STARS

University of Central Florida

STARS

Electronic Theses and Dissertations, 2004-2019

2016

Compressive and Coded Change Detection: Theory and Application to Structural Health Monitoring

Omid Sarayanibafghi
University of Central Florida

Part of the Electrical and Computer Engineering Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

STARS Citation

University of Central Florida

STARS
Showcase of Text, Archives, Research & Scholarship

COMPRESSIVE AND CODED CHANGE DETECTION: THEORY AND APPLICATION TO
STRUCTURAL HEALTH MONITORING

by

OMID SARAYANIBAFGHI
B.S. Sharif University of Technology, 2013

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the Department of Electrical and Computer Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2016

Major Professor: George Atia

ABSTRACT

In traditional sparse recovery problems, the goal is to identify the support of compressible signals using a small number of measurements. In contrast, in this thesis the problem of identification of a sparse number of *statistical changes* in stochastic phenomena is considered when decision makers only have access to compressed measurements, i.e., each measurement is derived by a subset of features. Herein, we propose a new framework that is termed Compressed Change Detection. The main approach relies on integrating ideas from the theory of identifying codes with change point detection in sequential analysis. If the stochastic properties of certain features change, then the changes can be detected by examining the covering set of an identifying code of measurements. In particular, given a large number N of features, the goal is to detect a small set of features that undergoes a statistical change using a small number of measurements. Sufficient conditions are derived for the probability of false alarm and isolation to approach zero in the asymptotic regime where N is large.

As an application of compressed change detection, the problem of detection of a sparse number of damages in a structure for Structural Health Monitoring (SHM) is considered. Since only a small number of damage scenarios can occur simultaneously, change detection is applied to responses of pairs of sensors that form an identifying code over a learned damage-sensing graph. Generalizations of the proposed framework with multiple concurrent changes and for arbitrary graph topologies are presented.

I dedicate my thesis work to my family and many friends. I am very grateful to my parents whose unlimited support both financially and intellectually made this thesis possible. I dedicate this work and special thanks to my advisor.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Physical and natural processes often exhibit variability over several dimensions such as space and time. Thus, one of the goals of monitoring these processes is to discover the deviation from normal behavior. This goal is of paramount importance in order to give decision makers enough margin to take possible preventive actions. It may be crucial to detect and localize changes as soon as they occur. For example, structural health monitoring (SHM) systems are intended to quickly detect structural changes and damage in the civil infrastructure for the safety assessment of the monitored structures [1]. Similarly, in surveillance applications it is important to recognize anomalous interactions in social processes for security purposes [2]. On the other hand, the information acquired during the monitoring phase is often of tremendous volume and high dimensional. For instance data collected from sensors measuring acceleration, strain, temperature, and displacement of the structure, provide a huge amount of data on a daily basis. Analyzing all the information is either expensive or computationally infeasible. Thus solving inference problems while processing less amount of data is always preferable.

One of the inference problems is change point detection. In the classical theory, the goal is to detect a change in the statistical properties of sequentially collected data, as soon as it occurs subject to false alarm constraints. The pioneering work of Shiryaev [3] provides optimal stopping rules for a single stream of data in Bayesian settings, where the change-point itself is random with a prior geometric distribution. Recently, Tartakovsky et. al. [4] considered a setting in which information is collected from multiple sensors. In this work the statistical properties of all sensors change after the change-point. In another work, Premkumar et al. in [5] consider the problem of event detection and localization in wireless sensor networks. In their setup each sensor is driven by white noise and the sensing model at each sensor depends on distance.

Our work is motivated by two observations. First, the statistical changes in stochastic phenomena are typically sparse. Second, the effect of change is often localized, for example in the spatial domain in SHM and within neighbors on social network graphs. While traditional sparse signal processing is concerned with the recovery and/or reconstruction of the support of sparse signals, the focus here is on the recovery of *sparse statistical changes*. Leveraging sparsity can generally lead to significant reductions in the sampling rates and the storage requirements of modern algorithms and technologies dealing with multi-dimensional compressible signals. Now the question is whether similar gains can be achieved when the sparsity is associated with the change process, and how such gains can be achieved.

Classical change point detection has been used to identify changes in the distributions of stochastic processes and time series [6]. A dominant stream of research in this area focused on the design of efficient change detection rules with favorable delay and false alarm tradeoffs [4, 7]. A large number of algorithms, including CUSUM, MPCA were proposed to minimize the detection delay subject to constraints on the probability of false alarm in different settings [6, 8, 9]. More recently, [5, 10] studied the change detection problem in a distributed setting using multi-sensor observations. Also, decentralized change detection was considered in [4, 7, 10] wherein sensors send their local decisions rather than their raw observations to a fusion center to decide if a change has occurred. The problem of detecting a change at an unknown subset of multiple data streams using centralized detection rules is considered in [11]. They assume each sensor can be influenced with a certain probability. The problem in which only an unknown subset of data undergoes a change by imposing a constraint on measurement and communication cost is investigated in [12].

Amongst numerous applications of the proposed method, we focus on Structural Health Monitoring (SHM). SHM deals with the continuous monitoring of infrastructures for safety and condition assessment [1]. One of the main goals of SHM is to quickly detect structural damages to take possible preventive actions. However, given the massive amounts of multi-dimensional data pro-

duced daily from multi-modal sensors installed on a structure, it is computationally inefficient to process all the available information. Various methods deal with extracting the damage sensitive features (DSF) from the sensors installed on structures in a training mode and then in the monitoring phase, meaningful deviations in these features will result in declaration of damage. For instance, [13] considers the autoregressive coefficient of vibration data as DSF. [14, 15] consider the cross correlation between measurement of pair of sensors as DSF. In this method, the mean and variance of the DSF in the training set is calculated and a damage is declared when its DSF exceeds a threshold. Furthermore, structural damages are often localized, and hence only affect a small subset of measurements. Moreover, damage is typically sparse since only a small number of changes can occur simultaneously. We develop and analyze a new approach for damage detection and sensor placement in civil infrastructures, and provide a case study using a 4-span laboratory bridge model [16].

Contrast to Prior Work

Related work considered the problem of location detection in sensor networks using identifying codes [17]. Finding identifying codes over graphs was considered in [18] and different greedy algorithms were proposed to find near-optimal identifying codes. Similar ideas were used to detect faults in multiprocessors and communication systems in [19]. In contrast to prior work, the focus here is not on deterministic or static phenomena, but rather on stochastically changing phenomena. In particular, identifying codes, which have been traditionally used to provide unique graph covers, are used herein to provide a unique covering of the changes in the stochastic phenomenon. The main idea is that if the stochastic properties of certain features change, then the changes can be detected by examining the covering set of an identifying code using distributed sequential change point detection. In other words, if the statistical properties of certain stochastic phenomena change,

then the changes can be detected by examining a small subset of measurements that form a covering set for this phenomenon. This approach integrates ideas from the theory of identifying codes [19] with change point detection in sequential analysis [20, 7, 6].

Contribution of Thesis

In this thesis we introduce a new framework, which we term compressed change detection [21]. Specifically, we show that the aforementioned sparsity and locality in the change process can be leveraged to compress the change detection so that sparse statistical changes can be identified and localized using a small number of measurements. The main approach relies on integrating ideas from the theory of identifying codes [19] with change point detection in sequential analysis.

Organization of Thesis

The rest of the thesis is organized as follows. In Chapter 2, the basic problem setup is introduced and some preliminary background about identifying codes over graphs and change point detection is provided. In Chapter 3, we present the proposed approach, which is based on integrating coding over graphs with sequential change detection. A case study for a 4-span bridge model is presented in Chapter 4. An asymptotic performance analysis of the proposed approach is presented in Chapter 5 and numerical results are presented in Chapter 6. We discuss several generalizations in Chapter 7 and conclude in Chapter 8.

CHAPTER 2: PROBLEM SETUP

Suppose there are N random features, $X_1(n), X_2(n), \ldots, X_N(n)$, where $n = 0, 1, \ldots$ denotes a discrete time index. Each feature X_i is independent and identically distributed (i.i.d) across time. Furthermore, the features are statistically independent. Decision makers do not have direct access to features but instead observe measurements $Y_1(n), Y_2(n), \ldots, Y_M(n)$. Each measurement is affected by a subset of features, i.e.,

$$Y_i(n) = g_i(\mathbf{X}_{S_i}(n)) \text{ for } i = 1, \ldots, M \tag{2.1}$$

where $\mathbf{X}_{S_i}(n) = \{X_{s_1}(n), X_{s_2}(n), \ldots, X_{s_k}(n)\} \ \forall s_j \in S_i$ denotes a set of features affecting the ith measurement and g_i denotes an arbitrary function. At an unknown point in time λ, statistical properties of an unknown subset of features indexed by S changes. The change point is assumed to be random with prior distribution π, i.e., $\pi_k = P(\lambda = k), \ k = 0, 1, \cdots$. Let X_S denote the set of features indexed by the set S, i.e., $X_S = \{X_j\}_{j \in S}$. By assumption, features are independent and identically distributed with a pre-change probability density function $f_{X_j}(x) = f_0(x), j = 1, \ldots, N$. Conditioned on the change point, $X_j(n), j \in S$ are i.i.d after the change point $n > \lambda$ with post-change distribution $f_1(x)$ for the set of features in S where f_0 and f_1 are two distinct probability density functions. The cardinality $|S|$ of the set S is assumed to be less than or equal to K and it is assumed that $K << N$, hence the sparsity of the change process. The goal is twofold, namely, we would like to identify the set S while minimizing the detection delay, and to understand the role of sparsity in reducing the number of measurements needed to identify S.

Background on Identifying Codes

Before describing the proposed approach we provide some necessary preliminary background. Let $G = (V, E)$ be a graph with node set V and edge set E. Consider any subset $\mathcal{C} \subseteq V$ and define the identifying set of a vertex $v \in V$ with respect to $\mathcal{C}$ as

$$\mathcal{I}(v, \mathcal{C}) = \mathcal{N}(v) \cap \mathcal{C}, \tag{2.2}$$

where $\mathcal{N}(v) = \{i \in V : e_{iv} \in E\}$, is the neighborhood of vertex $v \in V$. We can readily define an identifying code [19, 17].

