

# A KERNEL-BASED NONPARAMETRIC TEST FOR ANOMALY DETECTION OVER LINE NETWORKS

Shaofeng Zou, Yingbin Liang

Syracuse University  
Department of EECS  
{szou02,yliang06}@syr.edu

H. Vincent Poor

Princeton University  
Department of EE  
poor@princeton.edu

## ABSTRACT

The nonparametric problem of detecting existence of an anomalous interval over a one-dimensional line network is studied. Nodes corresponding to an anomalous interval (if one exists) receive samples generated by a distribution  $q$ , which is different from the distribution  $p$  that generates samples for other nodes. If an anomalous interval does not exist, then all nodes receive samples generated by  $p$ . It is assumed that the distributions  $p$  and  $q$  are arbitrary, and are unknown. In order to detect whether an anomalous interval exists, a test is built based on mean embeddings of distributions into a reproducing kernel Hilbert space (RKHS) and the metric of maximum mean discrepancy (MMD). It is shown that as the network size  $n$  goes to infinity, if the minimum length of candidate anomalous intervals is larger than a threshold which has the order  $O(\log n)$ , the proposed test is asymptotically successful. An efficient algorithm to perform the test with substantial computational complexity reduction is proposed, and is shown to be asymptotically successful if the condition on the minimum length of candidate anomalous interval is satisfied. Numerical results are provided, which are consistent with the theoretical results.

## 1. INTRODUCTION

In this paper, we are interested in a type of problem, the goal of which is to detect existence of an anomalous object over a network. Each node in the network is associated with a random variable. An anomalous object, if it exists, corresponds to a cluster of nodes in the network that take samples generated by a distribution  $q$ . All other nodes in the network take samples generated by the distribution  $p$  that is different from  $q$ . If an anomalous interval does not exist, then all nodes receive samples generated by  $p$ . Detection of no anomalous object (i.e., the null hypothesis  $H_0$ ) against the anomalous event (i.e., hypothesis  $H_1$ ) is a compound hypothesis testing problem due to the fact that the anomalous object may correspond to one of a number of candidate clusters in the network.

Such a problem models a variety of applications. For ex-

ample, in sensor networks, sensors are deployed over a large range of space. These sensors take measurements from the environment in order to determine whether or not there is intrusion of an anomalous object. Such intrusion typically activates only a few sensors that cover a certain geometric area. An alarm is then triggered if the network detects an occurrence of intrusion based on sensors' measurements. Other applications can arise in detecting an anomalous segment of DNA sequences, detecting virus infection of computer networks, and detecting anomalous spot in images.

Detecting the existence of geometric objects in large networks has been extensively studied in the literature. A number of studies have focused on networks with nodes embedded in a lattice such as one dimensional line or square, e.g., [1, 2]. Further generalization of the problem has also been studied, when network nodes are associated with a graph structure, and existence of an anomalous cluster or an anomalous subgraph of nodes needs to be detected, e.g., [3–5].

The majority of previous studies on this topic have taken parametric or semiparametric models on probability distributions, i.e., random variables are generated by known distributions such as Gaussian or Bernoulli distributions, or the two distributions are known to have mean shift relationship. However, parametric models may not always be available in applications. In many cases, distributions can be arbitrary, and may not be Gaussian or Bernoulli. They may not differ in mean either. Furthermore, distributions may not be known in advance. Hence, it is desirable to develop nonparametric tests that are distribution free.

In contrast to previous studies, in this paper, we study a nonparametric model for anomalous interval detection, in which distributions can be arbitrary and unknown a priori. We focus on the problem of detecting existence of an anomalous interval over a one-dimensional line network. Although this is a simple network, it already captures the essence of the problem, and the same approach can be extended to studying more general network models.

In order to deal with nonparametric models, we apply mean embedding of distributions into a reproducing kernel

Hilbert space (RKHS) [6, 7] (see [8] for an introduction to RKHS). The idea is to map probability distributions into an RKHS associated with an appropriate kernel such that distinguishing between two probabilities can be carried out by evaluating the distance between the corresponding mean embeddings in the RKHS. This is valid because the mapping is shown to be injective for certain kernels [9] such as Gaussian and Laplacian kernels. The main advantage of such an approach is that the mean embedding of a distribution can be easily estimated based on samples. This approach has been applied to solving the two sample problem in [10], in which the quantity of *maximum mean discrepancy (MMD)* was used as a metric of distance between mean embeddings of two distributions.

