \neq m$ and $k > 1$, $l > 1$, $m > 1$. For any feasible solution $\eta$ such that $\eta_u > 0$, take the related, also feasible, solution $\eta'$ identical to $\eta$ except $\eta'_u = 0$, $\eta'_{(k,1,1)} = \eta_{(k,1,1)} + \eta_u$, $\eta'_{(1,l)} = \eta_{(1,l)} + \frac{1-p_0(\theta_k)}{p_0(\theta_1)} \eta_u$, $\eta'_{(1,m)} = \eta_{(1,m)} + \frac{p_0(\theta_k)}{p_0(\theta_1)} \eta_u$. Then, the difference between the cost functions of $\eta$ and $\eta'$ is $\eta_u$ multiplied by

\[
(\mu^* - \mu(k, \ell, m)) - (\mu^* - \mu(k,1,1))
= \mu(k,1,1) - \mu(k,\ell,m)
- \frac{1-p_0(\theta_k)}{p_0(\theta_1)}(\mu^* - \mu(1,\ell)) - \frac{p_0(\theta_k)}{p_0(\theta_1)}(\mu^* - \mu(1,m))
= \mu(k,1,1) - \mu(k,\ell,m)
- \frac{1-p_0(\theta_k)}{p_0(\theta_1)}p_0(\theta_1)(\theta_2 - \theta_\ell) - \frac{p_0(\theta_k)}{p_0(\theta_1)}p_0(\theta_1)(\theta_2 - \theta_m)
= (1 - p_0(\theta_k))(\theta_1 - \theta_\ell - (\theta_2 - \theta_\ell))
+ p_0(\theta_k)(\theta_1 - \theta_m - (\theta_2 - \theta_m)) > 0
\]

so clearly $\eta$ is suboptimal and there exists an optimal solution $\eta^*$ with $\eta_u^* = 0$. Furthermore, if $k > 1$, $\ell > 1$ and $k \neq \ell$, highly similar arguments can be used to show that $\eta^*_{(k,\ell)} = \eta_{(k,\ell,k)} = \eta^*_{(k,1,\ell)} = \eta^*_{(k,\ell,1)} = 0$. It can be concluded that for any $k > 1$, $\ell > 1$, $k \neq \ell$ if both $k \in u$ and $\ell \in u$ then $\eta_u^* = 0$. With this in mind, along with the pruned constraints, solving the optimization problem (A.2) comes down to, for all $k \notin u^*$, identifying for which $u$ such that $k \in u$ we have $\eta_u > 0$. Proving the general statement of the solution in equation (2.6) then comes down to showing that if $k \in u$ $\eta_u^* = 0$ unless $u \in \mathcal{U}_I(k)$. In other words, we wish to show that $\eta^*_{(k,k,k)} = 0$ (which is trivial, as $\mu((k,k,k)) < \mu((k))$ but the two policies carry the same information about arm $k$) and that $u_1 = 1 \implies \eta_u^* = 0$. When $u^*(\theta) = (1,2)$, we also need to show $\eta_{(k)} = 0$ but this is completely analogous to Theorem [1].

First, analogous to $u = (k,k,k)$, we easily find $\eta_{(1,k,k)}^* = 0$. Next, consider $u = (1, k, 1)$. We find that redistributing the weight to $(1, k)$ gives less regret per average observation of arm $k$, as

\[
\frac{\mu^* - \mu(1,1,k)}{p_0(\theta_1)I(\theta_k, \theta)} - \frac{\mu^* - \mu(1,k,1)}{p_0(\theta_1)(1-p_0(\theta_1))I(\theta_k, \theta)}
= -\frac{(1-2\varepsilon)(\theta_1(1-\theta_2) + \theta_2(1-\theta_1))}{p_0(\theta_1)(1-p_0(\theta_1))I(\theta_k, \theta)} < 0
\]

and so $\eta_u^* = 0$. Finally, consider $u = (1,k)$. For any $\eta$ with $\eta_u > 0$, consider $\eta'$ with $\eta'_u = 0$, $\eta'_v = \eta_v + \frac{p_0(\theta_k)}{KL(\nu_\theta(u)||\nu_{\theta(k,-}\theta)(u))} \eta_u$ for $v = (1,k)$ and $\eta'_w = \eta_w$ for all $w \neq u,v$. Clearly, $\eta'$ is feasible and the difference in objective functions is $\eta_u$.
multiplied by
\[
(\mu^* - \mu((1, k))) - p_0(\theta_1)I(\theta_k, \bar{\theta})\]
\[
KL(\nu_{\theta(u)}||\nu_{\theta(\bar{\theta})}(u))(\mu^* - \mu((k, 1)))
\]
\[
= (\mu^* - \mu((1, k)))\left(1 - \frac{p_0(\theta_1)I(\theta_k, \bar{\theta})}{KL(\nu_{\theta(u)}||\nu_{\theta(\bar{\theta})}(u))}\right).
\]

Note that
\[
KL(\nu_{\theta(u)}||\nu_{\theta(\bar{\theta})}(u)) = (1 - \varepsilon)\theta \log \left(\frac{\theta}{\theta}\right)
\]
\[
+ \varepsilon(1 - \theta_k) \log \left(\frac{1 - \theta_k}{1 - \theta}\right) + p_0(\theta_k) \log \left(\frac{p_0(\theta_k)}{p_0(\theta)}\right).
\]

It thus suffices to prove that for \(\theta_k < \theta_2\), \(f(\theta_k) := p_0(\theta_1)I(\theta_k, \bar{\theta})KL(\nu_{\theta(u)}||\nu_{\theta(\bar{\theta})}(u)) \leq 0\). Since, by the properties of the KL-divergence, \(f(\theta) = f'(\theta) = 0\), this follows if for all \(\theta_k < \theta, f''(\theta) = 0\).

