

Test Plan of the Anticipatory Wireless Sensor Network for the Critical Energy Infrastructure

Report Type: Topical

Reporting Period: 1 October 2004 – 31 December 2006

Principal Author: Dr. Carlos Rentel

CarlosHRentel@eaton.com

Report Issue Date: 6 September 2006

DOE Agreement # DE-FC26-04NT42071

Deliverable 1: Preliminary Performance Metrics Report

Eaton Corporation
Innovation Center
4201 North 27th Street
Milwaukee, WI 53216

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Abstract.

The test plan for the performance of the Anticipatory Wireless Sensor Network (A-WSN) is presented. The results of the test campaigns will be obtained after actual measurements are taken in the field with the Wireless Sensor Network developed by The Innovation Center - Eaton Corp., and the Anticipatory algorithms developed by ORNL.

Table of Contents

TEST PLAN OF THE ANTICIPATORY WIRELESS SENSOR NETWORK FOR THE CRITICAL ENERGY INFRASTRUCTURE	1
ABSTRACT.	3
EXECUTIVE SUMMARY.....	5
1.1 PART I: EXPERIMENTAL METHOD FOR A-WSN TEST.....	6
1.1.1 Format of log file for testing	8
1.1.2 Measurement Campaigns.....	8
1.2 PART II: EXPERIMENTAL METHODS FOR THE ANTICIPATORY ALGORITHM TEST.....	9
1.3 CONCLUSION.....	12

Executive Summary

The Anticipatory Wireless Sensor Network (A-WSN) system has been designed, modeled and implemented for the Anticipatory Threat Aware Application for Security of Critical Physical Energy Infrastructure. The design and modeling is based on the Unified Modeling Language (UML). Executable state machines were utilized to test the functionality of the system before actual development and deployment.

The A-WSN gathers data from passive infrared (PIR), acoustic and seismic sensors and sends this data to a coordinator node that is part of the network, and which is attached to a PC workstation via serial interface. The coordinator node forwards the data to the anticipatory threat aware application running in the same PC to which it is connected. We will deploy the system in an area of approximately 10 meters (30 feet) long and 3 meters (9 feet) wide. Hardware wise, the system is composed of one coordinator node connected to a PC workstation running the Anticipatory Threat Aware Application, at most four (4) extender nodes to increase the coverage of the network, eight (8) PIR/Acoustic sensor nodes and eight (8) seismic sensor nodes. The network is sub-divided in four (4) groups of two (2) PIR/Acoustic sensor nodes and two (2) Seismic sensor nodes each.

1.1 Part I: Experimental method for A-WSN Test¹

Each individual testing unit is called a testing campaign, which includes eight (8) hours of continuous field-testing.

During each testing, data sent from the network is logged in the PC as a text file for later post-processing and also it is available via sockets to a client in real-time. Other information, such as START/STOP/CONFIGURE/ACKS responses and commands sent to/from the nodes are also logged. All the logged information is time stamped for further analysis. To correlate the data gathered from A-WSN and real physical events, all the information of physical events needs to be time stamped and recorded manually. The relevant physical events include, but are not limited to: intentional targets (e.g., human beings, vehicles, etc.), unexpected human beings or vehicles, weather change. The physical and logical topologies need to be documented as well.

The key results we want to obtain from the testing scenarios are:

- **Functionally:** Whether the wireless sensor network is capable of capturing the relevant information from the target area so that the data is sufficient for the post-testing data processing.
- **Performance:**
 - Maximum network throughput without significant congestion and/or data loss.
 - False negative rates and false positive rates.
 - Data efficiency – number of transmissions for each target per unit of time.
 - Effect of system/sensor parameters and anticipatory algorithms on the target detection/classification performance.

Each testing campaign will be conducted with various system/sensor parameters to study the impacts of these parameters on system performance. Various offline algorithms with various parameters will be applied on the data gathered from the WSN to find the optimal processing algorithm and the optimal parameters. The network throughput, data delivery ratio, false negative/positive rates and data efficiency also need to be recorded. Attached is a template for recording all the relevant information during the post-test data analysis and correlation.

¹ Source: Eaton Corporation. Innovation Center. Contact: CarloshRentel@eaton.com

A-WSN Testing Campaign Worksheet (Conductor Name: _____)

CAMPAIGN NUMBER									
PURPOSE & SPECIAL SETTINGS									
Duration	From (YYYY:MM:DD:hh:mm:ss)				To (YYYY:MM:DD:hh:mm:ss)				
Target Area	Same as previous campaign (y/n)? if "y", campaign number: If "n", fill in right:			Location		Length (unit)		Width (unit)	
Log File	Name:								
Topology	Same as previous campaign (y/n)? if "y", campaign number: If "n", fill in below:								
	Node ID	Role (C/R/PA/S)	Location x	Location y	Relay ID	Direction	Other info:		
Targets	Type	Time	Size	Duration	Entry Location x	Entry Location y	Exit Location x	Exit Location y	Other Info
Other Incidents									
Weather Condition									

1.1.1 Format of log file for testing

All messages sent to and from the network needs to be recorded with timestamps. The timestamp has a YYYY:MM:DD:HH:MM:SS format of the local time. Each message constitutes an individual line. The timestamp for the message goes first in the line, then the message enclosed in < >. The message is recorded as raw hexadecimal bytes separated by colons. By this means, we can reconstruct the sequence of events in terms of messages with a precision of seconds.