Since the distributions can be arbitrary, it is in general difficult to exploit properties of the distributions such as mean shift to detect existence of an anomalous interval. Furthermore, as the network size becomes large (i.e., the number  $n$  of nodes goes to infinity), in contrast to parametric models in which the mean shift can scale with  $n$ , here it is necessary that the length of anomalous intervals (i.e., the number of samples over the anomalous distribution) is large enough to allow accurately identifying such an interval (see Remark 1 in Section 2). This implies that the scale of the anomalous object should enlarge as the detection range becomes larger in order for successful detection. Such a behavior also sets a clear difference of our nonparametric problem from previous studies of parametric models. Our goal is to characterize how the minimum length of candidate anomalous intervals should scale as the number of nodes goes to infinity in order to successfully detect existence of an anomalous interval.

We summarize our main contributions as follows.

(1) We address the nonparametric model of detecting existence of an anomalous interval over a line network. We identify the length of the anomalous interval as an essential characteristic that must enlarge with  $n$  to guarantee successful detection in asymptotically large networks (see Remark 1 in Section 2). Hence, requiring the length of the anomalous interval to scale with  $n$  is necessary, not an artificial assumption.

(2) We build a distribution-free test using MMD based on kernel embeddings of distributions into an RKHS. We analyze the performance guarantee of the proposed test, and show that as the network size  $n$  goes to infinity, if the minimum length of candidate anomalous intervals scales on the order  $O(\log n)^1$  or larger, the proposed test can successfully detect whether there exists an anomalous interval. Furthermore, we show that the test and the minimum length depends on prior knowledge of the MMD of the two distributions.

(3) We adopt the multi-scale method in [1] and propose an efficient algorithm to perform the nonparametric test, which reduces the number of intervals for which MMD needs to be

<sup>1</sup>In this paper,  $f(n) = O(g(n))$  denotes that  $f(n)/g(n)$  converges to a constant as  $n \rightarrow \infty$ .

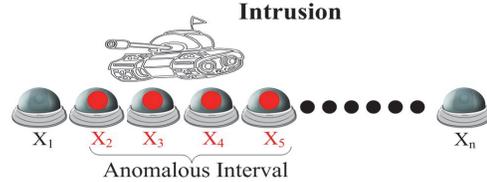


Fig. 1. A line sensor network for intrusion detection

computed from the order  $O(n^2)$  to  $O(n^{1+\rho})$ , for any  $\rho > 0$ . We further prove a performance guarantee for the proposed algorithm.

(4) We provide numerical results that are consistent with our theoretical assertions and demonstrate that the proposed test indeed provides guaranteed performance.

## 2. PROBLEM STATEMENT

We consider a line network, which consists of nodes  $1, \dots, n$ , as shown in Figure 1. We use  $I$  to denote a subset of consecutive indices of nodes, which is referred to as an *interval*. Here, the length of an interval  $I$  refers to the cardinality of  $I$ , and is denoted by  $|I|$ . We assume that each node, say node  $i$ , is associated with a random variable, denoted by  $Y_i$ , for  $i = 1, \dots, n$ . We use  $\mathcal{I}_n$  to denote a set that contains all intervals over the network. We further denote the set of all candidate anomalous intervals as

$$\mathcal{I}_n^{(a)} = \{I \in \mathcal{I}_n : |I| \geq I_{\min}\} \quad (1)$$

where  $I_{\min}$  denotes the minimum length of candidate anomalous intervals. The reason for imposing such a minimum length requirement is explained in Remark 1.

We consider two hypotheses about the distributions of the line network. For the *null hypothesis*  $H_0$ ,  $Y_i$  for  $i = 1, \dots, n$  are identical and independently distributed (i.i.d.) random variables, and are generated from a distribution  $p$ . For the *alternative hypothesis*  $H_1$ , there exists an interval  $I \in \mathcal{I}_n^{(a)}$  over which  $Y_i$  are i.i.d. and are generated from a distribution  $q \neq p$  for all  $i \in I$ , and otherwise,  $Y_i$  are i.i.d. and generated from the distribution  $p$ . We further assume that under both hypotheses, each node generates only one sample.

