

COMPRESSED CHANGE DETECTION

Omid Sarayanibafghi and George Atia

Department of Electrical Engineering and Computer Science
University of Central Florida

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 paper the problem of identification of a sparse number of *statistical changes* in stochastic phenomena is considered. This framework, which is newly introduced herein, is termed Compressed Change Detection. 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. 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. Sufficient conditions are derived for the probability of detection to approach 1 in the asymptotic regime where N is large. Several applications and generalizations of the proposed framework are presented.

Index Terms— Identifying codes, Change detection, Sparsity.

1. INTRODUCTION

Physical and natural processes often exhibit variability over several dimensions such as space and time. 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].

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*. Since 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, 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.

Contrast to Prior Work: Classical change point detection has been used to identify changes in the distributions of stochastic processes and time series [3]. A dominant stream of research in this area has been focused on the design of efficient change detection rules with favorable delay and false alarm tradeoffs [4, 5]. A large number of algorithms, including CUSUM, MPCA have been proposed to minimize the detection delay subject to constraints on the probability of false alarm in different settings [3, 6, 7]. More recently, [8, 9] studied the change detection problem in a distributed setting using multi-sensor observations. Also, decentralized change detection was considered in [4, 5, 8] wherein sensors send their local decisions rather than their raw observations to a fusion center to decide if a change has occurred.

In this paper we introduce a new framework, which we term compressed change detection. 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 [10] with change point detection in sequential analysis, which to the best of our knowledge is explored here for the first time.

Related work considered the problem of location detection in sensor networks using identifying codes [11]. Finding identifying codes over graphs was considered in [12] 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 [10]. 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.

This work was supported in part through NSF Grant CCF-1320547.

The rest of the paper is organized as follows. In Section 2, the basic problem setup is introduced and some preliminary background about identifying codes over graphs is provided. In Section 3, we present the proposed approach, which is based on integrating coding over graphs with sequential change detection. An asymptotic performance analysis of the proposed approach is presented in Section 4 and numerical results are presented in Section 5. We discuss several generalizations in Section 6 and conclude in Section 7.

2. PROBLEM SETUP

For ease of exposition, we first describe a simplified version of the problem. Generalizations are discussed in Section 6. Suppose there are N random features, X_1, X_2, \dots, X_N , and M sensors measuring these features. The measurement, $Y_i, i = 1, \dots, M$, of sensor i is only affected by a subset of features. At some unknown point in time, an unknown subset S of features undergoes a statistical change. Let X_S denote the set of features indexed by the set S , i.e., $X_S = \{X_j\}_{j \in S}$. It is further assumed that the features are independent and identically distributed with a pre-change probability density function $f_{X_j}(x) = f_0(x), j = 1, \dots, N$. The post change density for the set features in S is $f_{X_j}(x) = f_1(x), j \in S$, 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 $K \ll N$, 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 .

2.1. Identifying Codes: Preliminaries

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

$$\mathcal{I}(v, \mathcal{C}) = \mathcal{N}(v) \cap \mathcal{C}, \quad (1)$$

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 [10, 11].

Definition 2.1 A collection of vertices $\mathcal{C} \subseteq V$ is called an identifying code if $\forall v_i, v_j \in V$ and $v_i \neq v_j$

$$\mathcal{I}(v_i, \mathcal{C}) \neq \emptyset, \quad (2)$$

$$\mathcal{I}(v_i, \mathcal{C}) \neq \mathcal{I}(v_j, \mathcal{C}). \quad (3)$$

Equation (2) in Definition 2.1 means that the identifying set for each vertex with respect to \mathcal{C} is non-empty, and (3) 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 [11] for source localization, in [10] to detect faults in a communication network, and in [13] 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 [14] 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 [11]. 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) and (3) 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.

3. PROPOSED APPROACH

3.1. Mapping to a Bipartite Graph

The first step of the proposed approach is to map our problem to a bipartite graph, $G = (V, E)$, with vertex set $V = A \cup B$, where A corresponds to the set of sensors and B to the set of features, i.e., $|A| = M, |B| = N$ and $A \cap B = \emptyset$. An edge, $e_{ij} \in E$, designates that sensor i measures feature j .

3.2. 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/features 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 [15, 16], 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 B . To illustrate the proposed approach, we consider an example in Fig. 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 CUSUM change detectors placed at these nodes in Fig. 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 si-

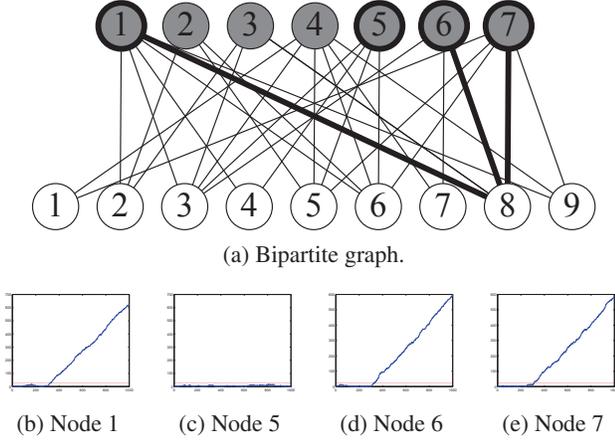


Fig. 1. Our proposed approach for a set of 9 features (white circles) and 7 sensors (gray circles). The bolded circles denote the ID-code. CUSUM statistics at sensors 1, 6, and 7 exceed the threshold, thus feature 8 is detected.

multaneously, 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 Section 6 using K -identifying codes.