1.1.2 Measurement Campaigns

- Maximum Throughput Campaign

Deploy the system with 4 relay nodes, 8 PIR/Acoustic sensor nodes, and 8 Seismic sensor nodes. Disable the sensors and send data with predefined data rates. Increase the attempted data rate for each sensor node from 1 packet per second to 50 packets per second. For each data rate among 1, 2, 3, 4, 5, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45 and 50 packets per second, Log the messages received.

Do offline analysis of the correlations of the attempted data rates and the actual delivery rates. Find the knee point where the network congestion starts to deteriorate the network delivery rate. To obtain an optimal data rate, this value should be taken into account, as well as energy consumption/lifetime constraints.

- Operational/Functional Campaign

These measurements are for validating the proper functionality of START/STOP/CONFIGURE operations. The sensors are enabled while their sensory data is ignored. The topology remains the same as the Maximum Throughput Measurements. The operator manually sends START/STOP/CONFIGURE commands from the PC workstation through the Graphical User Interface (GUI). All the messages from/to the PC workstation are logged.

Do offline analysis of the percentage of expected responses and qualitatively observe whether START/STOP commands properly start or stop the system.

- Control Target Campaign

These measurements are to validate the performance and efficiency of the system sensing targets for each type of sensors. Data is also gathered for post-testing analysis. These measurements need to be conducted in a highly controlled fashion. No objects that may be potentially detected by the sensors should appear in or close to the target field. All sources of noise or potential noise need to be properly identified and documented in detail. The weather condition, including temperature, humidity, wind among others, needs to be documented. Bushes, grass and other plants in or close to the target area also need to be documented for analysis of noise sources and effects. Unexpected incidents, such as animals crossing the target area, need to be documented with time, location and duration.

Two types of targets will be sent through the target area: human beings will walk through the field and vehicles will drive through the field. A timing device such as a stopwatch properly synchronized with the PC clock will be used for correlating the actual targets and sensory data offline. Each target needs to cross the target area along the long side and along the short side from different entry points. The intervals between two rounds of target crossing the area need to be long enough for the sensors to go back to a stabilized “no target” state. Shorter intervals will also be tested to figure out the effect of shorter intervals on the target detection performance. The speeds of human beings are computed from the distance he or she goes and the duration. The speeds of vehicles are read from the speedometer on the vehicle.

More complicated scenarios may also be conducted, which include:

- Two human beings moving together
- Two human beings entering the system simultaneously but from opposite entry positions, and moving towards each other, then leaving each other to the other ends of the target area
- A human being and a vehicle moving together

1.2 Part II: Experimental methods for the Anticipatory Algorithm Test²

This project is based on using Bayes’ decision theory to derive various formulae for the classification and the parameter estimation of the given data. The Bayes’ rule is given as follows:

$$P(w_j/x) = p(x/w_j)P(w_j)/p(x)$$

The parameters are estimated using the following formulae:

$$\mu = \frac{1}{n} \sum_{i=0}^n X_i$$

$$\sum = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^t$$

From the parameters obtained the decision rule is designed to classify the given data. The designed decision rule is run on the testing data said using the pre-calculated parameters. After running the classification each data is classified as category 1 or category 2. The obtained results are compared with the available data and the error rate and the accuracy are calculated.

² Source: Oak Ridge National Laboratory. Contact: kurugantipv@ornl.gov and allgoodgo@ornl.gov

(a) Maximum Likelihood ratio

This ratio is given by the formula:

$$\frac{p(x/w_1)}{p(x/w_2)} = \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(w_2)}{P(w_1)}$$

In the above equation a zero-one condition of the maximum likelihood is assumed thus making $\lambda_{11} = \lambda_{22} = 0$ and $\lambda_{12} = \lambda_{21} = 1$ and the prior probabilities $P(w_1) = P(w_2) = 0.5$. Since there are only two classification the probability of each is taken as the same. Thus from the given it makes sense that if $p(x/w_1) > p(x/w_2)$ then the sample falls under category one. To find the probabilities the parameters with the corresponding data is plugged in to the following Gaussian equation.

$$p(x) = \frac{1}{\sqrt{\sum (2\pi)^{d/2}}} \exp\left(-\frac{1}{2}(x - \mu)' \sum^{-1}(x - \mu)\right)$$

with two values of μ and Σ the probabilities are calculated and compared to be classified.