We evaluate \(f''(\theta_k)\) for \(0 < \theta_k < \theta\). We have that \(\frac{\partial^2}{\partial \theta^2} \theta \ln(\frac{\theta}{\theta}) = \frac{1}{\theta^2}\) and that
\[
\frac{\partial^2}{\partial \theta^2} (1 - \theta_k) \ln(\frac{1 - \theta_k}{1 - \theta}) = \frac{1}{1 - \theta_k}.\]
From this, it is relatively straightforward to find that
\[
\frac{\partial^2}{\partial \theta^2} (\varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k)) \ln(\frac{\varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k)}{\varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k)}) = \frac{(1 - 2\varepsilon)^2}{(\varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k))}.\]
Then, we can write
\[
f''(\theta_k) = \frac{1}{\theta_k}(\varepsilon \theta_1 + (1 - \varepsilon)(1 - \theta_1) - (1 - \varepsilon))
\]
\[
+ \frac{1}{1 - \theta_k}(\varepsilon \theta_1 + (1 - \varepsilon)(1 - \theta_1) - \varepsilon)
\]
\[
- \frac{(1 - 2\varepsilon)^2}{(\varepsilon \theta_k + (1 - \varepsilon)(1 - \theta_k))}
\]
\[
\leq (1 - 2\varepsilon)\left(\frac{1 - \theta_1}{1 - \theta_k} - \frac{\theta_1}{\theta_k}\right) \leq 0
\]
where the first inequality is removal of a non-positive term and the second inequality comes from \(\theta_k < \theta_1\) as well as \(\varepsilon \leq \frac{1}{2}\). Since we have eliminated all possibilities, we have now found that \(u_1 = 1 \implies \eta_u^* = 0\), which directly leads to the result in Theorem 2.

**Proof of Theorem 3**

Our proof strategy is similar to that used in Combes et al. [34] or other analyses of the regret of bandit algorithms. Namely, we decompose the set of rounds into several subsets, and upper bound the regret generated in each of the subsets. In the following lemma, we show that thanks to the aggressive exploitation behavior of SPAM, the expected number of rounds where the leading policy is not \(u^*(\theta)\) is finite.
Lemma 3. Choose \( \delta \in (0, \delta_0) \), with \( \delta_0 \) defined in the statement of Theorem 3. We define the following sets:

\[
\mathcal{A} = \{ t \in \mathbb{N} : \mathcal{L}(t) \neq u^*(\theta) \} \\
\mathcal{D} = \{ t \in \mathbb{N} : (\exists i \in \mathcal{L}(t) : |\hat{\theta}_i(t) - \theta_i| \geq \delta) \}
\]

and \( \mathcal{C} = \mathcal{A} \cup \mathcal{D} \). Furthermore, we denote \( \beta = (1 - \theta_1)^{-1} \). Then, under Algorithm 1, we have

\[
\mathbb{E}[|\mathcal{C}|] \leq 4K[4(K + \beta^2) + \delta^{-2}(\beta + 1)] + 30. \tag{A.4}
\]

Proof of Lemma 3. Introduce the sets

\[
\mathcal{E} = \{ t \in \mathbb{N} : (\exists i \in u^*(\theta) : b_i(t) \leq \theta_i) \} \\
\mathcal{G} = \{ t \in \mathcal{A} \setminus (\mathcal{D} \cup \mathcal{E}) : (\exists i \in u^*(\theta) : i \notin \mathcal{L}(t), |\hat{\theta}_i(t) - \theta_i| > \delta \}.
\]

We will show that \( \mathcal{C} \subseteq \mathcal{D} \cup \mathcal{E} \cup \mathcal{G} \). Take \( t \in \mathcal{A} \) which does not fulfill \( \forall i \in u^*(\theta) \setminus \mathcal{L}(t) : |\hat{\theta}_i(t) - \theta_i| \leq \delta \). Clearly, \( t \in \mathcal{E} \cup \mathcal{G} \). Now take \( t \in \mathcal{A} \) such that this is fulfilled. First, treat the case \( u^*(\theta) = (1, 2) \). Since \( t \in \mathcal{A} \), either (a) \( \mathcal{L}(t) = (1) \) or (b) there exists \( i, j \) such that \( i \in u^*(\theta) \setminus \mathcal{L}(t) \), \( \theta_i > \theta_j \) and \( \hat{\theta}_j(t) > \hat{\theta}_i(t) \). Without loss of generality let \( \hat{\theta}_j(t) > \hat{\theta}_i(t) \). Then, since \( \theta_i > \theta_j \), we have that \( \hat{\theta}_i(t) > \theta_i + \delta \) and \( \hat{\theta}_j(t) > \theta_j + \delta \). Thus, if (a) is true, \( t \notin \mathcal{D} \) and the last inequality follows from the definition of \( \delta_0 \). Therefore we have \( t \in \mathcal{D} \). If (b) is true, \( t \notin \mathcal{A} \setminus (\mathcal{D} \cup \mathcal{E} \cup \mathcal{G}) \). Thus, if (a) is true, \( t \notin \mathcal{A} \) and the last inequality follows from the definition of \( \delta_0 \). Therefore we have \( t \in \mathcal{D} \).