Definition 2.0.1. *A collection of vertices $\mathcal{C} \subseteq V$ is called an identifying code if $\forall v_i, v_j \in V$ and $v_i \neq v_j$*

$$\mathcal{I}(v_i, \mathcal{C}) \neq \emptyset, \tag{2.3}$$

$$\mathcal{I}(v_i, \mathcal{C}) \neq \mathcal{I}(v_j, \mathcal{C}). \tag{2.4}$$

Equation (2.3) in Definition 2.0.1 means that the identifying set for each vertex with respect to $\mathcal{C}$ is non-empty, and (2.4) refers to the fact that the identifying sets for different nodes are distinct. As such, it is not hard to see that an identifying code provides a unique covering for every node. This powerful property of identifying codes was exploited in [17] for source localization, in [19] to detect faults in a communication network, and in [22] for identifying spectrum violators in cognitive radio networks. Intuitively, a target or fault can be localized by examining the signature of the fault as revealed through the identifying set.

Finding an optimal (minimum cardinality) identifying code $\mathcal{C}^*$ over a graph is generally NP hard, however, different greedy algorithms were proposed to find irreducible near-optimal identifying

codes (See [23] and references therein). One approach is based on the observation that a graph admits an identifying code if and only if the neighborhoods $\mathcal{N}(v)$ for different nodes $v \in V$ are distinct and non-empty [17]. In such cases, the code can be initialized as the whole set V, then reduced by sequentially removing vertices. In each iteration, one vertex is eliminated based on predetermined order of the vertices in V, and the remaining code is verified. If it satisfies (2.3) and (2.4) then it is an ID-code and the procedure continues. If not, the node is not eliminated and the elimination step is repeated with the subsequent node in the predefined order. The resulting code $\mathcal{C}$ is an irreducible ID-code.

The Generalization is to use K-identifying codes which would provide a unique signature for up to K simultaneous changes. This idea is further discussed in Chapter 7.

Background on Change Detection

In classical change point detection, we observe a sequence of independent and identically distributed (i.i.d) random variables $Y_1, Y_2, \cdots$ with distribution f_0 before an unknown point in time λ, $\lambda \in \{1, 2, \cdots\}$. After the change point λ, observations are still i.i.d but with a different distribution f_1. In a Bayesian setting, the change point is assumed to be random with distribution $\pi_k = P(\lambda = k), k = 0, 1, 2, \cdots$. The goal is to detect the change in distribution as soon as it occurs subject to a false alarm probability constraint [24].

A sequential detection procedure is a stopping time τ on the sequence of observations $\{Y_n\}_{n \geq 1}$ such that for every n, the event $\{\tau \leq n\}$ is measurable, i.e., it is in sigma-algebra generated by $\{Y_i\}_{i=1}^{n}$ which is $(\mathcal{F}_n^Y = \sigma(Y_1, Y_2, \cdots, Y_n))$. A false alarm happens when the stopping time is less than the change point , i.e., $\tau < \lambda$. As defined in [24] a reasonable measure for the detection delay is Average Detection Delay (ADD(τ)) and for false alarm rate is defined as Probability of False

Alarm (PFA(τ)),

$$\text{ADD}(\tau) = \text{E}^{\pi}(\tau - \lambda | \tau \geq \lambda) = \frac{\text{E}^{\pi}((\tau - \lambda)^+)}{\text{P}^{\pi}(\tau \geq \lambda)} \tag{2.5}$$

$$\text{PFA}(\tau) = \text{P}^{\pi}(\tau < \lambda). \tag{2.6}$$

P^{π} is the average probability measure with respect to π and E^{π} is the expectation, respectively. We define $\text{P}^{\pi}(\Omega) = \sum_{k=0}^{\infty} \pi_k \text{P}_k(\Omega)$, where P_k is probability measure when $\lambda = k$ and E_k is the respective expectation.

An optimal change detection procedure minimizes ADD subject to PFA being below a given threshold $\alpha \in (0, 1)$.

The problem can be viewed as a simple hypothesis test, with the following hypotheses,

- $H_0 : \lambda > n$ (i.e. change has not happened before time n.)

- $H_1 : \lambda \leq n$ (i.e. change has happened before or at time n.)

It is easy to show that the likelihood ratio, Λ_n, for this test is as follows for each $\{n \geq 0\}$ [24],

$$\Lambda_n = \frac{\pi_0}{1 - \pi_0} + \frac{1}{\text{P}(\lambda > n)} \sum_{k=1}^{n} \pi_k \prod_{t=k}^{n} \frac{f_1(y_t)}{f_0(y_t)} \tag{2.7}$$

In particular, Shiryaev showed that for the case of change point λ with prior geometric distribution, it is optimal to stop at the first time ν_B that the likelihood ratio Λ_n crosses a threshold B which depends on the threshold α [25],

$$\nu_B = \inf\{n \geq 1 : \Lambda_n \geq B\} \tag{2.8}$$

where threshold $B = B_\alpha$ should be chosen so that $\text{PFA}(\nu_B) = \alpha$. By setting $B_\alpha = \frac{1-\alpha}{\alpha}$, it is not

hard to show that $\nu_B \in \Delta(\alpha) = \{\tau : \text{PFA}(\tau) \leq \alpha\}$ [24]. Shiryaev's procedure is *asymptotically optimal* when $\alpha \to 0$.

CHAPTER 3: METHODOLOGY

In this chapter we describe the methodology of Compressed Change Detection formally. we also provide a simple example with synthetic data.

Detection Graph Mapping

The function g_i in equation (2.1) is typically unknown. The key point is that the change in the statistical properties of any of the features indexed by S_i will result in some change in the statistical behavior of Y_i. In other words, the detection of a change in Y_i can be viewed as a detection of change in a subset of the random features indexed by S_i. Now, the problem can be transformed to a bipartite graph $G = (U, V, E)$ where U denotes the measurements and V denotes the features, i.e., $|U| = M$ and $|V| = N$. The edge set E describes the relation between the measurements and the features.

$$E = \{e_{ij} | \forall j \in S_i\}, i = 1, 2, \ldots, N \tag{3.1}$$

where S_i is the index set related to Y_i. An edge e_{ij} means that measurement i can be affected by feature j. Henceforth, the graph G is called *Detection Graph*. A typical detection graph is depicted in Figure 3.1 (a). This detection graph is the key to our approach. It can be obtained from the system dynamics or by data driven techniques during a training phase.

Change Fingerprints

The second step is the detection of statistical changes to identify the set S. The main idea we are exploring here is that change point detection can be applied to the measurements of a subset

of sensors that define an identifying code over the graph G of sensors/features, thereupon the detection of sparse statistical changes can be identified from a smaller number of measurements. In the spirit of compressive sensing [26, 27], the responses of different sensors are driven by more than one feature. Only measurements from sensors measuring features in the set S will have a change in their statistical distributions, and thus can be detected using techniques from sequential change detection. On account of the sparsity of the change process and the covering property of the identifying code, the detection of change can be efficiently compressed. Thus, the set S can be identified from the change fingerprints. Note that the identifying code together with the sequential change detection will provide a unique covering for any change at any node in V.

More specifically, we wait until the first sensor in the identifying code fires. We then wait for a grace period to allow other sensors in the identifying set of change(s) to fire. Setting the grace period is a crucial design parameter to guarantee that the affected sensors in the identifying set fire with high probability and that no other sensors outside the identifying set gives a false alert. In our algorithm once a sensor fires we consider it as a change sensor until the end. At the end of the grace period, the sensors which have already fired and raised a flag form the change fingerprint. If the change fingerprint matches the actual identifying set for the change we have a true alert.

To illustrate the proposed approach, we consider an example in Figure 3.1 (a) with 7 sensors and 9 features. First, we find an ID-code consisting of nodes 1, 5, 6, and 7. Second, we observe the statistics for Shiryaev change detectors placed at these nodes in Fig. 3.1 (b)-(e). Sensors 1, 6, and 7 detect a change as their statistics exceed a given threshold. The only feature corresponding to this change pattern is feature 8, which is thus declared as the changed feature.

However, if multiple features, say up to K, change simultaneously, then the idea can be extended by using a K-identifying code. Such codes provide *unique* covering for any subset of nodes of cardinality less than or equal to K. Henceforth, and without loss of generality, we focus on a

single change and generalization to multiple changes is further discussed in Chapter 7 using K-identifying codes.
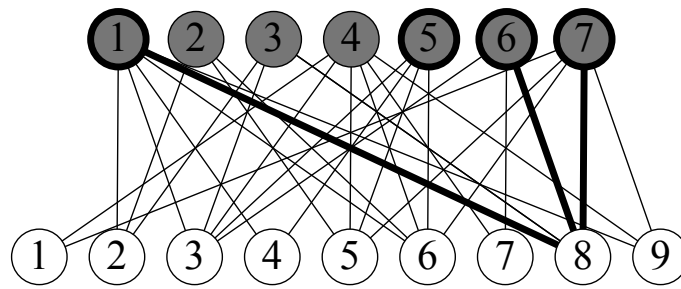
(a) Bipartite graph.
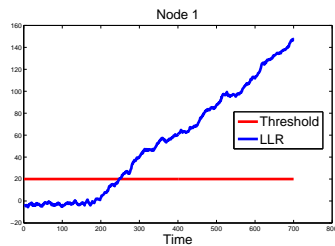
(b) Node 1
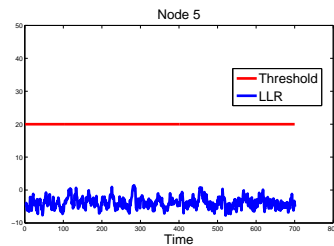
(c) Node 5

(d) Node 6

(e) Node 7

Figure 3.1: Our proposed approach for a set of 9 features (white circles) and 7 sensors (gray circles). The bolded circles denote the ID-code. The logarithm of Shiryaev statistics at sensors 1, 6, and 7 exceed the threshold, thus feature 8 is detected.

