The University of Kansas

KU INFORMATION & TELECOMMUNICATION TECHNOLOGY CENTER
The University of Kansas

Technical Report

Rail Sensor Testbed Program: Active Agents in Containers for Transport Chain Security: Algorithms

Ruoyi Jiang, Brian Quanz,
Hongliang Fei, Jun Huan

ITTC-FY2011-TR-47750-09

March 14, 2011

Project Sponsor:
Office of Naval Research
Contract: N00014-07-1-1042
The University of Kansas

20110324093

Copyright © 2011:
The University of Kansas
2335 Irving Hill Road, Lawrence, KS 66045-7559
All rights reserved.

DEFENSE TECHNICAL INFORMATION CENTER

Information for the Defense Community

DTIC® has determined on 3/28/2011 that this Technical Document has the Distribution Statement checked below. The current distribution for this document can be found in the DTIC® Technical Report Database.

☑ **DISTRIBUTION STATEMENT A.** Approved for public release; distribution is unlimited.

☐ **© COPYRIGHTED.** U.S. Government or Federal Rights License. All other rights and uses except those permitted by copyright law are reserved by the copyright owner.

☐ **DISTRIBUTION STATEMENT B.** Distribution authorized to U.S. Government agencies only (fill in reason) (date of determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT C.** Distribution authorized to U.S. Government Agencies and their contractors (fill in reason) (date determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT D.** Distribution authorized to the Department of Defense and U.S. DoD contractors only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT E.** Distribution authorized to DoD Components only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT F.** Further dissemination only as directed by (insert controlling DoD office) (date of determination) or higher DoD authority.

Distribution Statement F is also used when a document does not contain a distribution statement and no distribution statement can be determined.

☐ **DISTRIBUTION STATEMENT X.** Distribution authorized to U.S. Government Agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoDD 5230.25; (date of determination). DoD Controlling Office is (insert controlling DoD office).

Abstract

This effort focused on improving transportation security by making the objects (e.g. containers) being transported active agents in their own protection. Here, the objects are equipped with sensing and communication capabilities and are able to determine and communicate their sense of security throughout the dynamic transportation chain in a distributed manner. As part of the project, we have developed several data mining algorithms to enable intelligent agents to detect changes from their environment state. We have also designed algorithms to allow agents to communicate with each other for enhancing safety for the group of agents. Research issues of the designed algorithms have been applied to the Transportation Security SensorNet(TSSN) real transportation chain. We have tested all the new algorithms on real sensor data collected from transportation sensor network environments. The results have demonstrated the effectiveness of these algorithms for wireless sensor network security applications and provided useful insights regarding the challenges of the anomaly detection problem for distributed security in challenging environment.

Table of Contents

List of Figures

List of Tables

1. Overview

Our objective is to develop and test a fully distributed security monitoring systems, with wireless sensors, for transportation chains. In deriving a data-driven platform monitoring object security in transportation systems, we formalize the problem as an anomaly detection problem with data collected by wireless sensor networks. An anomaly in this report is defined as an object's observation of an event such as theft that deviates from the historical pattern or current consistent state in a group of sensors. There is a wide range of applications of anomaly detection in sensor network, for instance: ensuring the safety state of buildings such as bridges [19], and monitoring a parking lot [18]. In this report, as a driving application, we work on transportation chain security. Each physical object such as a container carries a computer and a wireless sensor monitoring the physical and environmental attributes of the object. Examples of such attributes include the object's moving acceleration and environment temperature among others.

Different from simple signature techniques to provide security for objects, our new approaches provide an active mechanism to endow objects with the ability to determine and communicate their sense of security in a dynamic environment. We investigate several approaches for performing fully distributed anomaly detection and examine their application to the task of transportation chain security. The goal behind our approaches is to embed anomaly detection techniques in each individual sensor node to endow them with the ability to determine their own security. The algorithms used should be able to automatically learn the "normal" concept, since coming up with rules from the perspective of the particular set of sensors would be difficult if not impossible in a dynamic environment. Also, to support real-time operations the algorithms should have the capability to continuously learn and adapt while running (online), to account for concept drift present in transportation environments. Furthermore, sensor nodes typically have limited resources in terms of memory and processing power, hence algorithms that can handle resource constraints and operate in an online training fashion are preferable.