(b) Discriminant Function

The discriminant function provides us with another way of classifying the given data. In this the probabilities $g(x)$ are compared as an index for classification. The $g(x)$ is given by:

$$g_i(x) = \ln p(x/w_i) + \ln P(w_i)$$

if $g_1(x) > g_2(x)$ then the data falls under category one else it falls under category two. The discriminant function for normal density is given by:

$$g_i(x) = -\frac{1}{2}(x - \mu_i)' \sum_i^{-1}(x - \mu_i) - \frac{1}{2} \ln |\sum_i| + \ln P(w_i)$$

(c). Bayesian Estimation of parameters and classification.

In this the parameters μ and Σ are estimated again by treating this parameters as random variables themselves. The new formulae for the μ and Σ are as follows:

$$\mu_n = \left[\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \right] \left(\frac{1}{n} \sum x_k \right) + \left(\frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \right) \mu_0$$

$$\sigma_n = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$$

The values of μ_0 and σ_0 are randomly chosen values by observing the given data. However I chose μ_0 by looking at the mean-matrix of the whole given data in the testing set , similarly the σ_0 .

The classification is done based on the probabilities obtained by plugging in the Gaussian equation with the new parameters.

$$p(x) = \frac{1}{\sqrt{\sum_n (2\pi)^{d/2}}} \exp\left(-\frac{1}{2}(x - \mu_n)^t \sum^{-1}(x - \mu_n)\right)$$

K-Nearest Neighborhood Estimation:

In KNN estimation, the volume is allowed to vary depending on the training data set. To estimate the $p(x)$ from n samples, we can center a cell at x and let it grow until it contains Kn samples, and Kn can be some function of n . Ususally Kn is taken as square root of n . If we wish to determine the statistics, a hypersphere of volume V which just encloses k points from the combined set. If within that volume, Km of these points belong to class m , then we estimate the density for class m by :

$$p(x/w_m) = k_m / n_m \quad p(w_m) = n_m/n \quad p(x) = k/nV$$

The decision rule tells us to look in a neighborhood of the unknown feature vector for k samples. If within that neighborhood, more samples lie in class i than any other class, we assign the unknown as belonging to class i .

(e) Principle Component Analysis

Principle component analysis is done to eliminate the correlated feature vector , thus making the feature space into useful minimum to help in dimensionality reduction. The algorithm of the PCA is as follows:

The covariance matrix of th feature space is calculated.

The Eigen values and the corresponding Eigen vectors are calculated.

The Eigen value matrix is arranged in ascending order of magnitude correspondingly arranging the Eigen vector matrix.

The sub matrix of Eigen vectors is chosen to be multiplied with the data set depending on K where is k is obtained from the following formula

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^M \lambda_i} \geq 1 - \eta$$

where n is the loss of energy.

The obtained set of the Eigen vectors is multiplied with the training set to reduce the dimensionality of the data set.

(f) Cross validation and ROC curve.

The given data set is divided into subsets. Cross validation is the process of selecting one of these sets as testing set and the rest as training set. This operation is performed statistically significant times and each time noting down the obtained output and the expected output. The output can be classified as four cases True positive, True negative, False positive, False negative. The classification is best described by the following table.

Truth	Decision taken	
	+	-
	+	True Positive
-	False Positive	True Negative

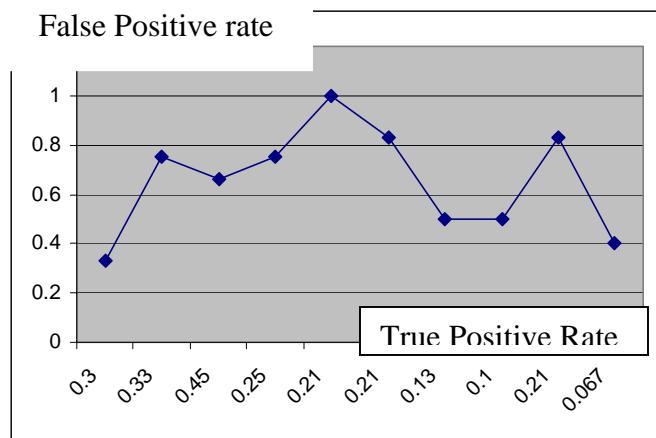


Fig 1. Typical ROC curve

The more the ROC curve tending towards the left hand top corner of the graph area the more efficient is the classifier.

(g) Other techniques that are considered for implementation are:

Independent Component Analysis for dimensionality reduction without loss of data abstraction
The features that are going to be used will be primarily statistical (mean, variance, skew, kurtosis) and wavelet-based

K-means, winner-take-all, kalman filter and kohonen feature maps are the other techniques that can be persuaded

The final algorithm would be a hybrid-approach purely based on the observations from trying out algorithms on the experimental sensor data. The ROC curves plotted on the data sets give a fair indication of the classifier accuracy.

1.3 Conclusion

The methods to test the performance and functionality of the A-WSN and the anticipatory algorithms were described. Actual results will be gathered from real data in the field (i.e., TVA sub-station). The main purpose is to verify the capability of the AWSN to detect simulated targets such as humans and vehicles in a target environment.