In contrast to previous work, we assume that the distributions  $p$  and  $q$  are arbitrary and are unknown a priori. Instead, one sample  $X_i$  is independently generated from the distribution  $p$  for each node as a reference sample for the null hypothesis. This is reasonable because in practical scenarios, systems typically start under  $H_0$  and it is not difficult to collect samples at this stage. The distributions  $p$  and  $q$  can be either continuous or discrete. Therefore, it is hard to estimate the probability distribution functions directly from the samples with guaranteed convergence rate, and measure the difference between the probability distribution functions.

For this problem, we are interested in the asymptotic scenario, in which the number of nodes goes to infinity, i.e.,  $n \rightarrow \infty$ . The performance of a test for such a system is captured by the two types of errors. The *type I error* refers to the event in which samples are generated from the null hypothesis, but the detector determines an anomalous event occurs. We denote the probability of such an event as  $P(H_1|H_0)$ , or  $P_{H_0}(\text{error})$ . The *type II error* refers to the case in which an anomalous event occurs but the detector claims that the samples are generated from the null hypothesis. We denote the probability of such an event as  $P(H_0|H_1)$ , or  $P_{H_1}(\text{error})$ .

**Definition 1.** A test is said to be asymptotically successful if

$$\lim_{n \rightarrow \infty} P(H_1|H_0) + P(H_0|H_1) \rightarrow 0. \quad (2)$$

Although we are interested in the asymptotic scenario, we also have the results for finite  $n$  in the intermediate steps of our proof.

This hypothesis testing problem has a compound nature in that  $H_1$  includes events corresponding to all candidate intervals where an anomalous object can be located, i.e., for all  $I \in \mathcal{I}_n^{(a)}$ . In general, an anomalous interval with shorter length is more difficult to detect due to the smaller number of samples from the anomalous distribution  $q$ . As  $n \rightarrow \infty$ , the total number of intervals goes to infinity in the order of  $O(n^2)$ . In this case, to achieve successful detection, each candidate anomalous interval should provide more accurate information about the corresponding distribution. This requires that the length of candidate anomalous intervals grow with  $n$ . This suggests that as the network becomes larger, it can detect only a sufficiently large anomalous object.

**Remark 1.** We argue that it is necessary for the minimum length  $I_{\min}$  of candidate anomalous intervals to grow to infinity as  $n \rightarrow \infty$  in order to guarantee asymptotically successful detection in a nonparametric model. This is because in the nonparametric model, the distributions  $p$  and  $q$  are fixed as  $n$  changes. Now suppose  $p$  and  $q$  are both Gaussian but with different mean values. Since mean values do not scale with  $n$  in our model, following [1] Theorem 2.3, no test can be asymptotically successful if  $I_{\min}$  is bounded. Therefore, no distribution-free test exists if the minimum length  $I_{\min}$  is bounded as  $n \rightarrow \infty$ .

Therefore, our goal in this problem is to characterize how the minimum length  $I_{\min}$  of candidate anomalous intervals should scale with the network size (i.e., the number of nodes) in order for a detector to successfully distinguish between the two hypotheses.

### 3. MAIN RESULTS

#### 3.1. Introduction to MMD

We provide a brief introduction to the idea of mean embedding of distributions into an RKHS [6, 7] and the MMD met-

ric. Suppose  $\mathcal{P}$  includes a class of probability distributions, and suppose  $\mathcal{H}$  is the RKHS with an associated kernel  $k(\cdot, \cdot)$ . We define a mapping from  $\mathcal{P}$  to  $\mathcal{H}$  such that each distribution  $p \in \mathcal{P}$  is mapped into an element in  $\mathcal{H}$  as follows:

$$\mu_p(\cdot) = \mathbb{E}_p[k(\cdot, x)] = \int k(\cdot, x) dp(x).$$

Here,  $\mu_p(\cdot)$  is referred to as the *mean embedding* of the distribution  $p$  into the Hilbert space  $\mathcal{H}$ .