4. ASYMPTOTIC ANALYSIS

In this section, we analyze the performance of the proposed approach in the asymptotic regime where the number of features $N \rightarrow \infty$. 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 Erdős-Rényi bipartite graph with probability of edge placement p . The same approach can be extended to other graphs (See Section 6).

Let $v \in B$ denote the vertex corresponding to feature X_j . To correctly declare a change for X_j , every node in the identifying set $\mathcal{I}(v, \mathcal{C})$ should detect the change and all remaining nodes in the ID-code should not raise a false alarm. Let $d_u(k) = 1$ if node u declares a change at time k and 0 otherwise. Given that the observed pattern of change matches the identifying set $\mathcal{I}(v, \mathcal{C})$ of a node v , and assuming that only the sensors observing v sense the change, the detection delay associated with a true change at node $v \in B$ is

$$T_v = \inf\{k : d_u(k) = 1, \forall u \in \mathcal{I}(v, \mathcal{C})\}. \quad (4)$$

Thus, T_v is the first time that *all* the nodes in the identifying set of v declare a change. Nevertheless, due to the randomness of the measurements, false alarms may occur. We lower bound the probability $P_{ta}(v)$ that a randomly generated code \mathcal{C} will generate a true event alert at node v , thus establishing a sufficient condition for the expected detection time in terms of the number of nodes N . Define the event \mathcal{E} as:

$$\mathcal{E} = \{\exists u \in \mathcal{C} \setminus \mathcal{I}(v, \mathcal{C}) : d_u = 1\}.$$

We further assume that conditioned on the presence or absence of an event, the signals at the different nodes are independent. Hence,

$$P_{ta}(v) \geq P\{\text{true alert}, \mathcal{C} \text{ is an ID-code}\} \quad (5)$$

$$\geq P\{\mathcal{E}^c \mid \mathcal{C} \text{ is an ID-code}\} P\{\mathcal{C} \text{ is an ID-code}\} \quad (6)$$

$$\geq (1 - \alpha_f)^{|\mathcal{C}|} \cdot P\{\mathcal{C} \text{ is an ID-code}\} \quad (7)$$

where α_f is the probability of false alarm of any of the change detectors and c denotes the complement. Henceforth, $P\{\mathcal{E}^c \mid \mathcal{C} \text{ is an ID-code}\}$, is termed the probability of miss detection, P_{md} . Thus, P_{md} is the probability of misclassifying the change, given that the graph admits an ID-code. We can readily state the following theorem.

Theorem 4.1 *For an Erdős-Rényi bipartite graph, the probability of true alert $P_{ta}(v)$ for a change at node v approaches 1 asymptotically if $\alpha_f = o\left(\frac{1}{\log N}\right)$, where α_f is the probability of false alarm at any of the change detectors defined over the identifying code. Furthermore, under this condition, the expected detection delay $\bar{T} \in O(\log N \cdot \log \log N)$.*

To prove this theorem we establish the following lemma. For Erdős-Rényi bipartite graphs the following lemma establishes an upper bound on the probability that a randomly generated code of size $|\mathcal{C}|$ is not identifying.

Lemma 4.1 *Given bipartite graphs $G(A \cup B, p)$ with $|B| = N$ and edge placement probability p , then the probability $\Pr(\mathcal{C} \text{ not a code})$ that a code $\mathcal{C} = A$ of size $|\mathcal{C}|$ is not an ID-code is upper bounded by $N^2(1 - \min\{p, 2p(1 - p)\})^{|\mathcal{C}|}$.*

The proof of this lemma is omitted for brevity and we refer the reader to an extended version of this work for further details.

Corollary 4.1 *The probability that a randomly generated code \mathcal{C} is an ID-code goes asymptotically to 1 for some $|\mathcal{C}| = O(\log N)$, i.e., \exists constant $C_1 > 0$ such that:*

$$P\{\mathcal{C} \text{ is an ID-code}\} \rightarrow 1 \text{ if } |\mathcal{C}| \geq C_1 \log N. \quad (8)$$

Hence, from Corollary 4.1 it follows that for Erdős-Rényi bipartite graphs of size N , $P_{ta}(v)$ goes to 1 asymptotically if $\alpha_f = o\left(\frac{1}{\log N}\right)$. Under this sufficient condition, we can also consider the scaling of the expected detection delay \bar{T} . The

expected detection time \bar{T}_{cd} of the optimal change detector scales logarithmically with the False Alarm probability α_f as

$$\bar{T}_{cd} \sim \frac{-\log \alpha_f}{D(f_0, f_1)},$$

where $D(f_0, f_1)$ is the KL distance between the pre-change and post-change distributions, in which case the overall expected delay is $\bar{T} \in O(\log N \cdot \log \log N)$, by corollary 4.1.