13

CHAPTER 4: CASE STUDY

In this chapter, we have M measurement streams[1] $Y_i, i = 1, \ldots, M$. We consider N damage scenarios of concern that could potentially occur on a given structure. It is assumed that each measurement is o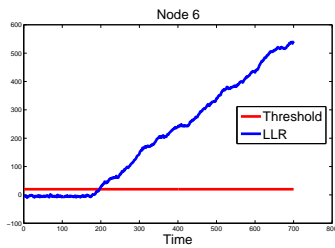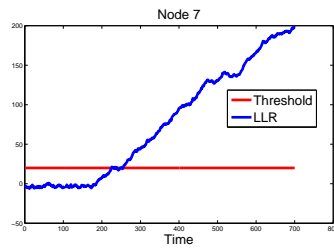nly affected by a subset of damage scenarios. For example, if the structure is equipped with multi-modal sensors (e.g. accelerometers, strain gauges, tiltmeters, etc.), sensors could respond differently to different changes depending on the sensing modality, the sensors' locations and the nature of damage. At some unknown point in time λ, an unknown subset S of damage scenarios occurs; thereafter, the statistics of all measurement streams that respond to the damage scenarios in S change.

The proposed approach consists of the following steps. First, we learn a damage-sensing graph, then we find an identifying code for this graph. Sequential change detection is then applied to the streams corresponding to this code to provide a unique signature for the set of damages S. Since structural changes are typically sparse, we assume that the cardinality of S is upper bounded by K and $K \ll N$. The goal is to identify S as soon as damages occur. In the following, we provide details for each of these steps.

Learning damage-sensing Graph

In the first step, we learn a bipartite graph $G = (V, E)$ with vertex set $V = A \cup B$ and edge set E. The set A corresponds to the streams of measurement and B to the damage scenarios. An edge $e_{ij} \in E$ designates that the jth damage affects the ith measurement. Hereafter, we refer to the graph G as the damage-sensing graph, since it specifies the sensitivity of different measurement

---

[1]These streams do not have to be raw measurements from individual sensors. For example, in our case study these streams correspond to measurements of pairs of sensors deployed on a structure.

streams to the various damage scenarios.

To decide if edge e_{ij} is in E, we apply a change-point detection algorithm to a sequence of measurements formed from nominal data (under no-damage) followed by measurements from the ith stream under damage j. If the change-point is detected, we conclude that the jth damage is detected through measurement stream i, and hence $e_{ij} \in E$. This procedure is repeated for all streams and all damage scenarios. Therefore, we need access to measurements under no-damage, as well as separate measurements for the remaining damage scenarios. Such measurements could be typically obtained from models of the monitored structures. In this chapter, we obtain these measurements from a finite element model (FEM) of a 4-span laboratory bridge model.

While various change point detection algorithms can be employed [10, 6], we use Moving Cross-Correlation Analysis (MCCA) [15, 28], since it is model-free which makes it suitable for our purpose. We briefly outline the steps of MCCA next.

Background on MCCA

This algorithm relies on the cross correlation between measurements from pairs of sensors over a moving window. In particular, let $y_i(t)$ and $y_j(t)$ denote the measurements of sensors i and j, respectively, at time t. The Moving Cross-Correlation (MCC) at time t, $\rho_{ij}(t)$, is defined as,

$$\rho_{ij}(t) = \frac{\sum_{k=t-T+1}^{t}(y_i(k) - \mu_i(t))(y_j(k) - \mu_j(t))}{\sqrt{\left(\sum_{k=t-T+1}^{t}(y_i(k) - \mu_i(t))^2\right)\left(\sum_{k=t-T+1}^{t}(y_j(k) - \mu_j(t))^2\right)}} \tag{4.1}$$

for $t \geq T$, where T is the size of a moving window. $\mu_i(t)$ and $\mu_j(t)$ are the moving averages for sensor measurements i and j over the window at time t. To calculate $\mu_s(t)$ for a sensor s we take

15

the sample average over the window,

$$\mu_s(t) = \frac{1}{T} \sum_{k=t-T+1}^{t} y_s(k). \tag{4.2}$$

The analysis consists of a training and a monitoring phase. In the training phase, we use the baseline data to find the average and variance for the MCC of pairs of sensors,

$$\mu_{ij} = \frac{1}{N_t - T + 1} \sum_{k=T}^{N_t} \rho_{ij}(k) \tag{4.3}$$

$$\sigma_{ij} = \sqrt{\frac{\sum_{k=T}^{N_t} (\rho_{ij}(k) - \mu_{ij})^2}{N_t - T}}, \tag{4.4}$$

where N_t is the length of the training data, μ_{ij} and σ_{ij} are the temporal average and variance of the MCC, $\rho_{ij}(t)$, between sensors i and j computed in (4.1). In the monitoring phase, a damage is declared whenever the MCC between two sensors deviates sufficiently from the average, i.e.,

$$|\rho_{ij}(t) - \mu_{ij}| > \gamma \cdot \sigma_{ij}, \ t > N_t \tag{4.5}$$

where γ is a tuning parameter.

This analysis is carried out for all pairs of sensors. In other words, every pair of sensors corresponds to a data stream in the learned damage-sensing graph.

Once a code $\mathcal{C}$ is found, we perform change point detection for the streams corresponding to the nodes in $\mathcal{C}$. In the monitoring phase, the damage is identified by comparing the pattern of the damage (damage fingerprint) to the identifying set of the different combinations of damage scenarios. This approach can significantly compress the process of damage detection since we only need to monitor a suitable covering set for the damage-sensing graph.

16

We consider a 4-span laboratory bridge model equipped with 9 strain gauges. These gauges are Fiber Bragg Grating (FBG) sensors that measure strain by means of the proportional shift in reflected wavelength under strain. It is feasible to simulate and test a variety of global and local damage scenarios that are commonly observed in bridge-type structures. An example of global damage is alteration in boundary conditions, including roller, pin, and fixed support as shown in Fig. 4.2(a) (Left). In addition, the bolts connecting the girders and deck can be loosened or removed at different locations to modify the stiffness of the system and simulate local damage as shown in Fig. 4.2(b) (Left). We identify 9 potential damage scenarios for the structure. To simplify the measurement process, we obtained a large number of measurements from an FEM of the described bridge model. This FEM was shown to faithfully represent the lab model [29]. A traffic load moving across the bridge is simulated and the strain at different sensors is recorded. Sample strain data is shown in Fig. 4.1 for the baseline condition (no damage), as well as different damage scenarios. For change detection, we use the MCCA method described earlier in Chapter 4. Fig. 4.2 shows the observed change in the cross correlation coefficient under global and local damage scenarios. In particular, in Fig. 4.2(a) (Left) a roller support is altered to a fixed support and in Fig. 4.2(b) (Left) few bolts are removed from the first span. These observed changes are then used to construct the damage-sensing graph based on the sensitivity of different streams to the different damages. The resulting damage-sensing graph consists of 36 streams corresponding to all pairs of sensors (set A) and 9 damage scenarios (set B). An ID-code is obtained using a greedy algorithm and is illustrated in Table 4 along with the identifying set for each damage scenario. For example, damage scenario 8 is only sensed by the pairs $(1, 8)$ and $(2, 4)$. By monitoring a small number of streams (6 streams), we are able to provide a unique covering for any of these damages over the bridge. Furthermore, it is observed that there is no need to use sensors 7 or 9 since they are not present in the pairs of sensors defining the ID-code. As such, this approach can be potentially used to optimize the sensor deployment.

An example of damage detection is shown in Fig. 4.3 where damage scenario 1 occurs at time 522. The algorithm correctly identifies the damage fingerprint $\{(1,4),(1,8),(2,4),(3,6)\}$, which is the unique signature for damage scenario 1, with a small detection delay. In this example we have fixed the tuning parameter γ, however, since the variance for different streams is different, a different threshold is used for the change statistic in (4.5).
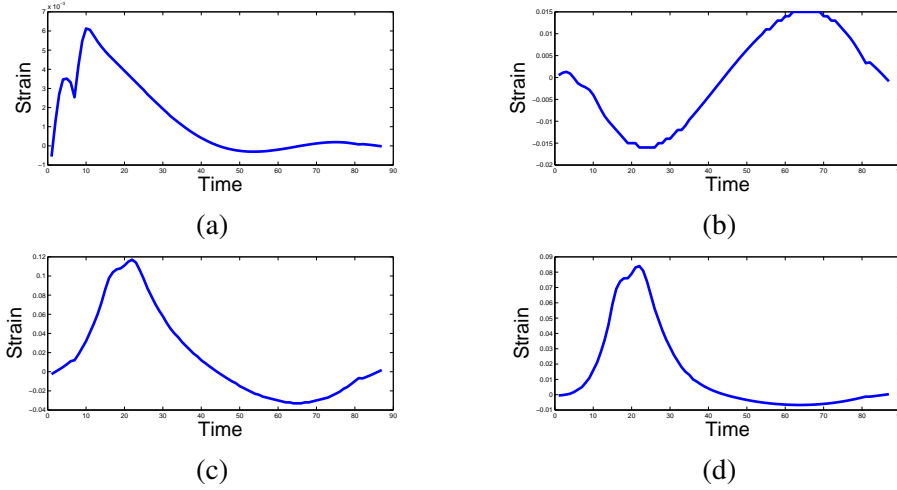
Figure 4.1: Typical Responses of strain gauges. (a) sensor 1 under baseline condition, (b) sensor 1 under damage scenario 1, (c) sensor 2 under baseline condition, (d) sensor 2 under damage scenario 2.

Table 4.1: ID-code and identifying sets for different damage scenarios. Having a 1 in the ijth entry designates that measurement j is in the identifying set of damage scenario i.

	(1,4)	(1,8)	(2,4)	(3,5)	(3,6)	(6,8)
Scenario 1	1	1	1	0	1	0
Scenario 2	1	0	0	0	0	0
Scenario 3	0	0	1	0	0	0
Scenario 4	1	1	1	1	1	1
Scenario 5	0	1	0	0	0	0
Scenario 6	0	1	1	1	1	1
Scenario 7	0	1	1	0	1	0
Scenario 8	0	1	1	0	0	0
Scenario 9	0	1	1	1	1	0

Figure 4.2: Change in MCC for (a) damage scenario 1, fixing the first boundary condition and (b) damage scenario 4, removing bolts from the first span.

Figure 4.3: Damage fingerprint. The revealed pattern consists of pairs $\{(1,4),(1,8),(2,4),(3,6)\}$, which corresponds to the identifying set of damage scenario 1.