Additionally, by allowing fully distributed security, we want to utilize sharing of information between objects, for instance, sensor nodes associated with cargo in the same container in the transport chain with capabilities to communicate with each other about their observations or sense of security. In this way the objects can work together and develop as a team to achieve greater security. The security is still fully distributed since each node can make predictions of its own and send alarms. The communication ability is essential since in many cases, a change is considered an anomaly only when it deviates from the consistency of the group. For example, a case in which all the objects in a container start to move because the container starts to move is usually an appropriate or safe situation although they are individually experiencing the change of location and velocity. Furthermore, we are able to form predictions for the network, or group, of sensor nodes by sharing opinions with nodes in the same group through communication.

Based on whether data samples are labeled or not, our approaches fall into two categories: supervised anomaly detection and unsupervised anomaly detection. In the domain of supervised anomaly detection, we focus on one class support vector machine and artificial immune systems [14]. We have investigated these two approaches by performing fully distributed anomaly detection and examining their application to the task of transportation chain security. After each object made a decision about its security, we considered three different ways of forming an overall group prediction in a distributed manner to achieve greater security: a baseline approach of individual detections, an event passing approach and an anomaly indication passing approach. For unsupervised anomaly detection, we considered distributed tracking

combined with Joint Space Principal Component Analysis (JSPCA) [13]. In short, each object uses local monitors that maintain parameterized sliding filters. These sliding filters detect local changes indicating potential unsafe states by simple threshold methods and share the changes among the objects. JSPCA uses the shared information to determine the consistent (normal) state and makes a final decision whether it is an unsafe state. In addition, at the same time as detection, JSPCA is able to identify the root object(s) where the anomaly occurred. All the approaches have been tested in the transportation security environment and experiment results have demonstrated the effectiveness of our approaches.

In addition to the general approaches and implementations of algorithms for performing anomaly detection and communication directly [14, 16, 13], we have pursued research directions addressing the challenges of applying these learning approaches to real world systems. These challenges include potentially large-scale applications (large sensor networks for monitoring large groups of objects) across various modes of transportation. Part of the challenge of making the sensor network monitoring a reality is anticipating the aspects of the data resulting from the future real-world and large-scale implementation that should be considered in constructing predictive models from and for the sensor data. In our work, we have identified several key research aspects for the real-world implementation of the distributed security sensor networks. Two of these aspects we consider are incorporating additional information about the sensor network structure (topology in the dimensions being sensed) [1, 5, 6], and addressing the dynamic nature of the sensor networks [15].

The work of data analysis fits into the sensor network architecture being developed by the SensorNet group at the University of Kansas. In Figure 1, we show the system design for TSSN. Here cargo monitoring operates over a mobile rail network (MRN) which in turn communicates with a virtual network operations center (VNOC) which handles transmitting alarm events and interfacing with the trade data exchange (TDE). TDE supplies information such as shipment data. The sensor anomaly detection fits into the sensor level of the MRN, responsible for detecting events which can then be reported and handled by higher levels, aside from any local action.

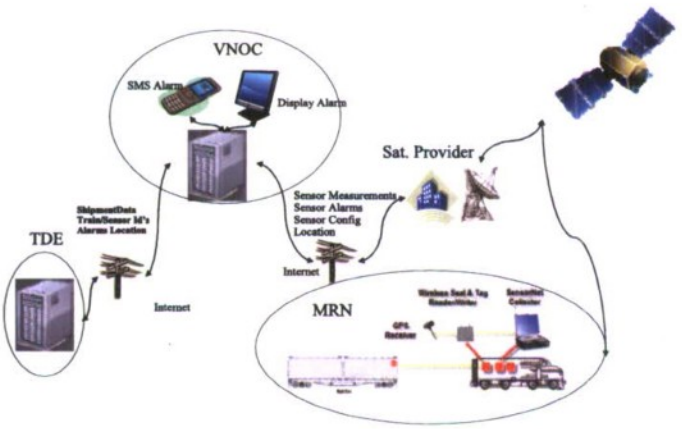

Figure 1. Transportation Security SensorNet Implementation

2. Supervised Anomaly Detection Techniques with Group prediction

The purpose of this work is to develop a detection-communication intelligent agent system. Distributed data mining algorithms are embedded into each agent to endow each one with the capability to learn its previous normal patterns and detect changes from such patterns. Once an agent detects a

change, a binary message will be broadcast to other agents and all of the agents work as a group to make a final decision if the change should be reported or not. Different decision rules can be designed and embedded in each agent to target different events. In the previous example when the train where a container is located starts to move, the container's location and velocity are changing. The changes would make the container agents assume that they are experiencing a potentially unsafe state, and would lead them to broadcast their individual decision to other agents. Since all the agents are experiencing the same change and broadcast the decisions within a specified time window the agents will be reassured and no anomaly reported.