Next, treat the case \( u^*(\theta) = (1, 2) \). Then, since \( t \in \mathcal{A} \) there exists \( i \in \mathcal{L}(t) \) such that either (a) \( \hat{\theta}_i(t) \geq \frac{\epsilon}{1 - \hat{\theta}_i(t)} \), or (b) \( \hat{\theta}_i(t) \geq \hat{\theta}_1(t) \). In both cases, either \( t \notin \mathcal{D} \) or \( |\hat{\theta}_i(t) - \theta_i| \leq \delta \). We focus on the case \( |\hat{\theta}_i(t) - \theta_i| \leq \delta \). If (a) is true, \( \hat{\theta}_i(t) \geq \frac{\epsilon}{1 - \hat{\theta}_i(t)} \geq \frac{\epsilon}{1 - \hat{\theta}_i + \delta} > \theta_i + \delta \) with the strict inequality following from the definition of \( \delta_0 \), thus we have \( t \in \mathcal{D} \). If (b) is true, \( t \in \mathcal{D} \) and the same reasoning as in the case \( u^*(\theta) = (1, 2) \). No matter what, we have \( t \in \mathcal{D} \), or in other words, \( \mathcal{C} \subseteq \mathcal{D} \cup \mathcal{E} \cup \mathcal{G} \).

Now, we wish to bound \( \mathbb{E}[|\mathcal{D}|] \), \( \mathbb{E}[|\mathcal{E}|] \) and \( \mathbb{E}[|\mathcal{G}|] \). The result will follow by a union bound.

Decompose \( \mathcal{D} = \bigcup_{i=1}^{K} \mathcal{D}_i \), where \( \mathcal{D}_i = \{ t \in \mathbb{N} : i \in \mathcal{L}(t), |\hat{\theta}_i(t) - \theta_i| \geq \delta \} \). Note that by the definition of the algorithm, the probability of observing arm \( i \) given that \( i \in \mathcal{L}(n) \) (and therefore, given that \( t \in \mathcal{D}_i \)), is at least \( \frac{\beta^{-1}}{2} \). Thus, by Lemma 5 of Combes et al. \[34\] with \( H = \mathcal{D}_1 \) and \( c = \frac{\beta^{-1}}{2} \) we have that \( \mathbb{E}[|\mathcal{D}_i|] \leq 4\beta[4\beta + \delta^{-2}] \) and by a union bound

\[
\mathbb{E}[|\mathcal{D}|] \leq 4K\beta[4\beta + \delta^{-2}].
\]
Next, for any $i \in u^*(\theta)$, let $E_i = \{t \in \mathbb{N} : b_i(t) \leq \theta_i\}$. It follows that $E = \bigcup_{i \in u^*(\theta)} E_i$. By Lemma 6 of Combes et al., we have that $\mathbb{E}[|E_i|] \leq 15$ and thereby, by a union bound,

$$\mathbb{E}[|E|] \leq 2 \times 15 = 30,$$

since there can be at most 2 distinct elements in $u^*(\theta)$.

Finally, for any $i \in u^*(\theta)$, let $E_i = \{t \in \mathbb{N} : b_i(t) \leq \theta_i\}$. It follows that $E = \bigcup_{i \in u^*(\theta)} E_i$. Then $G = \bigcup_{i \in u^*(\theta)} G_i$.

Consider $i = 1$ and choose $t \in G_1$. Then, since $t \notin E$ we have $b_1(t) \geq \theta_1$. Furthermore, since $t \notin D$ there exists $j \in \mathcal{L}(t) : j > 1$ such that

$$\hat{\theta}(t) \leq \hat{\theta}_j(t) \leq \theta_j + \delta \leq \frac{\theta_j + \theta_1}{2} < \theta_1 \leq b_1(t).$$

Thus, we have $1 \in \mathcal{B}(t)$.

Consider now $i = 2$ (in which case $u^*(\theta) = (1, 2)$) and choose $t \in G_2$. Then, since $t \notin A$, either there exists $j > 2$ such that $j \in \mathcal{L}(t)$ (in which case, the exact same argument as in the case $i = 1$ applies) or $\mathcal{L}(t) = \{1\}$. In the latter case, since $t \notin E$ we have $b_2(t) \geq \theta_2$ and since $t \notin D$ we have $|\theta_1(t) - \theta_1| \leq \delta$. Thus, by definition of $\delta_0$

$$\hat{\theta}(t) = \frac{c}{\theta_1(t)} \leq \frac{c}{1 - \theta_1 - \delta} < 2 \leq b_2(t).$$