CHAPTER 5: ASYMPTOTIC ANALYSIS

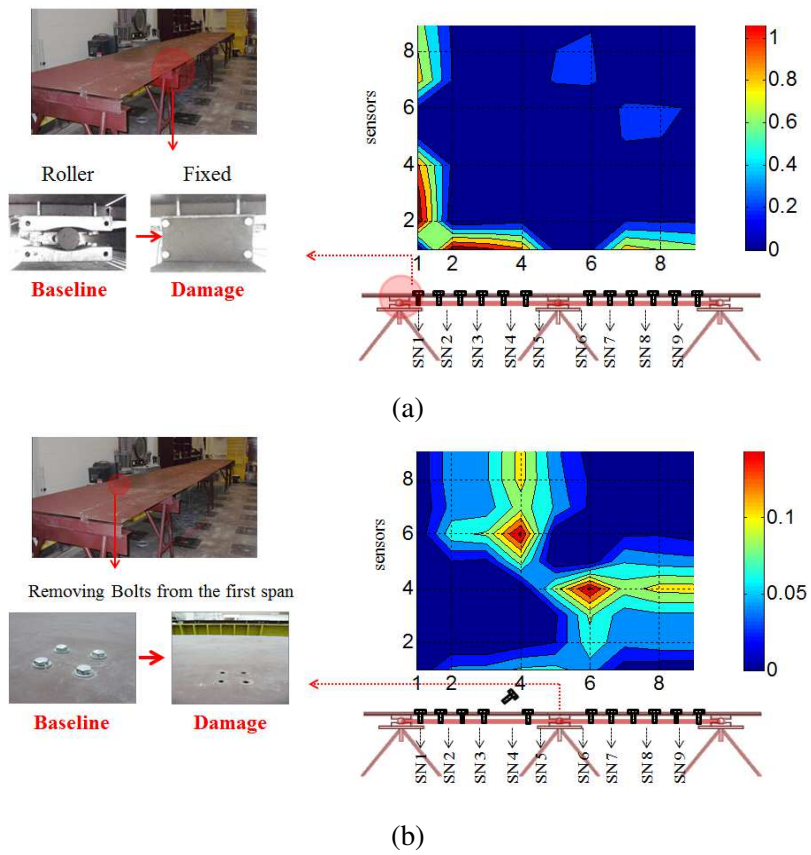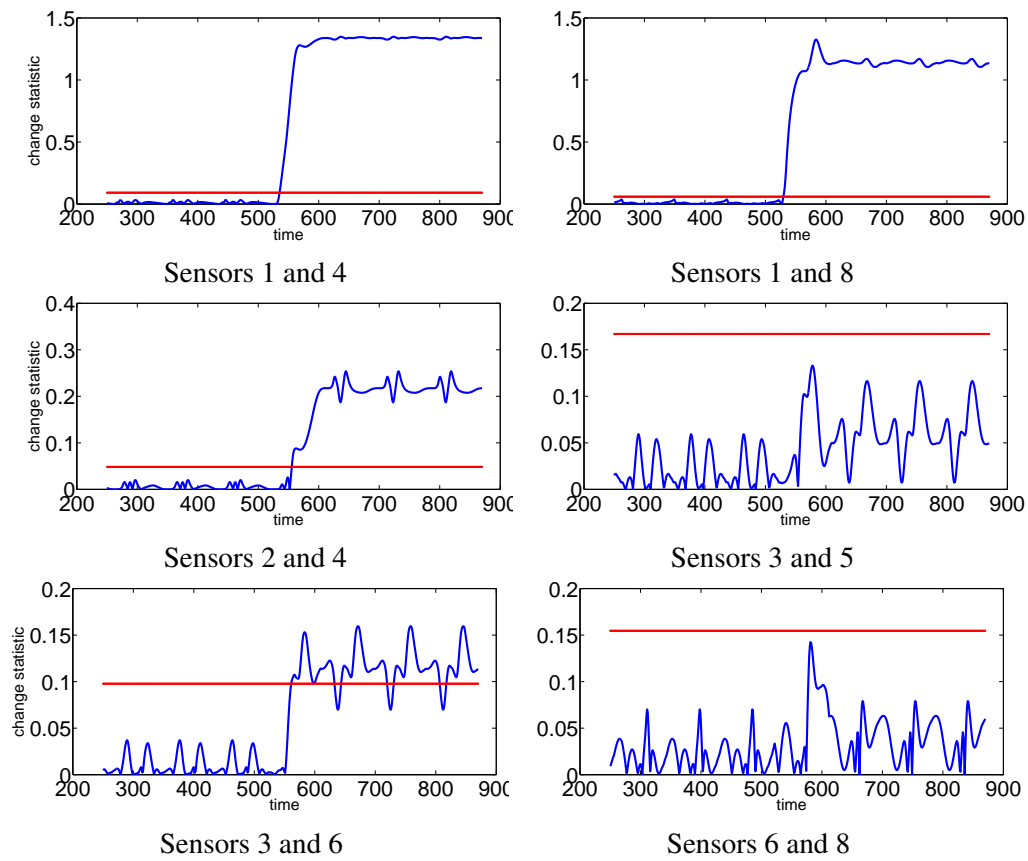In this chapter, we analyze the performance of the proposed approach in the asymptotic regime where the number of features $N \to \infty$. This analysis gives insight into the scaling of the detection delay as a function of the size of the graph. In our analysis we assume that G is an Erdös-Rényi bipartite graph with probability of edge placement p. In this section we further assume that the random features $X_1, X_2, \ldots, X_N$ have a pre-change distribution $Normal(0, 1/(Np))$. After the change point, the affected feature will have the post-change distribution $Normal(1, 1/(Np))$. This setup corresponds to a fixed SNR regime for an additive sensing model, i.e., when g_i in (2.1) is simply a summation. We define a random time as follows,

Definition 5.0.2. *We Consider N_0 as the first time that the Shiryaev statistic of a sensor in the identifying code crosses the threshold B and T_g as the grace period. Then,*

$$\tau = N_0 + T_g \tag{5.1}$$

is our random time.

In other words, we wait for a duration T_g – known as grace period – after the first time a sensor in the identifying raises a flag, then we stop. This duration is a design parameter that should be long enough to allow for the affected sensors to declare the change, but not too long to alleviate false alarms from those sensors unaffected by the change.

Theorem 5.0.1. *The random time τ defined in Definition 5.0.2 is a stopping time.*

Proof. WLOG assume that the identifying code is the whole sequence of nodes representing the measurement side. Define, $\mathcal{F}_n = \sigma(Y_1(1), ..., Y_1(n), Y_2(1), ..., Y_2(n), Y_M(1), ..., Y_M(n))$, that is

the σ-algebra generated by measurements up to time n. Consider the event $\{\tau = n\}$. If $n \leq T_g$ then $\{\tau = n\} = \emptyset \in \mathcal{F}_n$. Now if $n > T_g$,

$$
\begin{aligned}
\{\tau = n\} &= \{N_0 + T_g = n\} & \text{(5.2)} \\
&= \{\inf_{u \in [1,...,M]} \{T^u(B) + T_g\} = n\} & \text{(5.3)} \\
&= \{\inf_{u \in [1,...,M]} \{T^u(B)\} = n - T_g\} & \text{(5.4)} \\
&= \bigcup_{u=1}^{M} \{T^u(B) = n - T_g\} \in \mathcal{F}_{n-T_g} \subseteq \mathcal{F}_n. & \text{(5.5)}
\end{aligned}
$$

where $T^u(B)$ denotes the Shiryaev stopping time of node u with threshold B. $\square$

The Probability of false alarm and isolation of a change at node v on the feature side X is defined as,

$$
PFAI(v) := P^\pi(N_0 \leq \lambda) + P(d \neq d_{G^{-1}}(v)) \tag{5.6}
$$

where d is change fingerprint at the time of detection and $d_{G^{-1}}(v)$ is the identifying set of the changed node v. Since the change happens uniformly at random at the nodes, the total probability of false alarm and isolation is,

$$
PFAI := \sum_{i=1}^{N} P(v = i) PFAI(i) = \frac{1}{N} \sum_{i=1}^{N} PFAI(i) \tag{5.7}
$$

By symmetry, $PFAI(i)$ for all $1 \leq i \leq N$ are equal. Therefore, $PFAI = PFAI(1)$ for instance.

Theorem 5.0.2. *For an Erdös-Rényi bipartite graph, when a change occurs at one uniformly random chosen feature with pre-change distribution $f_0 \sim \mathcal{N}(0, \frac{1}{Np})$ and post-change distribution $f_1 \sim \mathcal{N}(1, \frac{1}{Np})$, the probability of false alarm and isolation(PFAI) goes to zero asymptotically if*

- $\alpha_{typ} \in O(\frac{1}{\log^3 N})$, *where α_{typ} is the Probability of False Alarm (PFA) of an individual change*

detector in a typical graph (which will be defined in the sequel).

- *The grace period $T_g \in O(\log^{1.1} N)$.*

- *Change point has a prior geometric distribution with parameter $\rho \in O(\frac{1}{T_g})$*

Under these conditions the average detection delay would be in the order of $O(\log^{1.1} N)$.

To prove Theorem 5.0.2 we first establish the following lemmas.