We used three base methods for distributed online anomaly detection, a resource-constrained online one-class support vector machine(OCSVM), a real-valued artificial immune system(AIS), and a simple feature threshold approach for comparison[14]. We extended all three detection methods to an online setting. Then we considered three different rules of forming an overall group prediction in a distributed manner through communications that targeted different events, a baseline approach of individual detections, an event passing approach and an anomaly indication passing approach.

For the first base detection method, we use the one-class SVM (OCSVM) technique. OCSVM utilizes one-class learning techniques for SVM and learns a region that contains the training (normal) data instances (a boundary) [17]. Kernel functions, such as the radial basis function (RBF) kernel, are widely used to learn complex regions. For each test instance, the basic technique determines whether the test instance falls within the learned region. If a test instance falls within the learned region, it is declared as normal, otherwise it is declared as anomalous. We extended OCSVM to an online version by testing each new training instance with the current model; if the training case is misclassified, the case is used to update the current model along with the stored training instances [20, 3].

For the second base detection method, we use the real-valued artificial immune system for anomaly detection. A key component of AIS is the negative selection where the goal is to define nonself (anomalies) by randomly generating a coverage of detectors around the normal, or *self* region of the event space. We have extended the real-valued negative selection to an online learning algorithm. The approach incorporates aspects of real-valued negative selection [12] and the online components of an AIS framework [9]. The third detection method is a simple threshold approach, where the max. and min. value for each feature in the training data is stored, and an anomaly is detected if a value is exceeded in the test instance.

The three approaches we used for forming the group predictions from individual detectors are individual detection method(corresponding to group OR), group event passing and group indication passing. In Group OR, no information or opinion is shared between the nodes. An anomaly is counted for the whole group whenever an individual node detects an anomaly. In effect this corresponds to an "OR" operation of the current prediction for each sensor node at a given time point. Group event passing follows a similar approach as used in [2]. Here when a sensor node detects an anomaly, it sends the event vector for which the anomaly was detected to its neighbors, who test the event vector with their own trained models. When a specified threshold of positive detections over the total number of predictions is reached, the event is considered an anomaly. Group indication passing focuses on anomaly events that target individual objects in the group, for instance theft of a single package. If a threshold fraction of the other nodes (we use the majority vote in our experiments) also detect a change and broadcast messages within a specified time window the node will be reassured and no anomaly reported. Thus this method focuses on anomaly events that target individual objects in the group, for instance theft of a single package.

For further algorithmic details, please see [14].

2.1. Data Set

In the experiments, we tested the above algorithms on the sensor data collected for a short haul rail trial[7]. We used seven Sun Small Programmable Object Technology devices (or SPOTs) to collect data; six were in the box together and experimented on, and the seventh was held separately as a control to gain an idea of baseline noise. The container with the SPOTs was placed on the floor of the engine compartment, near the experimenters traveling in the engine. Periodically throughout the trip, three events were performed in sequence, a SPOT was removed from the box, about ten minutes later, returned to the box, and ten minutes after that, the box was moved. Also during one such box movement event during the trip, the box is rotated by 90 degrees, and during another, each SPOT's orientation is changed in the box. This sequence of three actions was repeated six times throughout the trip. Acceleration data from 3-axis accelerometers was used for the anomaly detection.

2.2. Experimental Results

First we obtained the results for detecting both object removal, an event affecting individual objects, and box movement, an event affecting the entire group of objects. We generated receiver operating characteristic (ROC) [14] curves for three anomaly detection algorithms: the resource-constrained online one-class SVM, the AIS method with a buffer, and the feature threshold method, for three group prediction approaches: individual detector ("Group OR") and two group methods ("Event Passing" and "Event Passing v2") used in concert with each base anomaly detection method. In an ROC curve the true positive rate (TPR), given as the total number of true detected anomaly events over the total number of anomaly events, is plotted against the false positive rate (FPR), which is the total number of incorrect detected anomaly events over the total number of received non-anomaly event strings from the sensor data, as a tuning parameter that trades off between these two goals is varied. The area under the curve (AUC) is usually taken as a summary of the ROC curve with a score closer to one indicating better performance of the algorithm.