Thus, no matter what, $i \in \mathcal{B}(t)$. By the definition of the algorithm, the probability of observing $i$ given that $i \in \mathcal{B}(t)$ is at least $\frac{1}{2K}$. Then, we can once again employ Lemma 5 of Combes et al. with $H = G_i$ and $c = \frac{1}{2K}$ to find that $\mathbb{E}[|G_i|] \leq 4K(4K + \delta^{-2})$. This immediately yields

$$\mathbb{E}[|G|] \leq 4K[4K + \delta^{-2}].$$

By a union bound, we find

$$\mathbb{E}[|C|] \leq \mathbb{E}[|D|] + \mathbb{E}[|E|] + \mathbb{E}[|G|] \leq 4K[4(K + \beta^2) + \delta^{-2}(\beta + 1)] + 30,$$

which is the desired result. \hfill \blacksquare

**Proof of Theorem** Define $\mathcal{K}_i^1 = \{t \in [1, T] : t \notin C, u(t) = (i, 1)\}$ and $\mathcal{K}_2^2 = \{t \in [1, T] : t \notin C, u(t) = (i)\}$. By design of the algorithm, if $t \notin C$, the algorithm will either play the optimal policy or it will play $(i, 1)$ or $(i)$ for some $i \notin u^*(\theta)$. Since $\mu^* - \mu(u) \leq 1 + c$ for all $u$, we can decompose the regret as

$$R_{\delta}^{\text{SPAM}}(T) \leq (1 + c)\mathbb{E}[|C|] + \sum_{i \notin u^*(\theta)} [\mu^* - \mu((i, 1))]\mathbb{E}[|\mathcal{K}_i^1|] + \sum_{i \notin u^*(\theta)} [\mu^* - \mu((i))]\mathbb{E}[|\mathcal{K}_i^2|].$$
We now bound $\mathbb{E}\left[|K_i^1|\right]$ (the bound on $\mathbb{E}\left[|K_i^2|\right]$ will be analogous). Recall that $g(\theta_1, \theta_2, \delta)$ is defined such that

\[
g(\theta_1, \theta_2, \delta) := \begin{cases} 
\theta_1 - \delta, & \bar{\theta} = \theta_1 \\
\theta_2 - \delta, & \bar{\theta} = \theta_2 \\
c / (1 - \bar{\theta} + \delta), & \bar{\theta} = c / (1 - \bar{\theta}).
\end{cases}
\]

Note that $g(\theta_1, \theta_2, 0) = \bar{\theta}$. Now, choose $\epsilon \in (0, 1)$, define the number of elements in $K_i^1$ up to time $t$ as $k_i(t) := \sum_{s=1}^{t} I(s \in K_i^1)$, and define $n_0 = f(T) / (I(\theta_1 + \delta, g(\theta_1, \theta_2, \delta)))$. Then, we wish to decompose $K_i^1$ into $K_{i,1}^1 \cup K_{i,2}^1$, where

\[
K_{i,1}^1 = \{ t \in K_i^1 : n_i(t) \leq (1 - \epsilon)k_i(t) \text{ or } |\hat{\theta}_i(t) - \theta_i| \geq \delta \} \\
K_{i,2}^1 = \{ t \in K_i^1 : n_0 \geq (1 - \epsilon)k_i(t) \text{ and } |\hat{\theta}_i(t) - \theta_i| < \delta \}.
\]

Now we show that this decomposition is valid, by contradiction. Take $t$ in $K_{i,1}^1 \setminus (K_{i,1}^1 \cup K_{i,2}^1)$. Since $t \notin K_{i,1}^1$, $n_i(t) \geq (1 - \epsilon)k_i(t)$ and since $t \notin K_{i,2}^1$, $(1 - \epsilon)k_i(t) \geq n_0$, so $n_i(t) \geq n_0$, which we call inequality (a).

Furthermore, since $t \in K_i^1$ and by design of the algorithm, we get $i \in B(t)$ which in turn implies $b_i(t) \geq \hat{\theta}(t)$. Since $t \notin C$ and $\delta < \delta_0$ we must have (by definition of $\delta_0$) $\hat{\theta}(t) = g(\hat{\theta}_1(t), \hat{\theta}_2(t), 0) \geq g(\theta_1, \theta_2, \delta)$. Therefore we have $b_i(t) \geq g(\theta_1, \theta_2, \delta)$, which we call inequality (b).

Putting inequalities (a) and (b) together with the definition of $b_i(t)$ we obtain

\[
n_0I(\hat{\theta}_i(t), g(\theta_1, \theta_2, \delta)) \leq n_i(t)I(\hat{\theta}_i, g(\theta_1, \theta_2, \delta)) \leq f(t) \leq f(T)
\]

and so, by definition of $n_0$, we obtain $I(\hat{\theta}_i(t), g(\theta_1, \theta_2, \delta)) \leq I(\theta_i + \delta, g(\theta_1, \theta_2, \delta))$ which by monotonicity of $I(x, g(\theta_1, \theta_2, \delta))$ on the interval $[0, g(\theta_1, \theta_2, \delta)]$ implies that $\hat{\theta}_i(t) \geq \theta_i + \delta$. But then $t \in K_{i,1}^1$, which is a contradiction. Therefore $K_i^1 \subseteq K_{i,1}^1 \cup K_{i,2}^1$.