Lemma 5.0.1. *Given bipartite graphs $G(U \cup V, p)$ with $|V| = N$ and edge placement probability p, the probability $\Pr(\mathcal{C} \text{ not a K-ID code})$ that a code $\mathcal{C} = U$ of size $|\mathcal{C}|$ is not an K-ID code is upper bounded by $N^{2K}(1 - \min\{p, 2p(1-p)\}(1-p)^{K-1})^{|\mathcal{C}|}$.*

Proof. Similar to the proof of [23] but for a bipartite graph, consider $\mathcal{M}$ as all maximal pairs (X, Y) in $\{X \subseteq V, Y \subseteq V, |X| \leq K, |Y| \leq K\}$.

$$Pr\{\mathcal{C} \text{ is not } K\text{-ID code}\} \leq \sum_{(X,Y)\in\mathcal{M}} Pr\{\mathcal{I}(X,\mathcal{C}) = \mathcal{I}(Y,\mathcal{C})\} \tag{5.8}$$

$$\leq \sum_{(X,Y)\in\mathcal{M}} \prod_{z\in\mathcal{C}} Pr[\{z \in \mathcal{I}(X,\mathcal{C}) \cap \mathcal{I}(Y,\mathcal{C})\}$$

$$\cup \{z \notin \mathcal{I}(X,\mathcal{C}) \cup \mathcal{I}(Y,\mathcal{C})\}] \tag{5.9}$$

$$\leq \sum_{(X,Y)\in\mathcal{M}} \prod_{z\in\mathcal{C}} Pr[\{z \in \mathcal{I}(X \cap Y,\mathcal{C})\}$$

$$\cup \{z \in \mathcal{I}(X \backslash Y,\mathcal{C}) \text{ AND } z \in \mathcal{I}(Y \backslash X,\mathcal{C})\} \text{ AND } z \notin \mathcal{I}(X \cap Y,\mathcal{C})$$

$$\cup \{z \notin \mathcal{I}(X \cup Y,\mathcal{C})\}] \tag{5.10}$$

Without loss of generality assume $|X| \leq |Y|$. If $X \subset Y$ because (X, Y) is maximal then $|X| =$

$K - 1$ and $|Y| = K$.

$$Pr[\{z \in \mathcal{I}(X,\mathcal{C}) \cap \mathcal{I}(Y,\mathcal{C})\} \cup \{z \notin \mathcal{I}(X,\mathcal{C}) \cup \mathcal{I}(Y,\mathcal{C})\}]$$

$$\leq 1 - (1-p)^{K-1} + (1-p)^K \tag{5.11}$$

$$= 1 - p(1-p)^{K-1} \tag{5.12}$$

If $X \not\subset Y$ we have $|X| = K$ and $|Y| = K$ due to the fact that the pair are maximal and then it is easy to show that the upper bound is maximized when $|X \cap Y| = K - 1$

$$Pr[\{z \in \mathcal{I}(X,\mathcal{C}) \cap \mathcal{I}(Y,\mathcal{C})\} \cup \{z \notin \mathcal{I}(X,\mathcal{C}) \cup \mathcal{I}(Y,\mathcal{C})\}]$$

$$\leq 1 - (1-p)^{K-1} + (1-p)^{K-1}(1 - (1-p))^2 + (1-p)^{K+1} \tag{5.13}$$

$$= 1 - 2p(1-p)^K \tag{5.14}$$

Therefore,

$$Pr\{\mathcal{C} \text{ is not } K\text{-ID code}\} \tag{5.15}$$

$$\leq N^{2K}(\max\{1 - p(1-p)^{K-1}, 1 - 2p(1-p)^K\})^{|\mathcal{C}|} \tag{5.16}$$

$$\leq N^{2K}(1 - \min\{p, 2p(1-p)\}(1-p)^{K-1})^{|\mathcal{C}|} \tag{5.17}$$

$\square$

For the case of 1-ID code or simply ID-code we have $Pr(\mathcal{C} = U.$ is not an ID-code$) \leq N^2(1 - min\{p, 2p(1-p)\})^{|\mathcal{C}|}$.

Corollary 5.0.1. *The probability that a randomly generated code $\mathcal{C}$ is an ID-code goes asymptoti-*

cally to 1 *for some* $|\mathcal{C}| = O(\log N)$, *i.e.,* $\exists$ *constant* $C_1 > 0$ *such that:*

$$P\{\mathcal{C} \text{ is an ID-code}\} \to 1 \ \text{ if } |\mathcal{C}| \geq C_1 \log N. \tag{5.18}$$

Now, we establish another lemma.

Lemma 5.0.2. *If the process* $\{Z_n\}_{n=1}^{\infty}$ *is independent of the past, that is* Z_n *is independent of* $\mathcal{F}_{n-1}^Z = \sigma(Z_1, Z_2, \cdots, Z_{n-1})$, *and finite dimensional distribution is stationary, that is given* $m, n \in \mathbb{N}$ *and* $B_1, \cdots, B_n \in \mathcal{B}_{\mathbf{R}}$, *where* $\mathcal{B}_{\mathbf{R}}$ *is the Borel sigma algebra on real line,*

$$P(Z_1 \in B_1, \cdots, Z_n \in B_n) = P(Z_{m+1} \in B_1, \cdots, Z_{m+n} \in B_n) \tag{5.19}$$

Then, for stopping time N,

$$P(Z_1 \in B_1, \cdots, Z_n \in B_n) = P(Z_{N+1} \in B_1, \cdots, Z_{N+n} \in B_n | N < \infty) \tag{5.20}$$

Proof. Consider event $\mathcal{A} \in \mathcal{F}_N$, where $\mathcal{F}_N$ is the sigma algebra generated by stopping time N. Then,

$$P(\mathcal{A}, N < \infty, Z_{N+n} \in B_n, Z_{N+n-1} \in B_{n-1}, \cdots, Z_{N+1} \in B_1) \tag{5.21}$$

$$= \sum_{m=1}^{\infty} P(\mathcal{A}, N = m, Z_{m+n} \in B_n, Z_{m+n-1} \in B_{n-1}, \cdots, Z_{m+1} \in B_1) \tag{5.22}$$

$$= \sum_{m=1}^{\infty} P(\mathcal{A}, N = m) P(Z_{m+n} \in B_n, Z_{m+n-1} \in B_{n-1}, \cdots, Z_{m+1} \in B_1) \tag{5.23}$$

$$= \sum_{m=1}^{\infty} P(\mathcal{A}, N = m) P(Z_n \in B_n, Z_{n-1} \in B_{n-1}, \cdots, Z_1 \in B_1) \tag{5.24}$$

$$= P(\mathcal{A}, N < \infty) P(Z_n \in B_n, Z_{n-1} \in B_{n-1}, \cdots, Z_1 \in B_1) \tag{5.25}$$

26

By placing $\mathcal{A} = \Omega$, where Ω is the sample space, then,

$$P(Z_{N+1} \in B_1, \cdots, Z_{N+m \in B_m} | N < \infty) \tag{5.26}$$

$$= \frac{P(\Omega, N < \infty) P(Z_n \in B_n, Z_{n-1} \in B_{n-1}, \cdots, Z_1 \in B_1)}{P(N < \infty)} \tag{5.27}$$

$$= P(Z_1 \in B_1, \cdots, Z_n \in B_n). \tag{5.28}$$

$\square$

As a corollary of lemma 5.0.2, we have,

Corollary 5.0.2. *For positive constants C_1, C_2 and arbitrary density functions $f_0(x)$ and $f_1(x)$ we have*

$$P(C_1 \sum_{k=N}^{n} \Pi_{t=k}^{n} \frac{f_1(Y_t)}{f_0(Y_t)} > C_2 | N < \infty) = P(C_1 \sum_{k=1}^{n-N+1} \Pi_{t=k}^{n-N+1} \frac{f_1(Y_t)}{f_0(Y_t)} > C_2) \tag{5.29}$$

Before we write the proof, we define a typical graph. First, we define an indicator function of measurement i as $Z^i(N) := (z_1^i, z_2^i, \cdots, z_N^i)$ where z_j^i is 1 if measurement Y_i is connected to X_j. z_j^i are i.i.d Bernoulli random variables with parameter p. Then the typical graph G_{typ}^{ε} is the graph with indicator functions $Z^i(N)$ such that

$$2^{-N(H(p)+\varepsilon)} \le P(Z^i(N)) \le 2^{-N(H(p)-\varepsilon)}, 1 \le i \le M \tag{5.30}$$

where $H(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy [30].

Lemma 5.0.3. *For the typical graph G_{typ}^{ε}, if $\varepsilon \in o(\frac{1}{N})$ the probability of each event $(z_1^i, z_2^i, \cdots, z_N^i)$ goes to $2^{-NH(p)}$ asymptotically.*

27

Proof. $P(z_1^i, z_2^i, \cdots, z_N^i) = \lim_{N \to \infty} 2^{-N(H(p)+\varepsilon)} = \lim_{N \to \infty} 2^{-N(H(p)-\varepsilon)} = 2^{-NH(p)}$ □

Corollary 5.0.3. *As $N \to \infty$, the number of features measured by sensors in typical graph goes to Np.*

Lemma 5.0.4. *The probability that a graph is typical goes to 1 for sufficiently large N.*

Proof. By the weak law of large numbers,

$$\frac{-1}{n} \log P(Z^i(N)) \to -E[\log P(z_1^i)] = H(p) \text{(in probability)} \tag{5.31}$$

Therefore, given $\varepsilon > 0$, there exists N_i, such that for all $N > N_i$,

$$Pr(|\frac{-1}{n} \log P(Z^i(N)) - H(p)| > \varepsilon) < \frac{\varepsilon}{M}, 1 \leq i \leq M \tag{5.32}$$

By choosing $N > \max(N_1, \cdots, N_M)$ the probability of union of the sets in (5.32) must be less than ε. □

Now we can readily write the proof of Theorem 5.0.2.

Proof. The proof consists of three parts. First we establish an upper bound on the probability of false alarm ($P^\pi(N_0 \leq \lambda)$) and show that it goes to zero asymptotically. Then we find an upper bound on the probability of false isolation ($P(d \neq d_{G^{-1}}(v))$) in two parts, namely, by bounding the probability of false alarm of unaffected nodes and the probability of missed detection of affected nodes. We show that those upper bounds go to zero asymptotically as well.

Now we bound the probability that the first sensor fires before the change point.

$$P^\pi(N_0 \le \lambda) \le P^\pi(N_0 \le \lambda \,|\, \mathcal{C} \text{ is ID-code AND } G \text{ is typical})$$

$$+ P^\pi(\mathcal{C} \text{ is not ID-code }) + P^\pi(G \text{ is not typical}) \tag{5.33}$$

$$\le \sum_{i=1}^{|\mathcal{C}|} P^\pi(T^i \le \lambda \,|\, \mathcal{C} \text{ is ID-code AND } G \text{ is typical})$$

$$+ P^\pi(\mathcal{C} \text{ is not ID-code }) + P^\pi(G \text{ is not typical}) \tag{5.34}$$

$$\le \sum_{i=1}^{|\mathcal{C}|} P^\pi(T^{typ} \le \lambda) + P^\pi(\mathcal{C} \text{ is not ID-code }) + P^\pi(G \text{ is not typical})$$

$$= |\mathcal{C}|\alpha_{typ} + P(\mathcal{C} \text{ is not ID-code }) + P(G \text{ is not typical}) \tag{5.35}$$

where, T^i is the individual Shiryaev stopping time for sensor i. α_{typ} is the probability of false alarm (PFA) for a sensor in a typical graph. In equation (5.35), since the event $\{\lambda = k\}$ is independent of the event $\{\mathcal{C} \text{ is not ID-code}\}$ and $\{G \text{ is not typical}\}$ for all k, we can replace the probability measure P^π with P. The upper bound we obtain goes to zero as $N \to \infty$ by the assumptions of Theorem 5.0.2, Corollary 5.0.1, and Lemma 5.0.4.