Based on the experiment study, we had the following observations. First, three different detection algorithms (online one-class SVM, AIS and feature threshold) have similar performance achieving a perfect (area under curve of 1) ROC score. Second, in each figure, we found that the individual detection method (corresponding to "Group OR") was able to perform as well as group event passing methods, "Event Passing" and "Event Passing v2" (corresponding to the mean re-centering version), except for the simple feature threshold approach which had slightly less than an AUC of 1 (corresponding to 3 false positives at perfect true positive rate), but not enough of a difference to be significant. The group event passing methods, "Event Passing" and "Event Passing v2" (corresponding to the mean re-centering version), also both allowed perfect performance and were both able to correct the slight false positive rate of the feature threshold approach. In order to assess the benefits that the event passing methods produce, we have performed an analysis of the true positive and false positive reduction. Even though ideal performance was attainable without event passing for most methods, we found event passing generally decreased the false positive rate with only slight reduction in true positive rate, and tended to increase the number of tuning parameter values for which ideal performance was obtained, thus potentially contributing to robustness. We summarize the reduction in Table 1, showing the reductions averaged over all methods, using event passing. More details, results and analysis can be found in [14].

In this work, we designed a system where physical objects are able to determine their sense of security

Table 1. Averaged Performance Changes from Using Event Passing

	FPR Reduction	TPR Reduction	Increase in # Perfect Scores
Avg.	0.1458	0.0287	1.3
Std. Dev.	0.2403	0.0526	1.4181

in a distributed manner and then communicate such knowledge with other objects. Final decisions of safety would be made through consistency of information combined with sensor observation of their environment. The three embedded data mining techniques we investigated can learn a local consistency from the previous data and detect changes from the historical patterns. Once a change is detected, the potential unsafe agent would share its sense of local security with other agents and then three rules are applied to examine the group consistency. Experiments on the data set collected from rail trial demonstrate the effectiveness and feasibility of the two supervised learning approaches we used. The communication between agents improves the security as well.

3. Unsupervised Anomaly Detection Techniques by Joint Sparse Principal Component Analysis (JSPCA) with Shared Information

In this section, we describe our work on a communication-detection intelligent agent system. Different from the detection-communication system in the previous section, each agent in this system determines the security of the whole environment directly based on shared observation of sensors. JSPCA anomaly detector is used to detect a global anomaly and identify the root agent of it. In order to reduce the communication cost, a local filter is embedded in each agent with a local constraint. Observations will be shared and stored in each agent only if the constraint is violated in a test instance. Otherwise the previous update will be used for further analysis of anomaly detection.

Huang [10] has pointed out that the detector only needs to have a good approximation of the state when an anomaly is near because the purpose is to catch anomalies deviated from consistent state, rather than to track ongoing states. He has proven that, with such an approach, we could achieve a false detection rates below 4% while the data sent through the group is reduced by more than 90%. In our work, we follow his result on approximated detection and develop a JSPCA detector for anomaly detection and identification.

JSPCA detects anomalies of the whole network by the construction of a normal and abnormal subspace. Here the normal subspace is a representation of group state consistency extended by the top few principal components, while the abnormal subspace is extended by the last few principal components. All these principal components are generated by data matrix decomposition, where the data matrix is generated from the shared information gathered in a period of time. Anomalies are detected by projecting the data matrix to the abnormal subspace and comparing the projection to some threshold. If the projection is greater than the threshold, an anomaly will be reported. Since the data used is coming from all the group members, JSPCA can detect the deviation from the group consistency directly.

Additionally, JSPCA can identify the root cause of anomalies together with an anomaly detection. As the analysis in [13] shows, there exists a mapping between the data from each node and the corresponding entries in principal components. Such a mapping can help us to identify the node that should be responsible for the anomaly. In particular, the projection to one principal component(PC) is a summation of all the node observations weighted by corresponding entries in this PC. If one entry of one PC

is zero, the corresponding node has no projection onto this direction(PC). If the entries in the same positions of all the PCs representing the abnormal subspace are zero, the corresponding node observation can be projected to the normal subspace completely. If that happens, we claim that the node is a normal one. Thus, our key insight of the anomaly localization is that once an anomaly is detected, we check the entries across the principal components in the abnormal subspace: if the entries in the same position across the abnormal space are (close to) zero, the corresponding node is an innocent node which is not responsible for the abnormal event. However, in most situations, we cannot observe a joint zero (sparse) across the abnormal subspace, if it is directly generated from PCA. For most cases, the jth entry in one PC is close to zero, while in another PC is a large absolute value. Furthermore, the noise is expressed in the abnormal subspace as well, which causes even more difficulty in achieving simultaneous zero entries. Therefore, using PCA directly in anomaly localization is not practical in most situations. In order to overcome the challenge to get a joint sparsity in the abnormal subspace, we propose joint sparse PCA (JSPCA). JSPCA is an extension of PCA with regularization to constrain the entries in the same position of principal components to share the same sparsity pattern. Sparsity can enforces the unimportant entries to be zero or close to zero, which releases the influence of noise. Thus, the abnormal node will be located by a series of greater value across the abnormal principal components. Therefore, we can efficiently localize the anomalous node(s).