It is desirable that the embedding is *injective* such that each  $p \in \mathcal{P}$  is mapped to a unique element  $\mu_p \in \mathcal{H}$ . It has been shown in [9] that for many RKHSs such as those associated with Gaussian and Laplacian kernels, the mean embedding is injective. In order to distinguish between two distributions  $p$  and  $q$ , [10] introduced the MMD based on the mean embeddings  $\mu_p$  and  $\mu_q$  of  $p$  and  $q$  in RKHS:

$$\text{MMD}[p, q] := \|\mu_p - \mu_q\|_{\mathcal{H}}. \quad (3)$$

Due to the reproducing property of the kernel, it can be easily shown that

$$\begin{aligned} \text{MMD}^2[p, q] &= \mathbb{E}_{x, x'}[k(x, x')] - 2\mathbb{E}_{x, y}[k(x, y)] \\ &\quad + \mathbb{E}_{y, y'}[k(y, y')], \end{aligned} \quad (4)$$

where  $x$  and  $x'$  are independent with the same distribution  $p$ , and  $y$  and  $y'$  are independent with the same distribution  $q$ . An unbiased estimate of  $\text{MMD}^2[p, q]$  based on  $n$  samples of  $x$  and  $m$  samples of  $y$  is given by

$$\begin{aligned} \text{MMD}_u^2[X, Y] &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n k(x_i, x_j) \\ &\quad + \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j \neq i}^m k(y_i, y_j) - \frac{2}{nm} \sum_{i=1}^n \sum_{j=1}^m k(x_i, y_j). \end{aligned} \quad (5)$$

The properties of MMD and the estimator of MMD are analyzed in detail in [10].

#### 3.2. Test and Performance Analysis

We construct our test using the unbiased estimate  $\text{MMD}_u^2[X, Y]$  of  $\text{MMD}^2[p, q]$  given in (5). In particular, for each interval  $I \in \mathcal{I}_n^{(a)}$ , we compute  $\text{MMD}_{u, I}^2[X, Y]$  using (5) based on samples  $(y_j, j \in I)$  and the reference sequence  $(x_1, \dots, x_n)$  generated by  $p$ .

If there exists an anomalous interval  $I$ , we expect the corresponding  $\text{MMD}_{u, I}^2[X, Y]$  to be large, because the sequence of the anomalous interval is generated by a distribution  $q$  differently from the reference sequence. Otherwise, under the null hypothesis  $\text{MMD}_{u, I}^2[X, Y]$  should be small for all candidate  $I$ . Hence, we build our test as follows:

$$\max_{I \in \mathcal{I}_n^{(a)}} \text{MMD}_{u, I}^2(X, Y) \begin{cases} \geq t, & \text{determine } H_1 \\ < t, & \text{determine } H_0 \end{cases} \quad (6)$$

where  $t$  is a threshold parameter, which is determined in Corollaries 1 and 2 below.

The following theorem provides the performance of the above test.

**Theorem 1.** *Suppose the test (6) is applied to the nonparametric problem given in Section 2. Further assume that the kernel in the test satisfies  $0 \leq k(x, y) \leq K$  for all  $(x, y)$ . Then the test (6) is asymptotically successful if*

$$I_{\min} \geq \frac{16K^2(1+\eta)}{t^2} \log n \quad (7)$$

where  $\eta$  is any positive constant, and  $t$  is the threshold of the test that satisfies  $t < \text{MMD}^2[p, q]$ .

We note that the boundedness assumption on  $k(x, y)$  is satisfied for many kernels such as Gaussian and Laplacian kernels. This boundedness property is exploited to analyze the probability of error when using concentration inequality in the proof. We further note that the above theorem implies that the minimum length  $I_{\min}$  can be in the order  $O(\log n)$ . This implies that the number of candidate anomalous intervals in the set  $\mathcal{I}_n^{(a)}$  is on the order  $O(n^2)$ , which is the same as the number of all intervals. Thus, in the order sense, not many intervals are excluded from being candidate anomalous intervals.

Theorem 1 requires that the threshold  $t$  in the test (6) be less than  $\text{MMD}^2[p, q]$ . Knowledge of  $\text{MMD}^2[p, q]$  may or may not be available depending on specific applications. If  $\text{MMD}^2[p, q]$  is known, the threshold  $t$  can be set as a constant smaller than  $\text{MMD}^2[p, q]$ . If  $\text{MMD}^2[p, q]$  is not known, then the threshold  $t$  needs to scale to zero as  $n$  gets large in order to be asymptotically smaller than  $\text{MMD}^2[p, q]$ . We summarize these two cases in the following corollaries, which follows directly from Theorem 1.