Next, we analyze the probability of false isolation. The event $\{d \ne d_{G^{-1}(v)}\}$ occurs when either

- $\{$at least one sensor in $d_{G^{-1}}(v)$ does not fire before the stopping time $\tau\} := \mathcal{E}$,

- or $\{$at least one sensor *not* in $d_{G^{-1}}(v)$ falsely fires before the stopping time $\tau\} := \mathcal{F}$.

Now we want to show that $P^\pi(\mathcal{E}) \to 0$ asymptotically. Henceforth, all probabilities are conditioned on the event $\{N < \infty\}$.

$$P^\pi(\mathcal{E}) \le P^\pi(\mathcal{E}|N_0 > \lambda \text{ AND } \mathcal{C} \text{ is ID-code AND } G \text{ is typical})$$

$$+ P^\pi(N_0 \le \lambda) + P^\pi(\mathcal{C} \text{ is not ID-code}) + P(G \text{ is not typical}) \tag{5.36}$$

By the assumptions in the statement of Theorem 5.0.2, corollary 5.0.1, and the result of previous part the last three terms on the RHS go to zero asymptotically. It remains to show that $P^\pi(\mathcal{E}|N_0 > \lambda \ AND \ \mathcal{C}$ is ID-code AND G is typical$)$ goes to zero asymptotically.

$$P^\pi(\mathcal{E}|N_0 > \lambda \text{ AND } \mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.37}$$

$$\leq \sum_{i \in d_{G-1}} P^\pi(T^i > N_0 + T_g | N_0 > \lambda \text{ AND } \mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.38}$$

$$\leq \sum_{i \in d_{G-1}} P^\pi(T^i > \lambda + T_g | \mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.39}$$

$$\leq |\mathcal{C}| P^\pi(T^{typ} - \lambda > T_g) \tag{5.40}$$

$$\leq |\mathcal{C}| \frac{E^\pi[T^{typ} - \lambda]}{T_g} \tag{5.41}$$

$$\leq |\mathcal{C}| \frac{E^\pi[(T^{typ} - \lambda)^+]}{P^\pi(T^{typ} \geq \lambda) T_g} \tag{5.42}$$

$$\leq |\mathcal{C}| \frac{ADD^{typ}}{T_g} \tag{5.43}$$

$$\sim \frac{|\mathcal{C}||\log \alpha^{typ}|}{T_g(\frac{1}{2} + |\log(1 - \rho)|)} \tag{5.44}$$

In inequality (5.40), T^{typ} is associated with a typical measurement in the typical graph G. Inequality (5.42) is attained using Chebyshév's inequality. Equation (5.44) is a result of Theorem 4 in [24]. By the assumptions of the Theorem, equality (5.44) goes to zero asymptotically.

For bounding the probability of the event $\mathcal{F}$, note that $P^\pi(\mathcal{F}) = P_\infty(\mathcal{F})$ since the change never affects the corresponding set of nodes and P_∞ is the corresponding probability measure. In other terms, all data for these sensors are coming from the $f_0(x)$ distribution.

$$P^{\pi}(\mathcal{F}) = P_{\infty}(\mathcal{F}) \tag{5.45}$$

$$\leq P_{\infty}(\mathcal{F}|\mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.46}$$

$$+ P(\mathcal{C} \text{ is not ID-code}) + P(G \text{ is not typical}) \tag{5.47}$$

The last two terms on the right hand side of inequality (5.47) go to zero by Corollary 5.0.1 and Lemma 5.0.4. Now it remains to prove that $P_{\infty}(\mathcal{F}|\mathcal{C} \text{ is ID-code AND } G \text{ is typical})$ goes to zero asymptotically.

$$P_{\infty}(\mathcal{F}|\mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.48}$$

$$\leq \sum_{i \notin d_{G^{-1}}(v)} P_{\infty}(N_0 \leq T^i \leq N_0 + T_g|\mathcal{C} \text{ is ID-code AND } G \text{ is typical}) \tag{5.49}$$

$$\leq |\mathcal{C}|P_{\infty}(N_0 \leq T^{typ} \leq N_0 + T_g) \tag{5.50}$$

$$\leq |\mathcal{C}|(P_{\infty}(\max_{N_0 < n \leq N_0 + T_g} \Lambda_n^{typ} > B| \max_{0 \leq n \leq N_0} \Lambda_n^{typ} < \sqrt{B}) + P(\max_{0 \leq n \leq N_0} \Lambda_n^{typ} \geq \sqrt{B})) \tag{5.51}$$

T^{typ} in inequality (5.50) is a typical node among unaffected sensors. It is easy to show that the second term of the right hand side of (5.51) goes to zero asymptotically.

The rest is to show that, $|\mathcal{C}|P_{\infty}(\max_{N_0 < n \leq N_0 + T_g} \Lambda_n^{typ} > B| \max_{0 \leq n \leq N_0} \Lambda_n^{typ} < \sqrt{B})$ goes to zero

asymptotically.

$$|\mathcal{C}|P_\infty(\max_{N_0 < n \le N_0+T_g} \Lambda_n^{typ} > B| \max_{0 \le n \le N_0} \Lambda_n^{typ} \le \sqrt{B})$$

$$\le |\mathcal{C}|P_\infty(\max_{N_0 < n \le N_0+T_g} \frac{P(\lambda > N_0)}{P(\lambda > n)}.\Lambda_{N_0}^{typ}.\Pi_{t=N_0+1}^n \frac{f_1(Y_t)}{f_0(Y_t)}$$

$$+ \frac{1}{P(\lambda > n)} \sum_{k=N_0+1}^n \rho(1-\rho)^{k-1}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)} > B) \qquad (5.52)$$

$$\le |\mathcal{C}|P_\infty(\max_{N_0 < n \le N_0+T_g} \frac{1}{(1-\rho)^{n-N_0}}.\sqrt{B}.\Pi_{t=N_0+1}^n \frac{f_1(Y_t)}{f_0(Y_t)}$$

$$+ \sum_{k=N_0+1}^n \frac{\rho}{(1-\rho)^{n-k+1}}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)} > B) \qquad (5.53)$$

$$\le |\mathcal{C}|P_\infty(\max_{N_0 < n \le N_0+T_g} \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=N_0+1}^n \frac{f_1(Y_t)}{f_0(Y_t)}$$

$$+ \sum_{k=N_0+1}^n \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)} > B) \qquad (5.54)$$

$$\le |\mathcal{C}|P_\infty(\max_{0 < n \le T_g} \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^n \frac{f_1(Y_t)}{f_0(Y_t)}$$

$$+ \sum_{k=1}^n \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)} > B) \qquad (5.55)$$

$$\le |\mathcal{C}|\frac{E_\infty(\frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^n \frac{f_1(Y_t)}{f_0(Y_t)} + \sum_{k=1}^{T_g} \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)})}{B} \qquad (5.56)$$

$$\le \frac{|\mathcal{C}|}{(1-\rho)^{T_g}}(\frac{1}{\sqrt{B}} + \frac{\rho T_g}{(1-\rho)B}) \qquad (5.57)$$

where inequality (5.55) is a result of Lemma (5.0.2) and inequality (5.56) is obtained from Doob's

Maximal inequality for martingales [31], since $Z_n = \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^n \frac{f_1(Y_t)}{f_0(Y_t)} + \sum_{k=1}^n \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^n \frac{f_1(Y_t)}{f_0(Y_t)}$

is a positive sub-martingale with respect to $\mathcal{F}_n^Y$ under probability measure P_∞. To show that, first

it is obvious that Z is adapted to $\mathcal{F}_n^Y$. Secondly, for all $n \in \mathbf{N}$

$$E_\infty |Z_n| = \frac{\sqrt{B}}{(1-\rho)^{T_g}} \prod_{t=1}^{n} E_\infty(\frac{f_1(Y_t)}{f_0(Y_t)}) + \sum_{k=1}^{n} \frac{\rho}{(1-\rho)^{T_g}} \prod_{t=k}^{n} E_\infty(\frac{f_1(Y_t)}{f_0(Y_t)}) \qquad (5.58)$$

$$= \frac{\sqrt{B}}{(1-\rho)^{T_g}} + \frac{\rho.n}{(1-\rho)^{T_g}} < \infty \qquad (5.59)$$

Thirdly, we need to prove that $E_\infty(Z_{n+1}|\mathcal{F}_n^Y) \geq Z_n$, i.e.

$$E_\infty(Z_{n+1}|\mathcal{F}_n^Y) = E_\infty(\frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^{n+1}\frac{f_1(Y_t)}{f_0(Y_t)} + \sum_{k=1}^{n+1} \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^{n+1}\frac{f_1(Y_t)}{f_0(Y_t)}|\mathcal{F}_n^Y) \qquad (5.60)$$

$$= \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^{n}\frac{f_1(Y_t)}{f_0(Y_t)}E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}|\mathcal{F}_n^Y)$$

$$+ E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}|\mathcal{F}_n^Y) \sum_{k=1}^{n} \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^{n}\frac{f_1(Y_t)}{f_0(Y_t)} + \frac{\rho}{(1-\rho)^{T_g+1}}.E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}|\mathcal{F}_n^Y)$$

$$(5.61)$$

$$= \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^{n}\frac{f_1(Y_t)}{f_0(Y_t)}E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})})$$

$$+ E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}) \sum_{k=1}^{n} \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^{n}\frac{f_1(Y_t)}{f_0(Y_t)} + \frac{\rho}{(1-\rho)^{T_g+1}}E_\infty(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})})$$

$$(5.62)$$

$$= \frac{1}{(1-\rho)^{T_g}}.\sqrt{B}.\Pi_{t=1}^{n}\frac{f_1(Y_t)}{f_0(Y_t)} \qquad (5.63)$$

$$+ \sum_{k=1}^{n} \frac{\rho}{(1-\rho)^{T_g+1}}\Pi_{t=k}^{n}\frac{f_1(Y_t)}{f_0(Y_t)} + \frac{\rho}{(1-\rho)^{T_g+1}} \qquad (5.64)$$

$$= Z_n + \frac{\rho}{(1-\rho)^{T_g+1}} > Z_n. \qquad (5.65)$$

The right hand side of the inequality (5.57) goes to zero asymptotically by the assumptions in the

statement of the Theorem. Now for the calculation of the detection delay we have,

$$E^\pi[N_0 + T_g - \lambda | N_0 > \lambda] = T_g + E^\pi[N_0 - \lambda | N_0 > \lambda] \tag{5.66}$$

$$= T_g + E^\pi[\inf_{1 \le i \le |\mathcal{C}|} T^i - \lambda | \inf_{1 \le i \le |\mathcal{C}|} T^i > \lambda] \tag{5.67}$$

$$\le T_g + \frac{E^\pi[(T^{typ} - \lambda)^+]}{P(\inf_{1 \le i \le |\mathcal{C}|} T^i > \lambda)} \le T_g + \frac{E^\pi[(T^{typ} - \lambda)^+]}{P(T^{typ} > \lambda)} \tag{5.68}$$

$$\le T_g + ADD^{typ} \sim T_g + \frac{|\log \alpha_{typ}|}{D(f_1^{typ} || f_0^{typ}) + |\log(1 - \rho)|} \tag{5.69}$$

The order of T_g dominates the second term on (5.69). This completes the proof.