3.1. Data Set

In these experiments, the sensor data was collected during a car trial along the campus of University of Kansas. Seven Sun SPOTs were fixed in separate boxes and loaded on the back seat of a car. During the trial, each sensor recorded the magnitude of accelerations along x,y,z axis, temperature and luminance with a sample rate 3.33Hz. To collect the data for the trip, the SPOTs were programmed to continuously read and aggregate the sensor value for each sensor. We used the overall acceleration $(x^2 + y^2 + z^2)^{\frac{1}{2}}$ as the feature to detect the designed anomalous events with our anomaly detection and localization algorithm. During the whole trip, we simulated box removal and replacement, box rotation and flipping. Each event was repeated twice, in approximately 5-minute intervals. The whole experiment lasted about 1 hour.

3.2. Experimental Results

In this section, we present the anomaly detection and localization performance of JSPCA and two other benchmark algorithms Eigen Equation Compression(EEC) [8] and the Stochastic Nearest Neighborhood(SNN) [11] on the Sun SPOT sensor data collected during the car transportation trial. EEC and SNN both localize anomalies by scoring each node based on the change of neighborhood graphs. For EEC, we clustered the whole network into 3 groups for each sensor. For SNN, the neighborhood graph size k is chosen as 2. More details about these two algorithms could be found in [8], [11]. In Figure 2, the abnormal scores computed by JSPCA are shown in the first row, and the second and third rows show the anomaly scores measured by EEC [8] and SNN [11] respectively. For each algorithm, we also give the results for three different events: removal and replacement of node 2 (in the first column), flipping of node 4 (in the second column), and rotation of node 6 (in the last column). X-axis represents different nodes (seven in total)and the y-axis shows the abnormal scores, which have been normalized to the range [0,1].

Based on the experiment study, we found that JSPCA was able to localize the abnormal node accurately. As shown, we have a clear contrast between the score value of abnormal nodes(close to one)

and normal nodes(close to zero). For example, in the event of node 2 removal and replacement (shown in the left figure of the top row), the abnormal score of node 2 is significantly higher than that of the other normal nodes. In the next two rows of Figure 2, we show the localization performance of EEC and SNN respectively. We found that these two methods have commonality with JSPCA: all of them are able to pick out the true abnormal nodes for three events. However, EEC and SNN introduce many true positives because the score values are close to each other. Compared with these methods, our algorithm is more effective to localize abnormal objects.