**Corollary 1.** *If  $\text{MMD}^2[p, q]$  is known a priori, then set the threshold  $t = (1 - \delta)\text{MMD}^2[p, q]$  for any  $0 < \delta < 1$ . In this case, the test (6) is asymptotically successful if*

$$I_{\min} \geq \frac{16K^2(1+\eta')}{\text{MMD}^4[p, q]} \log n \quad (8)$$

where  $\eta'$  is any positive constant.

**Corollary 2.** *If  $\text{MMD}^2[p, q]$  is unknown a priori, then set the threshold  $t$  to scale with  $n$  such that  $\lim_{n \rightarrow \infty} t_n = 0$ . In this case, the test (6) is asymptotically successful if*

$$I_{\min} \geq \frac{16K^2(1+\eta)}{t_n^2} \log n. \quad (9)$$

It is clear from Corollary 2 that for the case when  $\text{MMD}^2[p, q]$  is unknown, the growth rate of the minimum length  $I_{\min}$  is strictly larger than the order  $O(\log n)$ . This demonstrates that the prior knowledge about  $\text{MMD}^2[p, q]$  is

very important for network capability in identifying anomalous events. If  $\text{MMD}^2[p, q]$  is known, then the network can resolve an anomalous object with the length on the order  $O(\log n)$ . However, if such knowledge is unknown, the network can resolve only bigger anomalous objects with the lengths larger than  $O(\log n)$ .

### 3.3. Outline of Proof of Theorem 1

We first introduce the definition of dyadic intervals and their properties [1], which are useful for the proof of Theorem 1 and for understanding Algorithm 1 in Section 3.4. For convenience, assume that  $n = 2^J$ , where  $J$  is an integer, and define the dyadic intervals as

$$I_{j,k} = \{k2^j, \dots, (k+1)2^j - 1\}, 0 \leq j \leq \log_2 n, 0 \leq k \leq \frac{n}{2^j}.$$

We let  $\mathcal{I}_n^{(d)}$  denote the set that consists of all dyadic intervals. It was shown in [1] that for any interval  $I$ ,  $\frac{|I^{(d)}|}{|I|} \geq \frac{1}{4}$  where  $I^{(d)}$  is the largest dyadic interval contained in  $I$ .

We next define the  $l$ -level extensions of a dyadic interval  $I_{j,k}$  as follows. Starting from the interval  $I_{j,k}$  or the union of  $I_{j,k}$  and  $I_{j,k+1}$  where  $k$  is odd, at level  $q = 1, \dots, l$ , attach dyadic intervals of length  $2^{-q}|I_{j,k}|$  at either, or both ends of the interval resulting from the previous step, or do nothing. Let  $\mathcal{J}_{n,l}(I)$  denote the set that includes all  $l$ -level extensions of a dyadic interval  $I$ . Let  $\mathcal{J}_{n,l} = \cup_{I \in \mathcal{I}_n^{(d)}} \mathcal{J}_{n,l}(I)$ . We note the following useful properties.

**Lemma 1.** [1] (1)  $|\mathcal{J}_{n,l}| \leq n^{4l+1}$ ; (2) For any interval  $I$  and the corresponding maximum dyadic interval  $I^{(d)}$  contained in  $I$ , there exists one interval  $J$  in the  $l$ -level extension  $\mathcal{J}_{n,l}(I^{(d)})$  of  $I^{(d)}$  such that  $|I| - |J| \leq 2^{-(l-1)}|I^{(d)}|$ .

*Outline of Proof of Theorem 1:* We now provide the main idea to prove Theorem 1 with the complete proof provided in [11]. For each set  $I \in \mathcal{I}_n^{(a)}$ , we find the corresponding set  $J_I \in \mathcal{J}_{n,l}(I^{(d)})$  that satisfies Lemma 1 (2). We then collect all such intervals  $J_I$  into a set  $\mathcal{J}_{n,l}^{(a)}$ . The idea of the proof is to use  $\max_{J \in \mathcal{J}_{n,l}^{(a)}} \text{MMD}_{u,J}^2[X, Y]$  as a good approximation of the true test in (6). Since the number of intervals in  $\mathcal{J}_{n,l}^{(a)}$  is much smaller than the number of intervals in  $\mathcal{I}_n^{(a)}$ , a test based on this approximation helps to tighten the result. Based on this idea, under  $H_0$ , we bound  $P_{H_0}(\text{error})$  as,