$\square$

CHAPTER 6: NUMERICAL RESULTS

To back up our asymptotic analysis we first compute the length of the ID-code as N increases. In Figure 6.1 the length of the ID-code is plotted vs. $\log N$. We observe that for Erdös-Rényi bipartite random graphs, the length of the ID-code grows logarithmically with the size of the graph, which matches (5.18). This demonstrates that significant savings can be achieved for detecting sparse changes.

Next, we present some numerical results for the considered bipartite graph model. The pre-change distribution is $f_0(X_i) = \mathcal{N}(0, \frac{1}{Np})$. The post-change probability density function is $f_1 = \mathcal{N}(1, \frac{1}{Np})$. The Shiryaev [25] change detection algorithm is used for these simulations. The probability of edge placement is $p = 0.5$ and the Shiryaev threshold is chosen to ensure that $\alpha_{typ} = O(\frac{1}{\log^3 N})$. We also set the grace period T_g to be in $O(\log^{1.1} N)$ and the geometric distribution parameter ρ for the change point to be in $O(\frac{1}{\log^{1.1} N})$. The probability of false alarm and isolation $PFAI$ as a function of the total number of features N is plotted in Fig. 6.2. As shown, $PFAI$ approaches 0 as N increases when α_{typ} decreases as $O(\frac{1}{\log^3 N})$ in agreement with Theorem 5.0.2. Fig. 6.3 shows the scaling of the detection delay with N. The detection delay is shown to scale linearly with $\log^{1.1} N$, which is consistent with the derived upper bound in Theorem 5.0.2.
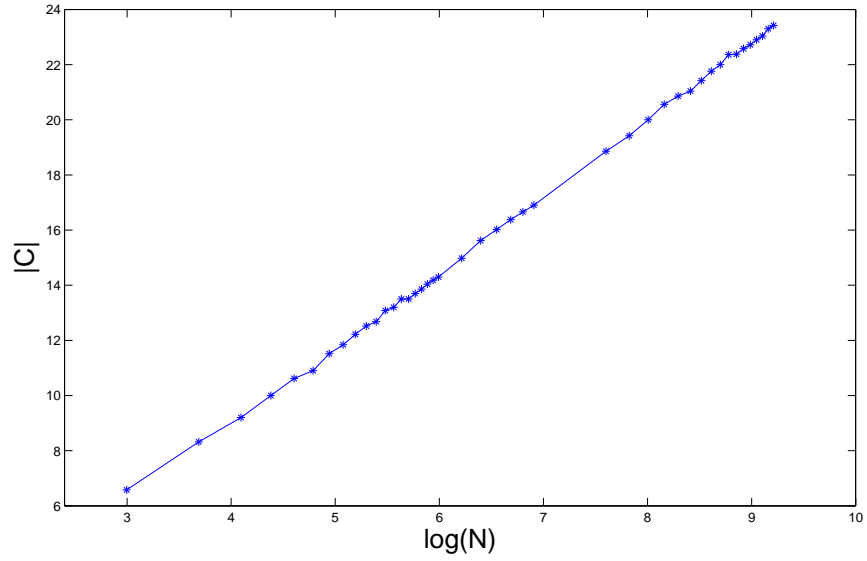
Figure 6.1: Length of the ID-code vs. number of nodes in a random bipartite graph with $p = 0.5$.
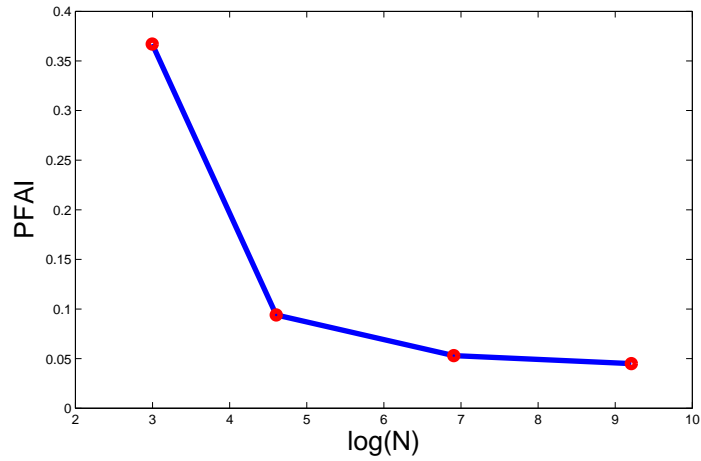

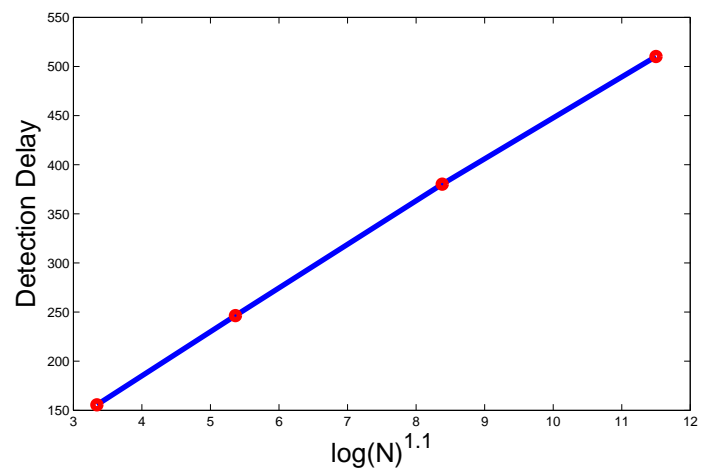Figure 6.2: Probability of false alarm and isolation $PFAI$ goes asymptotically to 0.

Figure 6.3: Detection delay grows almost linearly with $\log^{1.1} N$, which is consistent with the upper bound in Theorem 5.0.2.

CHAPTER 7: GENERALIZATIONS

In this chapter, we briefly discuss generalizations of our approach to different scenarios.

Multiple Concurrent Changes

This approach and principle can be extended to multiple concurrent changes. The idea is to use *K-Identifying codes* or simply *K-ID codes* which give unique signature to every subset of nodes in the graph with cardinality less than or equal to K. In particular, consider a bipartite graph $G = (U, V, E)$ where U and V are two disjoint sets of size m and n. E denotes the edge set and $e_{ij} \in E$ designates that the ith node in U is connected to the jth node in V. Consider any $\mathcal{C} \subseteq U$ and define the identifying set of $X \subseteq V$ with respect to $\mathcal{C}$ as

$$\mathcal{I}(X, \mathcal{C}) = \bigcup_{x \in X} \mathcal{N}(x) \cap \mathcal{C}, \tag{7.1}$$

where $\mathcal{N}(x) = \{i \in U : e_{ix} \in E\}$, is the neighborhood of vertex $x \in V$. We can readily define an K-identifying code [19, 23].

Definition 7.0.3. *A collection of vertices $\mathcal{C} \subseteq U$ is called a K-identifying code over V if $\forall X_i, X_j \subseteq V$, $1 \leq |X_i| \leq K$ and $X_i \neq X_j$*

$$\mathcal{I}(X_i, \mathcal{C}) \neq \emptyset, \tag{7.2}$$

$$\mathcal{I}(X_i, \mathcal{C}) \neq \mathcal{I}(X_j, \mathcal{C}). \tag{7.3}$$

Finding an optimal (minimum cardinality) K-ID code is a NP-hard problem in general. Therefore, some greedy algorithms were proposed to reduce complexity and provide sub-optimal solutions

[18, 17]. We follow the work of Ray et al. in [17] to find a greedy algorithm for determining a K-ID code.

First consider the whole set U as the code. Then check whether the code satisfies (7.2) and (7.3). If not, the graph cannot admit a K-ID code and if yes the algorithm proceeds and removes one vertex according to a predefined sequence of nodes at a time. At each step, if the remaining code does not satisfy (7.2) and (7.3), the deleted vertex is put back in the code, and if it does, we continue until every single node is examined. The remaining code is an irreducible K-ID code, i.e., no subset of the resulting code is a K-identifying code.

The following lemma guarantees that a K-ID code with addition of a vertex or more still remains K-ID code and is similar to lemma 1 in [17] for the case $K = 1$.