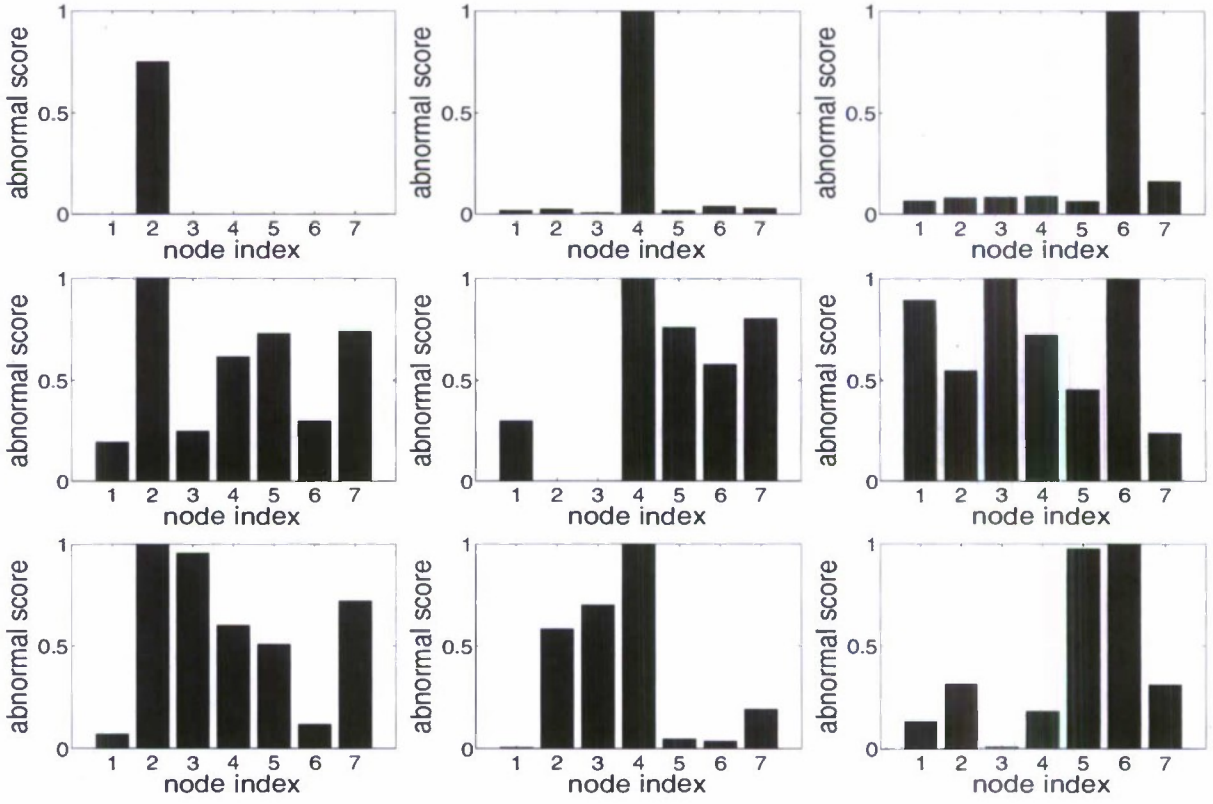

Figure 2. Comparison of JSPCA, Stochastic Nearest Neighbor and Eigen Equation Compression

We also tested this algorithm in other sensor data sets and demonstrated the stability of our method. See [13] for more detail.

In this work, we designed a detection algorithm JSPCA where physical objects were endowed with the ability to detect anomalies and further localize which objects have more responsibility for the detected anomalies. Different from the work in the previous section where the binary decisions(whether or not a change is detected) are made in each agent, the abnormal scores of each agent measure their contribution to a specific anomaly. Such algorithm helps us quickly localize the abnormal nodes and recover the abnormal situation.

4. Data Mining Foundation of Anomaly Detection in Wireless Sensor Networks

Aside from the above approaches and implementations directly applying to anomaly detection and localization in wireless sensor data [13, 14, 16], we have pursued research directions addressing the challenges of application of these predictive model learning approaches to real world systems, which

includes potentially large-scale applications (large sensor networks for monitoring large groups of objects) across various modes of transportation. As mentioned in the overview, our investigation of the challenges of anomaly detection consists of incorporating additional information about the sensor network structure (topology in the dimensions being sensed), and addressing the dynamic nature of the sensor networks.

4.1. Network Topology In Sensor Network Anomaly Detection

Our initial general approach to performing the anomaly detection for the sensor network data involved using classification models either globally on the sensor network or locally at each node [13, 14, 16]. However, building a model of predicted anomaly given sensed values only ignores an important source of information - the specific topological structure of the network. For example, for accelerometer data (e.g. measuring vibrations), we would expect physically nearby sensor nodes to sense similar values, for example as a person walks through an area with a senor network and walks close by a specific sensor node the accelerometer readings will be strongest in an area around where the person steps, so that nearby nodes will sense similar values; another example is temperature gradients. In general if we view our sensor network as a sensing field, sensed values will tend to vary smoothly over the field. Thus we investigated the idea that we may be able to improve the accuracy of predictive models if we incorporate this topological structure between the sensed values, or features, into the model inference process. We introduced the concept of using a feature graph, a graph between the features (corresponding to specific sensed values), that describes the topological structure of the features, and use it to bias the model learning under a regularized regression paradigm [1]. The idea is to use the feature graph Laplacian in a regularization term in the model optimization problem so that the model parameters vary smoothly over the feature graph. Afterward we extended the approach to the specific regression model of support vector machines as used in our general sensor network anomaly detection approach [14, 16], and incorporated feature selection to identify which values are the most important for the predictive model [6]. Motivated from [6], we investigated a boosting approach [5], in which a set of base learners are combined to achieve a more accurate prediction. In [5], a boosting algorithm considering structure information among base learners was proposed. The smoothness is imposed over the similar base learners. Suppose each sensor is an agent (running an anomaly detection algorithm individually) that can provide an opinion about the current status of the network, the algorithm investigates how to combine weak decision of each sensor into a stronger one. Also, by boosting on each learner, we can identify several sensors that contribute to the anomalies. The algorithm utilizes the topology among the sensors and assigns more weights to neighboring sensors if a sensor is suspicious about an anomaly. The method can be applied to decision fusion in sensor networks for anomaly detection.