$$P_{H_0}(\text{error}) \leq \exp\left(2 \log n - \frac{\epsilon^2 2^l I_{\min}}{c_1 + c_2 + c_3}\right) + \exp\left(\log n + (l+1) \log 4 - \frac{(t-\epsilon)^2(1-2^{1-l})I_{\min}}{16K^2}\right)$$

where  $c_1, c_2, c_3$  are constants. We further set  $\epsilon = (1 - \beta)t$ , where  $0 < \beta < 1$  is a constant. It can be shown that there exist  $\beta$  close enough to 1 and  $l$  large enough (but a constant) such that  $P_{H_0}(\text{error}) \rightarrow \infty$  as  $n \rightarrow \infty$  if  $I_{\min} > \frac{16K^2(1+\eta)}{t^2} \log n$ .

Under  $H_1$ , suppose  $\hat{I}$  is the anomalous interval. Using the fact that  $t < \text{MMD}^2[p, q]$ , we have the following bound:

$$P_{H_1}(\text{error}) \leq \exp\left(-\frac{(\text{MMD}^2[p, q] - t)^2 I_{\min}}{16K^2}\right) \quad (10)$$

which converges to zero as  $n \rightarrow \infty$ , if  $I_{\min} > \frac{16K^2(1+\eta)}{t^2} \log n$ .

### 3.4. An Efficient Algorithm

We describe an efficient algorithm to perform the proposed test (6). In general, since the number of intervals with length larger than  $I_{\min}$  has an order  $O(n^2)$ , the test (6) requires computation of  $\text{MMD}_{u,I}^2[X, Y]$  for  $O(n^2)$  intervals. We next provide Algorithm 1 that computes  $\text{MMD}_{u,I}^2[X, Y]$  for only  $O(n^{1+\rho})$  intervals for any  $\rho > 0$ . This algorithm adapts the multi-scale method in [1] for parametric models.

The basic idea of the algorithm is to use the set of dyadic intervals and their extensions (as introduced in Section 3.3) such that  $\text{MMD}_{u,I}^2[X, Y]$  over any interval is well approximated by an interval in such a set. A key property of such a set is that its cardinality is on the order  $O(n^{1+\rho})$  for any  $\rho > 0$ , which reduces computation of  $\text{MMD}_{u,I}^2[X, Y]$  to only  $O(n^{1+\rho})$  intervals.

---

#### Algorithm 1 Detect Existence of an Anomalous Interval

---

**Input:**  $n; t = t_n \rightarrow 0; t' < t; \eta > 0; I_{\min} \geq \frac{16K^2(1+\eta)}{t^2} \log n, \delta > \frac{\eta}{2}$ ; and  $l = \lceil \log_2\left(\frac{1+\eta}{\eta}\right) + 2 \rceil$ .

**Output:**

- Construct the set  $\mathcal{I}_n^{(p)} = \{I \in \mathcal{I}_n^{(d)} : |I| \geq \frac{I_{\min}}{4}, \text{ and } \text{MMD}_{u,I}^2[X, Y] \geq t'\}$ ;
  - If  $|\mathcal{I}_n^{(p)}| > n^{1 - \frac{\eta^2}{4t^2}(1+\eta) + \delta}$ , then determine  $H_1$ ;
  - If  $\max_{I \in \mathcal{I}_n^{(p)}} \text{MMD}_{u,I}^2[X, Y] > \frac{2t}{\sqrt{1+\eta/2}}$ , then determine  $H_1$ ;
  - Construct the set  $\mathcal{J}_{n,l}^{(p)}$  that includes level- $l$  extensions of all intervals in  $\mathcal{I}_n^{(p)}$  with lengths larger than  $\frac{16K^2(1+\eta/2)}{t^2} \log n$ ;
  - If  $\max_{I \in \mathcal{J}_{n,l}^{(p)}} \text{MMD}_{u,I}^2[X, Y] > t$ , then determine  $H_1$ ;
  - Otherwise, determine  $H_0$ .
- 

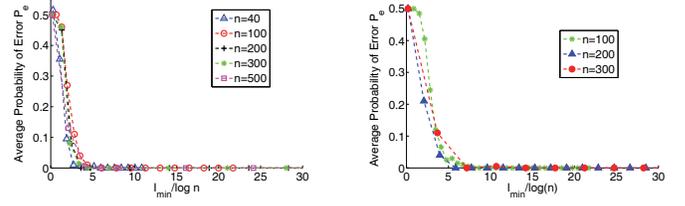
The following theorem provides a performance guarantee for Algorithm 1. The proof is proved in [11].