Next, we bound $\mathbb{E}\left[|K_{i,1}^1|\right]$ and $\mathbb{E}\left[|K_{i,2}^1|\right]$. First, note that the probability of observing arm $k$ given that $t \in K_i^1$ (or for that matter, given that $t \in K_i^2$) is 1. Next, we can use Corollary 1 in [34] with $H = K_{i,1}^1$ and $c = 1$ to bound $\mathbb{E}\left[|K_{i,1}^1|\right] \leq \epsilon^{-2} + (1 - \epsilon)^{-1}\delta^{-2}$.

Finally, note that by definition of $\delta_0$, if it is true that $(1 - \theta_i)\theta_1 < c$ then, if $t \in K_{i,2}^1$ we have $(1 - \hat{\theta}_1(t))\hat{\theta}_i(t) \leq (1 - \theta_i + \delta)(\theta_i + \delta) < c$ since $\theta_i < \theta_2$. But by design of the algorithm, $u(t) \neq (k, 1)$ which is a contradiction. Therefore, if $(1 - \theta_i)\theta_1 < c$, it follows that $\mathbb{E}\left[|K_{i,2}^1|\right] = 0$. We also have that if $t \in K_{i,2}^1$ then $k_i(t) \leq (1 - \epsilon)^{-1}n_0$. Since $k_i(t)$ is incremented at $t$, we then have that $\mathbb{E}\left[|K_{i,2}^1|\right] \leq (1 - \epsilon)^{-1}n_0$ and in total

\[
\mathbb{E}\left[|K_{i,2}^1|\right] \leq \mathbb{I}((1 - \theta_i)\theta_1 \geq c)(1 - \epsilon)^{-1}n_0
\]
Now we put everything together (with an analogous bound on \( E[|K_i^2|] \)) and we obtain

\[
R^{SPAM}_\theta(T) \leq (1 + c)(4K[4(K + \beta^2) + \delta^{-2}(\beta + 1)] + 30) \\
+ 2(1 + c)K[\epsilon^{-2} + (1 - \epsilon)^{-1}\delta^{-2}] \\
+ \sum_{i \notin u^*(\theta)} \frac{1}{(1 - \epsilon)I(\theta_i, g(\theta_1, \theta_2, \delta))} f(T) \\
+ \sum_{i \notin u^*(\theta)} \frac{1}{(1 - \epsilon)I(\theta_i, g(\theta_1, \theta_2, \delta))} f(T) \\
= 2K(1 + c)[8(K + \beta^2) + \epsilon^{-2} \\
+ \delta^{-2}(2(\beta + 1) + (1 - \epsilon)^{-1}] \\
+ \sum_{i \notin u^*(\theta)} \frac{H_i(\theta)I(\theta_i, \tilde{\theta})}{(1 - \epsilon)I(\theta_i, g(\theta_1, \theta_2, \delta))} f(T)
\]

which directly leads to the result of Theorem 3.

\[\blacksquare\]

A.2  Proofs from Paper II

Lower bound proofs

Proof of Proposition 2

Recall that \( \Lambda \) is the simplex of dimension \((K - 1) \). For any correct answer \( \ell \in C(\mu) \), denote by \( \text{Alt}(\ell) := \{ \lambda : \ell \notin C(\lambda) \} \), and denote by \( D(w, \mu, \lambda) := \sum_{k=1}^{K} w_k d(\mu_k, \lambda_k) \). We wish to solve the max-min problem

\[
w^*(\mu, \ell) \in \arg\max_{w \in \Lambda} \inf_{\lambda \in \text{Alt}(\ell)} D(w, \mu, \lambda).
\]

We begin with the case \( \mu^* > \gamma \). Then \( C(\mu) = \{0\} \) by definition. We have that \( \text{Alt}(0) = \{ \lambda : \exists k \in [K] : \lambda_k < \gamma \} \), and from this set, we can restrict ourselves to studying only \( \lambda_m^{(k)}(\nu) = \nu \mathbb{1}_{(m=k)} + \mu_m \mathbb{1}_{(m \neq k)} \) with \( \nu < \gamma \) (only a single arm \( k \) is changed compared to \( \mu \)). Indeed, for any instance \( \lambda \in \text{Alt}(\mu) \) and any weights \( w \in \Lambda \), there exists \( k \) and \( \nu \) such that \( D(w, \mu, \lambda) \geq D(w, \mu, \lambda^{(k)}(\nu)) \). It follows that for any \( w \in \Lambda \), \( \inf_{\lambda \in \text{Alt}(\ell)} D(w, \mu, \lambda) = \min_{k \in [K]} D(w, \mu, \lambda^{(k)}(\gamma)) = \min_{k \in [K]} w_k d(\mu_k, \gamma) \). We note then that for maximizing \( w = w^*(\mu, 0) \), it must be true that \( D(w^*(\mu, 0), \mu, \lambda^{(k)}(\gamma)) = D(w^*(\mu, 0), \mu, \lambda^{(m)}(\gamma)) \) \( \forall k \in [K], m \in [K] \).

Adding the condition \( \sum_{k=1}^{K} w_k = 1 \), this yields a linear equation system which is easily solved with

\[
w^*_k(\mu, 0) = \frac{d(\mu_k, \gamma)^{-1}}{\sum_{j=1}^{K} d(\mu_j, \gamma)^{-1}} \tag{A.5}
\]
APPENDIX A. OMITTED PROOFS

and therefore, \( T_1(\mu) = \sum_{k=1}^{K} d(\mu_k, \gamma)^{-1} \), which concludes this case.