We also investigate graph based anomaly detection, which is used for determining the whole state of the network. The assumption of this method is that most of the anomalies are caused by topological changes, such as node missing, unnecessary link and anonymous node invasion. For each time stamp, we model the network topology as a graph, in which each node represents a sensor; each edge represents a relationship between two nodes. The relationship can be captured by communication signal strength or sensor reading correlation. Our preliminary experiments show that the signal strength is very sensitive to sensor orientation and spatial position, hence the signal strength is not a good way to construct graphs, especially in a noisy environment. Correlation between sensor readings is a potential approach to build graphs. In [4], we evaluate the graph classification problem considering the internal structure of subgraph features using an L2 norm regularized kernel matrix.

Since our own sensor network data was limited to a small network and simulated experiments on a miniaturized scale, the utility of incorporating topology is not evident at this stage, so we chiefly validated our algorithms with a number of benchmark data sets on related types of tasks that had network structure between features. Subsequently this feature-graph regularization approach has become an emergent area of research in the data mining / machine learning communities. More details can be found in [5, 4, 6, 1]

4.2. Transfer Learning in Anomaly detection

Traditional machine learning/data mining approaches for learning predictive models from data rely on assumptions about near ideal data collection and generation conditions which are unrealistic for real world data such as sensor data. One typical assumption is that the collected data come from some fixed underlying distribution, i.e. that they are identically distributed. However this is generally not the case for sensor network monitoring data, in which the network itself, the environment it's in, and the nature of the monitoring tasks can be considered as dynamic entities that can change over time, and as a learned model is applied to different scenarios or specific sensor network systems. Thus we must be able to adapt learned predictive models for sensor networks to new situations as they arise, for example, as the network itself and the environment change over time, as the model is applied to different networks for potentially different but related monitoring tasks, as the normal and anomalous behavior the sensor network is monitoring changes, and as the sensor network moves between different modes of transportation. The process of transferring knowledge between data sources, e.g. from a set of collected supervised sensor data to a new monitoring situation, has been given the name transfer learning in the machine learning and data mining communities. We investigated the problem of transferring knowledge from one training data set to different test sets for which we did not yet have ground truth knowledge, i.e. the idea of adapting knowledge or a learned model from an initial training run for the distributed security sensor network to new situations and networks. We developed transfer learning algorithms based again on the regularized regression, and specifically support vector machine, paradigm [15]. The idea was to include a regularization term in the model learning optimization problem that would encourage the model to learn to generalize across data generating distributions - essentially introducing a bias toward a solution (model) that would allow knowledge transfer. We thus used a distribution distance measure in the regularization term that was efficient to implement and could be incorporated in the kernel learning framework, where the kernel is an inner product function implicitly mapping data points to a new feature space to allow nonlinear decision functions to be learned. Again since our own current sensor data is limited in complexity and scale [14] we validated our algorithms using a variety of data sets including benchmarks commonly used in the machine learning/data mining literature and data sets closely related to the anomaly detection tasks with shifting data generating distributions, for example detecting spam mail.

More recently we have been investigating an algorithm-free approach to transfer learning, by learning a feature embedding, which allows us to learn a common (feature) representation for different data sources while simultaneously addressing the heterogeneity among the data sources. Heterogeneous real-world data typically do not have all of the same features and missing values are common. For example, heterogeneous sensor networks are often made up of diverse combinations of sensor nodes that may have different sets of sensors, and sensor or transmission failure is common. The approach that we are investigating is to learn a set of basis that best explain the data, but also generalize across data generating distributions. More details can be found in [15]

5. Conclusions and Future Work

We have investigated several approaches to perform anomaly detection algorithms with sensor data for maintaining the security of transportation chains. We have evaluated these algorithms as a proof of concept on real sensor data collected during car and rail transport trials. Experiments on the data sets have demonstrated the feasibility of our approaches.