**Theorem 2.** *Algorithm 1 is asymptotically successful with computation of  $\text{MMD}_{u,I}^2[X, Y]$  for the order  $O(n^{1+\rho})$  intervals for any  $\rho > 0$ .*

## 4. NUMERICAL RESULTS

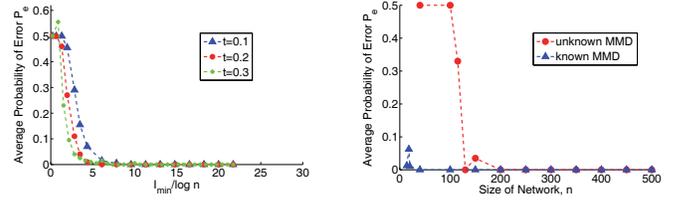
In this section, we provide four numerical experiments to demonstrate our theoretical results for both cases with known and unknown  $\text{MMD}[p, q]$ . In these numerical results, we average the two types of errors as

$$P_e = \frac{1}{2} (P_{H_1}(\text{error}) + P_{H_0}(\text{error})).$$



**Fig. 2.** Performance of detecting an anomalous interval distributed as a mixture of two Gaussian distributions out of a Gaussian distributed line network with Gaussian kernel and known MMD

**Fig. 3.** Performance of detecting an anomalous interval with Gaussian distribution that has a different variance from other nodes in a line network with Gaussian kernel and known MMD



**Fig. 4.** Impact of threshold  $t$  in the test on  $I_{\min}$

**Fig. 5.** Impact of knowledge of MMD

The first two experiments study how  $I_{\min}$  scales with the network size  $n$  in order to guarantee successful experiments. For these two experiments, we assume that a good estimate of  $\text{MMD}[p, q]$  is available.

In experiment 1, the distribution  $p$  is  $\mathcal{N}(0, \frac{1}{2})$ , and  $q$  is a mixture of two Gaussian distributions  $\mathcal{N}(-2, \frac{1}{2})$  and  $\mathcal{N}(2, \frac{1}{2})$  with equal probability. For detection, we use the Gaussian kernel  $k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$  with  $\sigma = 1$ , and we set the threshold  $t = 0.25$ . We run the experiment for networks with sizes  $n = 40, 100, 200, 300, 500$ , respectively.

In Figure 2, we plot how the average probability of error changes with the minimum length  $I_{\min}$ . For illustration convenience, we normalize  $I_{\min}$  by  $\log n$ . It can be seen that when  $\frac{I_{\min}}{\log n}$  is above a certain threshold, the probability of error converges to zero, which is consistent with our theoretical results. Furthermore, for different value of  $n$ , all curves drop to zero almost at the same threshold. Such behavior also agrees with Theorem 1, which implies that the threshold depends only on the bound of the kernel and the threshold of the experiment, and these parameters are the same for all curves.

In experiment 2, distributions  $p$  and  $q$  are respectively chosen to be  $\mathcal{N}(0, 1)$  and  $\mathcal{N}(0, 4)$ , i.e., they have the same mean but different variances. We use the Gaussian kernel with  $\sigma = 1$  for the experiment, and set the threshold  $t = 0.1$ . We run the experiment for networks with sizes  $n = 100, 200, 300$ . Figure 3 plots how the average probability of error changes with the minimum length  $I_{\min}$ , and demon-

strates a behavior similar to experiment 1.