Lemma 7.0.5. *If $\mathcal{C}$ is a K-ID code, any $\mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}' \subseteq U$ is a K-ID code.*

Proof. For the sake of contradiction, there should be a $\mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}'$ which is not K-ID code. Consequently, there should be two distinct subsets $X, Y \subseteq U$ with cardinality less than or equal to K such that

$$\mathcal{I}(X, \mathcal{C}') = \mathcal{I}(Y, \mathcal{C}') \tag{7.4}$$

$$\left(\bigcup_{x \in X} \mathcal{N}(x) \cap \mathcal{C}' \right) \cap \mathcal{C} = \left(\bigcup_{y \in Y} \mathcal{N}(y) \cap \mathcal{C}' \right) \cap \mathcal{C} \tag{7.5}$$

$$\bigcup_{x \in X} \left(\mathcal{N}(x) \cap \mathcal{C}' \cap \mathcal{C} \right) = \bigcup_{y \in Y} \left(\mathcal{N}(y) \cap \mathcal{C}' \cap \mathcal{C} \right) \tag{7.6}$$

$$\bigcup_{x \in X} \mathcal{N}(x) \cap \mathcal{C} = \bigcup_{y \in Y} \mathcal{N}(y) \cap \mathcal{C} \tag{7.7}$$

Equation (7.7) contradicts the assumption of $\mathcal{C}$ being a K-ID code. $\square$

The following Theorem provides guarantees for the output of the algorithm to be irreducible and appears in [17] as Theorem 1 for $K = 1$.

Theorem 7.0.3. *The code generated by the algorithm depicted in Figure 7.1 is irreducible.*

Proof. Assume $\mathcal{C}$, the output of algorithm in Figure 7.1 to be reducible. Therefore, there is at least one node $x \in \mathcal{C}$ such that $\mathcal{C} \setminus \{x\}$ is also a K-ID code. Consider iteration i in which the node x is being considered and the corresponding K-ID code is $\mathcal{C}_i$. Then $\mathcal{C}_i \setminus \{x\}$ is not an ID-code, otherwise the node x would have been deleted in iteration i. Since $\mathcal{C} \setminus \{x\} \subseteq \mathcal{C}_i \setminus \{x\}$, by lemma 7.0.5, $\mathcal{C} \setminus \{x\}$ is not a K-ID code which results in contradiction and hence completes the proof. $\square$

<center>Arbitrary Graph Topology</center>

So far, we have only considered Erdös-Rényi bipartite graphs. However, the approach extends beyond bipartite graphs to graphs with arbitrary topology. Such graphs could be used to represent other physical or social phenomena beyond the sensors/features problem we introduced in Chapter 2. For example, the nodes of a given graph can be used in an SHM application to represent a set of critical points on the bridge structure. The connectivity represents the load distribution across the structure. In other words, two points i and j are connected by an edge e_{ij} if a damage induced at node i is sensed by a sensor placed at node j.

Figure 7.2 shows an identifying code for a 20-node graph. To learn this code we used the polynomial-time ID-code algorithm, which can generate irreducible codes for arbitrary topologies. Change detectors deployed at the ID code were shown to detect a change in the distribution at any given node.

procedure GREEDY
    $\mathcal{C} \leftarrow U$
    $S \leftarrow$ predefined sequence of U
    if $\mathcal{C}$ does not satisfy (7.2) and (7.3) **then**
        G does not admit K-ID code
    else
        for i:=1 to m **do**
            $\mathcal{C}_{new} \leftarrow \mathcal{C} \setminus S(i)$
            if $\mathcal{C}_{new}$ satisfy (7.2) and (7.3) **then**
                $\mathcal{C} \leftarrow \mathcal{C}_{new}$
            end if
        end for
    end if
end procedure

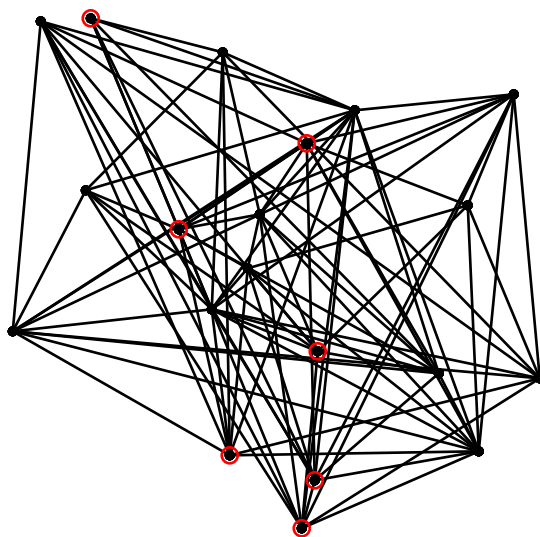Figure 7.1: Greedy algorithm for finding K-ID code.

Figure 7.2: Identifying code for a 20-node graph. The nodes encircled in red form an identifying code.

CHAPTER 8: CONCLUSIONS

In this thesis, we introduced a new framework for compressed change detection. The main approach integrates ideas from the theory of identifying codes over graphs and change point detection in sequential analysis to compress the detection of changes in stochastic phenomena. When the change process itself is sparse, we get a significant compression in the number of measurements since the size of ID code scales only logarithmically with the total number of nodes. We established asymptotic results on the average detection delay for the probability of false alarm and isolation to go to zero. It is shown that the detection delay increases gracefully with the number of damage scenarios. Generalizations to multiple concurrent changes and arbitrary graph topologies were also discussed.

We also leveraged the sparsity and locality of damage occurring in structures to efficiently compress the process of damage detection. This goal is achieved using the newly developed framework for compressed change detection. The effectiveness of the proposed approach is verified through a case study wherein different damage scenarios occurring on a 4-span bridge model equipped with strain gauges are correctly and quickly identified based on the damage fingerprints.

LIST OF REFERENCES

[1] D. Balageas, C. Fritzen, and A. Guemes, *Structural Health Monitoring*, ISTE. Wiley-ISTE, Feb 2006.

[2] N. A. Heard, D. Weston, K. Platanioti, and D. J. Hand, "Bayesian anomaly detection methods for social networks," *Annals of Applied Statistics*, vol. 4, no. 2, pp. 645–662, 2010.

[3] A. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability and Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[4] V.V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. on Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.

[5] K. Premkumar, A. Kumar, and J. Kuri, "Distributed detection and localization of events in large ad hoc wireless sensor networks," in *47th Annual Allerton Conference on Communication, Control, and Computing*, 2009, pp. 178–185.

[6] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[7] V.V. Veeravalli and A.G. Tartakovsky, "Asymptotics of quickest change detection procedures under a Bayesian criterion," in *Proceedings of the IEEE Information Theory Workshop*, 2002, pp. 100–103.

[8] G. Lorden, "Procedures for Reacting to a Change in Distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 18971908, 1971.

[9] Manabu Kano, Shinji Hasebe, Iori Hashimoto, and Hiromu Ohno, "A new multivariate statistical process monitoring method using principal component analysis," *Computers and Chemical Engineering*, vol. 25, no. 78, pp. 1103 – 1113, 2001.

[10] A.G. Tartakovsky and V.V. Veeravalli, "Quickest change detection in distributed sensor systems," in *Proceedings of the Sixth International Conference on Information Fusion*, 2003, vol. 2, pp. 756–763.

[11] "Sequential multi-sensor change-point detection," *Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.

[12] T. Banerjee and V.V. Veeravalli, "Data-efficient quickest change detection in distributed and multi-channel systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*, May 2013, pp. 3952–3956.

[13] Y Lei, AS Kiremidjian, KK Nair, JP Lynch, KH Law, TW Kenny, ED Carryer, and A Kottapalli, "Statistical damage detection using time series analysis on a structural health monitoring benchmark problem," in *Proceedings of the 9th International Conference on Applications of Statistics and Probability in Civil Engineering*, 2003, pp. 6–9.

[14] Mustafa Gul F Necati Catbas, Hasan B Gokce, "Nonparametric analysis of structural health monitoring data for identification and localization of changes: Concept, lab, and real-life studies," *Structural Health Monitoring*.

[15] M. Malekzadeh, M. Gul, and F. N. Catbas, "Application of multivariate statistically based algorithms for civil structures anomaly detection," *Conference Proceedings of the Society for Experimental Mechanics Series*, pp. 289–298, 2013.

[16] O. Sarayanibafghi, G. Atia, M. Malekzadeh, and N. Catbas, "Compressed change detection for structural health monitoring," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 1231–1235.

[17] S. Ray, D. Starobinski, A. Trachtenberg, and R. Ungrangsi, "Robust location detection with sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1016–1025, 2004.

[18] M. Laifenfeld and A. Trachtenberg, "Identifying codes and covering problems," *IEEE Trans. on Inf. Theory*, vol. 54, no. 9, pp. 3929–3950, 2008.

[19] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin, "On a new class of codes for identifying vertices in graphs," *IEEE Trans. on Inf. Theory*, vol. 44, no. 2, pp. 599–611, Mar 1998.

[20] A.G. Tartakovsky and V.V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Prob. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.

[21] O. Sarayanibafghi and G. Atia, "Compressed change detection," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) - Signal Processing Theory and Methods (ICASSP2014 - SPTM)*, Florence, Italy, May 2014, pp. 3429–3433.

[22] G. Atia, A. Sahai, and V. Saligrama, "Spectrum enforcement and liability assignment in cognitive radio systems," in *Proceedings of the IEEE DySpAN*, Chicago, IL, Oct. 2008.

[23] A. Frieze, R. Martin, J. Moncel, M. Ruszink, and C. Smyth, "Codes identifying sets of vertices in random networks," *Discrete Mathematics*, vol. 307, no. 910, pp. 1094–1107, 2007.

[24] Alexander G Tartakovsky and Venugopal V Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory of Probability & Its Applications*, vol. 49, no. 3, pp. 458–497, 2005.

[25] Albert N Shiryaev, *Optimal stopping rules*, vol. 8, Springer Science & Business Media, 2007.

[26] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[27] E. J. Candes, "Compressive sampling," in *the International Congress of Mathematicians*, Madrid, Spain, 2006.

[28] Daniele Posenato, Francesca Lanata, Daniele Inaudi, and Ian F.C. Smith, "Model-free data interpretation for continuous monitoring of complex structures," *Advanced Engineering Informatics*, vol. 22, no. 1, pp. 135 – 144, 2008, Intelligent computing in engineering and architecture.

[29] I. Kwon, M. Malekzadeh, Q. Ma, H. Gokce, T. Terrell, A. Fedotov, and F. N. Catbas, "Fiber optic sensor installation for monitoring of 4 span model bridge in UCF," in *Rotating Machinery, Structural Health Monitoring, Shock and Vibration, Volume 5*, Conference Proceedings of the Society for Experimental Mechanics Series, pp. 383–388. Springer New York, 2011.

[30] Thomas M Cover and Joy A Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[31] Rick Durrett, *Probability: theory and examples*, Cambridge university press, 2010.