In the future, we will continue the investigation of JSPCA algorithm and an algorithm-free approach to transfer learning. The future work of JSPCA will focus on two directions. First, PCA approaches are limited to linearly correlated data and may fail to deliver optimal results when non-linear correlation exists between the sensors. Table 1 summarizes the key results obtained from our research. Second, integrating the network topology information of a sensor network to JSPCA is also a major direction in future work. Third, for the future work for transfer learning we plan to focus on learning a feature embedding such that common (feature) representation for different data sources will be available.

Acknowledgments

This work has been partially supported by an Office of Naval Research award N00014-07-1-1042.

References

[1] *Proceedings of the SIAM International Conference on Data Mining, SDM 2009, April 30 - May 2, 2009, Sparks, Nevada, USA*. SIAM, 2009. 5, 11, 12

[2] S. G. Cheetancheri, J. M. Agosta, D. H. Dash, K. N. Levitt, J. Rowe, and E. M. Schooler. A distributed host-based worm detection system. In *Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense*, New York, NY, USA, 2006. ACM. 6

[3] M. Davy, F. Desobry, A. Gretton, and C. Doncarli. An online support vector machine for abnormal events detection. *Signal Processing*, 86:2009–2025, 2006. 6

[4] H. Fei and J. Huan. L2 norm regularized feature kernel regression for graph data. In *CIKM*, pages 593–600, 2009. 11, 12

[5] H. Fei and J. Huan. Boosting with structure information in functional space: an application to graph classification. In *KDD*, 2010. 5, 11, 12

[6] H. Fei, B. Quanz, and J. Huan. Glsvm: Integrating structured feature selection and large margin classification. In *ICDM Workshops*, pages 362–367, 2009. 5, 11, 12

[7] D. T. Fokum, V. S. Frost, D. DePardo, M. Kuchnhausen, A. N. Oguna, L. S. Searl, E. Komp, M. Zeets, J. B. Evans, and G. J. Minden. Experiences from a Transportation Security Sensor Network Field Trial. Technical Report ITTC-FY2009-TR-41420-11, Information and Telecommunication Technology Center, University of Kansas, Lawrence, KS, June 2009. 7

[8] S. Hirose, K. Yamanishi, T. Nakata, and R. Fujimaki. Network anomaly detection based on eigen equation compression. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1185–1194, New York, NY, USA, 2009. ACM. 9

[9] S. A. Hofmeyr and S. Forrest. Architecture for an artificial immune system. *Evolutionary Computation*, 8(4):443–473, 2000. 6

[10] L. Huang, M. I. Jordan, A. Joseph, M. Garofalakis, and N. Taft. In-network pca and anomaly detection. In *In NIPS*, pages 617–624, 2006. 8

[11] T. Idé, S. Papadimitriou, and M. Vlachos. Computing correlation anomaly scores using stochastic nearest neighbors. In *ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, pages 523–528, Washington, DC, USA, 2007. IEEE Computer Socicty. 9

[12] Z. Ji and D. Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In *Genetic and Evolutionary Computation Conference (GECCO)*, volume 3102, pages 287–298. Springer, 2004. 6

[13] R. Jiang, H. Fei, and J. Huan. Anomaly localization by joint sparse pca and its implementation in sensor network. In *Sensor KDD*, 2010. 5, 8, 10, 11

[14] B. Quanz and J. Huan. Aligned graph classification with regularized logistic regression. In *Proc. 2009 SIAM International Conference on Data Mining*, 2009. 4, 5, 6, 7, 10, 11, 12

[15] B. Quanz and J. Huan. Large margin transductive transfer learning. In *CIKM*, pages 1327–1336, 2009. 5, 12

[16] B. Quanz and C. Tsatsoulis. Determining object safety using a multiagent, collaborative system. In *Environment-Mediated Coordination in Self-Organizing and Self-Adaptive Systems (ECOSOA 2008) Workshop*, Venice, Italy, October 2008. 5, 10, 11

[17] B. Scholkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001. 6

[18] H. Song, S. Zhu, and G. Cao. Svats: A sensor-network-based vehicle anti-theft system. In *INFOCOM*, pages 2128–2136, 2008. 4

[19] N. Xu, S. Rangwala, and et al. A wireless sensor network for structural monitoring. In *IN SENSYS*, pages 13–24, 2004. 4

[20] Z. Zhang and H. Shen. Application of online-training svms for real-time intrusion detection with different considerations. *Computer Communications*, 28:1428–1442, 2005. 6

Appendix B

SensorNet-III Requirements Document: Container Transportation Security Network