Next, we study the case \( \mu_* < \gamma \). Take any arm \( \ell \in C(\mu) \), so \( \mu_\ell < \gamma \). For this case, \( \text{Alt}(\ell) \) are all problems \( \lambda \) such that \( \lambda_\ell > \gamma \). Similarly as before, it is easy to see that \( \inf_{\lambda \in \text{Alt}(\ell)} D(w, \mu, \lambda) = D(w, \mu, \lambda(\ell)(\gamma)) = w_\ell d(\mu_\ell, \gamma) \). Furthermore, this expression is maximized under \( w \in \Lambda \) by \( w^*(\mu, \ell) \) with \( w^*_k(\mu, \ell) = 1(k = \ell) \forall k \in [K] \). Maximizing over \( \ell \in C(\mu) \), we find that the above expression is maximized when \( \mu_\ell \) is minimized, which occurs when \( \ell = k_* \), so \( T_1(\mu) = d(\mu_\gamma, \gamma)^{-1} \). This concludes this case and the proof.

Proof of Proposition 3

We will assume that \( \mu_* < \mu^* < \gamma \) (for the cases \( \mu^* > \gamma \) or \( \mu_\gamma = \mu^* < \gamma \), refer to the proof of Proposition 2). In this case, \( C(\mu) = \{k^*\} \), with \( k^* \) unique by hypothesis. We notice that \( \text{Alt}(k^*) \) can be written as the union of three sets, \( \text{Alt}(k^*) \subseteq A_1 \cup A_2 \cup A_3 \) where

1. \( A_1 := \{ \lambda : \exists a : \mu_a > \gamma, \lambda_a < \gamma \} \),
2. \( A_2 := \{ \lambda : \exists a : \mu_a < \mu^*, \lambda_{k^*} < \lambda_a < \gamma \} \),
3. \( A_3 := \{ \lambda : \lambda_{k^*} > \gamma \} \).

Hence, \( \inf_{\lambda \in \text{Alt}(k^*)} D(w, \mu, \lambda) = \inf_{\lambda \in A_1 \cup A_2 \cup A_3} D(w, \mu, \lambda) \). We will first find the value of \( \inf_{\lambda \in A_i} D(w, \mu, \lambda) \) for \( i \in \{1, 2, 3\} \). This value is

1. for \( A_1 \), \( \min_{k: \mu_k > \gamma} w_k d(\mu_k, \gamma) \),
2. for \( A_2 \), \( \min_{k: \mu_k < \mu^*} w_k I_{w_{k^*}/(w_k + w_{k^*})}(\mu^*, \mu_k) \),
3. and for \( A_3 \), \( w_k d(\mu^*, \gamma) \).

Let \( D^*(w, \mu, \text{Alt}(k^*)) \) be the minimum of these three expressions. Since the sets of modified arms are non-overlapping between \( A_1 \) and \( A_2 \cup A_3 \), we see that for each arm \( k \) with \( \mu_k > \gamma \) the expression \( w_k d(\mu_k, \gamma) \) must be equal and identical to \( D^*(w, \mu, \text{Alt}(k^*)) \) for any maximizing \( w \), and thereby, for these arms \( w^*_k(\mu, k^*) = D^*(w^*(\mu, k^*), \mu, \text{Alt}(k^*))/d(\mu_a, \gamma) \).

For \( A_2 \), the best proportions \( w_k \) are known from Best Arm Identification problems \(^{50}\) and can be found as \( w_k = \frac{x_k(y^*)D^*(w, \mu, \text{Alt}(k^*))}{y^*} \) for all \( k \) with \( \mu_k < \gamma \), where \( x_k(y) \) and \( y^* \) are defined in the proposition.

Now, if \( y^* \leq d(\mu^*, \gamma) \) then \( D^*(w, \mu, \text{Alt}(k^*)) \geq D^*(w, \mu, \text{Alt}(k^*)) \) and so these proportions maximize the expression \( \min_{\lambda \in A_2 \cup A_3} D(w, \mu, \lambda) \).

If instead \( y^*> d(\mu^*, \gamma) \), by convexity of \( D^*(\cdot, \mu, \text{Alt}(k^*)) \), we see that \( w_k = \frac{D^*(w, \mu, \text{Alt}(k^*))}{d(\mu_k, \gamma)} \) is maximizing and so, since the expression \( I_{w_{k^*}/(w_a + w_{k^*})}(\mu^*, \mu_a) \) must be identical to \( D^*(w, \mu, \text{Alt}(k^*)) \) for all arms \( k \) with \( \mu_k < \gamma \), we obtain \( w_k = \)}
A.2. PROOFS FROM PAPER II

\[ D^*(w, \mu, \text{Alt}(k^*)) f \equiv D^*(w^*(\mu, k^*), \mu, \text{Alt}(k^*)) \forall k : \mu_k < \gamma. \]

In summary, recalling that \( z^* = \min(d(\mu^*, \gamma), y^*) \), we find that \( w^*_k(\mu, k^*) = D^*(w^*(\mu, k^*), \mu, \text{Alt}(k^*)) \) \( \forall k : \mu_k < \gamma. \)