5. NUMERICAL RESULTS

To back up our asymptotic analysis we first compute the length of the ID-code as N increases. In Fig. 2 the length of the ID-code is plotted vs. $\log N$. We observe that for Erdős-Rényi bipartite random graphs, the length of the ID-code grows logarithmically with the size of the graph, which matches (8). This demonstrates that significant savings can be achieved for detecting sparse changes. In Fig. 3, we ob-

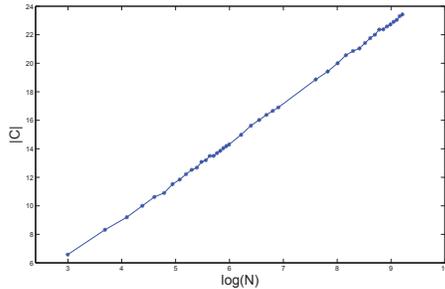


Fig. 2. Length of the ID-code vs. number of nodes in a random bipartite graph with $p = 0.5$.

tain the relation between the detection delay and P_{md} when $f_0(x) = \mathcal{N}(0, 1)$ and $f_1(x) = \mathcal{N}(1, 1)$ using 1000 iterations for a random bipartite graph with $N = 30$ and probability of edge placement $p = 0.5$. This shows that a high probability of detection can be achieved while incurring small detection delays.

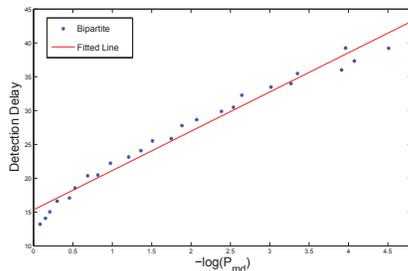


Fig. 3. Detection delay versus probability of miss detection for a random bipartite graph with $p = 0.5$.

6. GENERALIZATIONS

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

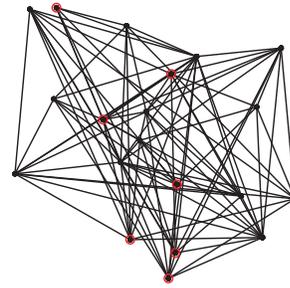


Fig. 4. Identifying code for a 20-node graph. The nodes encircled in red form an identifying code.

Multiple Concurrent Changes: This approach and principle can be extended to multiple concurrent changes. The idea is to use K -identifying codes which would provide a unique signature for up to K simultaneous changes. In such a case we choose a code \mathcal{C} such that $\forall \mathcal{I}, \mathcal{J} \subseteq B, 1 \leq |\mathcal{I}|, |\mathcal{J}| \leq K$ and $\mathcal{I} \neq \mathcal{J}$

$$\bigcup_{v_i \in \mathcal{I}} \mathcal{I}(v_i, \mathcal{C}) \neq \bigcup_{v_j \in \mathcal{J}} \mathcal{I}(v_j, \mathcal{C}) \quad (9)$$

$$\mathcal{I}(v_i, \mathcal{C}) \neq \emptyset, \quad \forall v_i \in B, \quad (10)$$

then sequential change detection can be implemented at \mathcal{C} .

Arbitrary Graph Topology: To this end, we have only considered Erdős-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 Section 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 4 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.

7. CONCLUSIONS

In this paper, 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 detection probability to go to 1. Generalizations to multiple concurrent changes and arbitrary graph topologies were also discussed.

8. 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] Michèle Basseville and Igor V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, NJ: Prentice Hall, 1993.
- [4] V.V. Veeravalli, “Decentralized quickest change detection,” *IEEE Trans. on Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.
- [5] 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.
- [6] G. Lorden, “Procedures for Reacting to a Change in Distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] S. Ray, D. Starobinski, A. Trachtenberg, and R. Ungangsi, “Robust location detection with sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1016–1025, 2004.
- [12] M. Laifenfeld and A. Trachtenberg, “Identifying codes and covering problems,” *IEEE Trans. on Inf. Theory*, vol. 54, no. 9, pp. 3929–3950, 2008.
- [13] 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.
- [14] Alan Frieze, Ryan Martin, Julien Moncel, Mikl Ruszink, and Cliff Smyth, “Codes identifying sets of vertices in random networks,” *Discrete Mathematics*, vol. 307, no. 910, pp. 1094 – 1107, 2007.
- [15] D. L. Donoho, “Compressed sensing,” *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.
- [16] E. J. Candes, “Compressive sampling,” in *the International Congress of Mathematicians*, Madrid, Spain, 2006.