In experiment 3, we study how the threshold  $t$  affects the scaling behavior of  $I_{\min}$  with  $n$ . We choose the same distributions  $p$  and  $q$  as in experiment 1. We also use the Gaussian kernel with  $\sigma = 1$  for our experiment. We study a network with size  $n = 100$ . We run the experiment for  $t = 0.1, 0.2, 0.3$ . In Figure 4, we plot the average probability of error versus  $\frac{I_{\min}}{\log n}$  corresponding to different values of  $t$ . Although the curves exhibit behavior similar to that in the first two experiments, the probabilities of error do not drop at the same threshold on  $\frac{I_{\min}}{\log n}$ . It can be seen that as  $t$  grows, the dropping threshold on  $\frac{I_{\min}}{\log n}$  gets smaller, implying that the network can detect smaller anomalous object. This is consistent with Theorem 1, which suggests that the threshold on  $I_{\min}$  to guarantee successful experiments is inversely proportional to  $t^2$ .

In experiment 4, we study the case in which the MMD is unknown, and compare its performance with the case when the MMD is known. We choose  $p$  to be  $\mathcal{N}(0, \frac{1}{2})$ , and choose  $q$  to be a mixture of two Laplacian distributions with the same variance  $\frac{1}{2}$  and different means of  $-3$  and  $3$  equally likely. We use the Gaussian kernel with parameter  $\sigma = 0.9$ . Since the MMD is unknown, we set the threshold  $t$  to change with  $n$  as  $t_n = 4\sqrt{\frac{\log n}{n^{0.9}}}$ , which goes to zero as  $n$  goes to infinity. As the network size  $n$  changes, we set the minimum length of anomalous intervals as  $I_{\min} = \lceil n^{0.9} \rceil$  suggested by Theorem 1. We also run a comparison experiment with MMD known in advance, for which we set the threshold  $t = 0.1$ . For a fair comparison, we also set  $I_{\min} = \lceil n^{0.9} \rceil$ .

In Figure 5, we plot how the average probability of error changes as a function of  $n$  with unknown MMD. It can be seen that as  $n$  becomes large, the probability of error goes to zero demonstrating that our experiment is asymptotically successful. This also agrees with Theorem 1 because we have chosen the minimum length to satisfy (7). It can also be seen that the probability of error for the case with known MMD converges much faster than the case with unknown MMD, demonstrating the importance of prior knowledge of the MMD.

## 5. CONCLUSIONS

In this paper, we have investigated the nonparametric problem of detecting existence of an anomalous interval over a line network. We have developed a distribution free test using the MMD, and have analyzed the performance of such a test. Furthermore, we have proposed an efficient algorithm to perform the test with reduced complexity, and have shown that the algorithm is asymptotically successful. Our results demonstrate that the MMD is a powerful metric that can be applied to a variety of detection problems involving distinguishing among distributions.

## Acknowledgment

The work of S. Zou and Y. Liang was supported by an NSF CAREER Award under Grant CCF-10-26565 and by the NSF under Grant CNS-11-16932. The work of H. V. Poor was supported by the ARO under MURI Grant W911NF-11-1-0036 and by the NSF under Grants DMS-11-18605 and EECS-13-43210.

## 6. REFERENCES

- [1] E. Arias-Castro, D. L. Donoho, and X. Huo, "Near-optimal detection of geometric objects by fast multiscale methods," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2402–2425, July 2005.
- [2] G. Walther, "Optimal and fast detection of spatial clusters with scan statistics," *Ann. Statist.*, vol. 38, no. 2, pp. 1010–1033, 2010.
- [3] L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi, "On combinatorial testing problems," *Ann. Statist.*, vol. 38, no. 5, pp. 3063–3092, 2010.
- [4] E. Arias-Castro, E. J. Candes, and A. Durand, "Detection of an anomalous cluster in a network," *Ann. Statist.*, vol. 39, no. 1, pp. 278–304, 2011.
- [5] J. Sharpnack, A. Rinaldo, and A. Singh, "Changepoint detection over graphs with the spectral scan statistic," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, Scottsdale, AZ, May 2013.
- [6] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer, 2004.
- [7] B. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, and B. Scholkopf, "Hilbert space embeddings and metrics on probability measures," *J. Mach. Learn. Res.*, vol. 11, pp. 1517–1561, 2010.
- [8] B. Scholkopf and A. Smola, *Learning with Kernels*, MA: MIT press, 2002.
- [9] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf, "Kernel measures of conditional dependence," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2008.
- [10] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, 2012.
- [11] S. Zou, Y. Liang, and H. V. Poor, "A kernel-based nonparametric test for anomaly detection over line networks," *Technical report*, 2013, available at <http://szou02.mysite.syr.edu/admin.html>.