Finally, we make use of the fact that \( \sum_k w_o k = 1 \), and find that

\[
D^*(w^*(\mu, k^*), \mu, \text{Alt}(k^*)) = \left( \sum_{k: \mu_k > \gamma} d(\mu_k, \gamma)^{-1} + \frac{1}{z^*} \sum_{k: \mu_k < \gamma} x^*_k(z^*) \right)^{-1}.
\]

But since these proportions \( w^*(\mu, k^*) \) are maximizing, we find that any \( \delta \)-PC algorithm must be sampled in expectation at least \( T_2(\mu) d_B(\delta, 1 - \delta) \) times, with \( T_2(\mu) = D^*(w^*(\mu, k^*), \mu, \text{Alt}(k^*))^{-1} = \sum_{k: \mu_k > \gamma} d(\mu_k, \gamma)^{-1} + \frac{1}{z^*} \sum_{k: \mu_k < \gamma} x^*_k(z^*). \)

This concludes the proof.

Proof of Proposition 4

We will assume that \( \mu_* < \gamma \) as the case \( \mu^* > \gamma \) can be treated as in Proposition 3. Thus \( C(\mu) = \{ k_* \} \) with \( k_* \) unique by hypothesis. As in the proof of Proposition 3, we note that \( \text{Alt}(k_*) \subseteq A_2 \cup A_3 \) with \( A_2 = \{ \lambda : \exists k : \lambda_k < \lambda_{k_*} \} \) and \( A_3 = \{ \lambda : \lambda_{k_*} > \gamma \} \), and furthermore

\[
\inf_{\lambda \in \text{Alt}(k^*)} D(w, \mu, \lambda) = \inf_{\lambda \in A_2 \cup A_3} D(w, \mu, \lambda).
\]

Next, like before we find that

\[
\inf_{\lambda \in A_2} D(w, \mu, \lambda) = \min_{k: \mu_k > \mu_*} w_{k_*, I_{w_{k_*/(w_k + w_{k_*)}}} (\mu_*, \mu_k)}
\]

and \( \inf_{\lambda \in A_2} D(w, \mu, \lambda) = w_{k_*} d(\mu_*, \gamma)^{-1}. \)

As in the proof of Proposition 3 if \( y_* \leq d(\mu_*, \gamma) \), then the optimal proportions are those of a Best Arm Identification problem and \( w^*_k(\mu, k_*) = \frac{x_{k_*/y_*}}{y_*} T_3(\mu) \) with \( T_3(\mu) := \frac{1}{y_*} \sum_{k=1}^K x_{k_*/y_*} \) for all \( k \) with \( y_* \) and \( x_{k_*/y_*} \) defined as in Section 3.4.

Otherwise, by convexity of \( D(\cdot, \mu, \text{Alt}(k_*)) \) and by \( w_{k_*} d(\mu_*, \gamma) = D(w, \mu, \text{Alt}(\mu)) \), we find that \( w^*_k(\mu, k_*) = d(\mu_*, \gamma)^{-1} D(w, \mu, \text{Alt}(\mu)) \). Since then

\[
d(\mu_*, \gamma)^{-1} D(w, \mu, \text{Alt}(\mu)) h(w_k / w_{k_*}) = D(w, \mu, \text{Alt}(\mu))
\]

it follows that \( w^*_k(\mu, k_*) = w^*_k(\mu, k_*) x_{k_*/k} d(\mu_*, \gamma) \) for all \( k \). Utilizing \( \sum_{k=1}^K w_k = 1 \) and defining in this case \( T_3(\mu) := D(w, \mu, \text{Alt}(\mu))^{-1} = \sum_{k=1}^K d(\mu_*, \gamma)^{-1} x_{k_*/k} d(\mu_*, \gamma) \), Equation (3.4) follows.

Upper bound proofs and correctness of TaS

In this section we prove Theorem 4. The proof logic closely follows that of [50].

We will show that this result is more generally applicable.
First, take $\epsilon > 0$ and define the event $E_T(\epsilon) = \{ ||\hat{\mu}(t) - \mu||_{\infty} < \xi(\epsilon) \}$, where $\xi(\epsilon) > 0$ is the greatest value such that
\[ ||\mu' - \mu||_{\infty} < \xi(\epsilon) \implies C^*(\mu) \subseteq C(\mu'), \]
\[ |w^*(\mu', k_E(\mu)) - w^*(\mu, k_E(\mu))| < \epsilon. \]
Such $\xi(\epsilon)$ necessarily exists by continuity of $w^*$ (Proposition 6 of [50], Theorem 4 of [51]). Given a sufficiently small value of $\epsilon$, (say $\epsilon < \epsilon'(\mu)$ for some function $\epsilon'$) it holds that on $E_T(\epsilon)$, $k_E(t) \in C^*(\mu)$. Under the specified uniqueness assumption, $C^*(\mu) := \{ k_E(\mu) \}$ is a singleton and so on $E_T(\epsilon)$ for $t \geq \sqrt{T}$,
\[ Q(t) = t \inf_{\lambda \in Alt(k_E(\mu))} D(N(t)/t, \mu, \lambda). \]

Further, by Lemma 20 of [50], for TAS there exists $T_\epsilon$ independent of $\delta$ such that for $T \geq T_\epsilon$ on $E_T(\epsilon)$ it holds that $\forall t \geq \sqrt{T}$, $||N(t)/t - w^*(\mu)||_{\infty} \leq 3(K-1)\epsilon$. Introducing $C_\epsilon^*(\mu) = \inf_{\mu': ||\mu' - \mu||_{\infty} \leq \xi(\epsilon), w': ||w' - w^*(\mu)||_{\infty} \leq 3(K-1)\epsilon} D(w', \mu', Alt(\lambda))$ it then follows that on $E_T(\epsilon)$ for every $T \geq T_\epsilon$ and $t \geq \sqrt{T}$
\[ Q(t) \geq tC_\epsilon^*(\mu). \]

Then, on $E_T(\epsilon)$ and with $T \geq T_\epsilon$, we have
\[ \min(t_\delta, T) \leq \sqrt{T} + \sum_{t=1}^{T} \mathbb{I}(t_\delta > t) \leq \sqrt{T} + \sum_{t=1}^{T} \mathbb{I}(Q(t) \leq f_\delta(t)) \]
\[ \leq \sqrt{T} + \sum_{t=1}^{T} \mathbb{I}((t) \leq f_\delta(t)) \leq \sqrt{T} + \frac{f_\delta(T)}{C_\epsilon^*(\mu)} \]
Introducing $T_0(\delta) = \inf \{ T : T \geq \sqrt{T} + \frac{f_\delta(T)}{C_\epsilon^*(\mu)} \}$, it follows that for every $T \geq \max(T_0(\delta), T_\epsilon)$, $E_T(\epsilon) \subseteq (t_\delta \leq T)$. As such, we have
\[ \mathbb{E}_\mu[\tau_\delta] = \sum_{T=1}^{\infty} \mathbb{P}(\tau_\delta > T) \leq T_0(\delta) + T_\epsilon + \sum_{T=1}^{\infty} \mathbb{P}(E_T^c(\epsilon)) \]
\[ \leq T_0(\delta) + T_\epsilon + \sum_{T=1}^{\infty} BT \exp(-CT^{1/8}) \]
where the final inequality follows from Lemma 19 in [50] for some constants $B$ and $C$ depending on $\epsilon$ and $\mu$ but not $\delta$. Following the remainder of the proof step by step, we find that
\[ \limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\log(1/\delta)} \leq \limsup_{\delta \rightarrow 0} \frac{T_0(\delta)}{\log(1/\delta)} \leq \frac{2(1 + \eta)}{C_\epsilon^*(\mu)} \]
for some $\eta > 0$ of our choice. By Theorem 4 of [51], the function $(w, \mu) \mapsto D(w, \mu, Alt(\ell))$ is continuous on $\Lambda \times \mathcal{M}$ for any $\ell \in C(\mu)$, so letting $\epsilon \rightarrow 0$,
A.2. PROOFS FROM PAPER II

$C^*_e(\mu) \to D(w^*(\mu, k_E(\mu)), \mu, \text{Alt}(k_E(\mu)))$. But by Propositions 2, 3 and 4, we see that this value is identical to $T_1(\mu)^{-1}$, $T_2(\mu)^{-1}$ and $T_3(\mu)^{-1}$ respectively, and we obtain the desired bound by letting $\eta \to 0$.

For the $\delta$-PC property, we will use a proof strategy similar to that of Proposition 12 in [50]. We will show that whenever an error occurs, $f_\delta(t) < Q(t) < \sum_{k=1}^{K} N_k(t)(\hat{\mu}_k(t), \mu_k)$. Then, $\delta$-correctness follows from Theorem 2 in [87].

The first inequality is an immediate consequence of the statement of TaS, so we focus on the second. For any-available-slice, packing-slice or least-loaded-slice, all errors can be divided into two categories, and we will show the inequality for both:

(i) For some arm $\ell$, $\hat{\mu}_\ell(t) < \gamma < \mu_\ell$ or $\mu_\ell < \gamma < \hat{\mu}_\ell(t)$. In either case, for this to be considered an error it must be true that $Q(t) \leq N_\ell(t)d(\hat{\mu}_\ell(t), \gamma) < N_\ell(t)d(\hat{\mu}_\ell(t), \mu_\ell) \leq \sum_{k=1}^{K} N_k(t)d(\hat{\mu}_k(t), \mu_k)$, which concludes this category.

(ii) For some arms $j, \ell: \hat{\mu}_j(t) < \hat{\mu}_\ell(t)$ and $\mu_\ell < \mu_j$. Defining $\alpha = N_j(t)/(N_j(t) + N_\ell(t))$, if this is considered an error, it can be shown (see the proof of Proposition 12 in [50]) that, $Q(t) < N_j(t)d(\hat{\mu}_j(t), \alpha \hat{\mu}_j(t) + (1-\alpha)\hat{\mu}_\ell(t)) + N_\ell(t)d(\hat{\mu}_\ell(t), \alpha \hat{\mu}_j(t) + (1-\alpha)\hat{\mu}_\ell(t)) \leq N_j(t)d(\hat{\mu}_j(t), \mu_j) + N_\ell(t)d(\hat{\mu}_\ell(t), \mu_\ell) \leq \sum_{k=1}^{K} N_k(t)d(\hat{\mu}_k(t), \mu_k)$. This concludes the proof.
References


