TARGET DETECTION IN SENSOR NETWORK USING A ZAMBONI AND SCAN STATISTICS

M. Guerriero, P. Willett

University of Connecticut
ECE Department
Storrs, CT 06269 USA

J. Glaz

University of Connecticut
Department of Statistics
Storrs, CT 06269 USA

ABSTRACT

In this article we introduce a sequential procedure for detecting a target using distributed sensors in a two dimensional region. The detection is carried out in a mobile fusion center (in a way familiar to hockey fans, we envision this as a *Zamboni* machine) which successively counts the number of binary decisions reported by local sensors lying inside its field of view. The proposed sequential detection procedure is based on a two-dimensional *scan statistic* – this is an emerging tool from the statistics field that has been applied to a variety of anomaly detection problems such as of epidemics or computer intrusion; but that seem to be unfamiliar within the signal processing community. Analytical and simulation results are presented for system-level detection.

***Index Terms*—** Sensor network; scan statistics; sequential detection.

I. INTRODUCTION

I-A. Sensors Networks with Mobile Agents

In [1] [2] the authors propose an appealing architecture for sensor networks in which information is accumulated in a traveling "rover," which sequentially queries the sensors that fall in its current (and changing) field of view. The Mobile Agent (MA), playing the role of the fusion center, takes the final decision about the presence of the target. There are sequential versions of this (e.g., [3]) that seek to use an MA to decide as quickly as possible which among a pair of hypotheses is true. However, the hypothesis is homogeneous: either H_0 is true at all sensors or H_1 is uniformly true. More realistically, a target is a *local disturbance*, and a MA-based anomaly detector is what is sought. Here we provide one: a sequential test in two dimensions via a *scan statistic*.

I-B. Scan Statistics

The detection and analysis of clustering of events is of great importance in many areas of science and technology including: epidemiology [4], bionformatics [5], biosurveillance [6], ecology [7], medicine [8], quality control and reliability [9]. Theory and applications of scan statistics, as well as recent advances, have been presented in [10], [11] and [12]. Since in this article we employ a scan statistic for observed counts of signals in a two dimensional region, we introduce below a two dimensional scan statistic.

THIS RESEARCH WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT N00014-07-1-0055.

Assume that a two-dimensional rectangular region $R = [0, T_1] \times [0, T_2]$ is under surveillance. Let $h_i = \frac{T_i}{N_i} > 0$, where N_i are positive integers, $i = 1, 2$. For $1 \le i \le N_1$ and $1 \le j \le N_2$, let X_{ij} be the number of events that have been observed in the rectangular subregion (cells) $[(i-1)h_1, ih_1] \times [(j-1)h_2, jh_2]$. We are interested in detecting unusual clustering of these events under the null hypothesis that X_{ij} are i.i.d. Bernoulli random variables with $P(X_{ij} = 1) = p_0$, where $0 < p_0 < 1$. For $1 \le i_1 \le N_1 - m_1 + 1$ and $1 \le i_2 \le N_2 - m_2 + 1$ define:

$$Y_{i_1,i_2} = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} X_{ij} \qquad (1)$$

to be the number of events in a rectangular region comprising m_1 by m_2 adjacent rectangular subregions with area $h_1 h_2$ and the south-west corner located at the point $((i_1-1)h_1, (i_2-1)h_2)$. If Y_{i_1,i_2} exceeds a preassigned value of k, we will say that k events are clustered within the inspected region. We define a *two-dimensional discrete scan statistics* as the largest number of events in any m_1 by m_2 adjacent rectangular subregions with area $h_1 h_2$ and the south-west corner located at the point $((i_1-1)h_1, (i_2-1)h_2)$:

$$S_{m_1 \times m_2; N \times N} = \max \{Y_{i_1,i_2}; \; 1 \le i_1 \le N_1 - m_1 + 1, 1 \le i_2 \le N_2 - m_2 + 1\} \qquad (2)$$

We assume that $N_1 = N_2 = N$ and $m_1 = m_2 = m$ for simplicity. We also abbreviate $S_{m \times m; N \times N}$ to $S_{m \times m}$.

The scan statistics $S_{m \times m}$, defined in equation (2) are used for testing the null hypothesis of randomness that assumes the $X'_{ij}s$ are i.i.d. Bernoulli random variables with $P(X_{ij} = 1) = p_0$, where $0 < p_0 < 1$. For the alternative hypothesis of clustering we assume a rectangular subregion in which X_{ij} with $P(X_{ij} = 1) = p_1 > p_0$.

For small and moderate values of $P(S_{m\times m} \geq k)$ a quite accurate approximation has been derived in [13]:

$$P(S_{m\times m} \geq k) \approx$$
$$1 - \left[\frac{[P\{S_{m\times m}(m,m) \leq k-1\}]^{(N-m-1)^2}}{[P\{S_{m\times m}(m,m+1) \leq k-1\}]^{2(N-m-1)(N-m)}}\right]$$
$$\times [P\{S_{m\times m}(m+1,m+1) \leq k-1\}]^{(N-m)^2} \quad (3)$$

where:

$$P\{S_{m\times m}(m,m) \leq k-1\} = F_b(k-1; m^2, p_0) \quad (4)$$

$$P\{S_{m\times m}(m,m+1) \leq k-1\}$$
$$= \sum_{s=0}^{k-1} F_b^2(k-1-s; m, p) b(s; m(m-1), p_0) \quad (5)$$

$$P\{S_{m\times m}(m+1,m+1) \leq k-1\}$$
$$= \sum_{s_1,s_2=0}^{k-1} \sum_{t_1,t_2=0}^{k-1} \sum_{i_j=0, 1\leq j\leq 4}^{1} b_1(s_1)b_1(s_2)b_2(t_1)b_2(t_2)$$
$$\times p_0^{\sum_{j=1}^4 i_j}(1-p_0)^{4-\sum_{j=1}^4 i_j} F_b(x; (m-1)^2, p_0) \quad (6)$$

where for $i = 1, 2$,

$$b_1(s_i) = b(s_i; m-1, p_0) \quad (7)$$
$$b_2(t_i) = b(t_i; m-1, p_0) \quad (8)$$
$$b(j; N, w) = \binom{N}{j} w^j (1-w)^{N-j} \quad (9)$$
$$F_b(i; N, w) = \sum_{j=0}^{i} b(j; N, w) \quad (10)$$
$$x = min(k-1-s_1-t_1-i_1, k-1-s_2-t_1-i_2,$$
$$k-1-s_2-t_1-i_2, k-1-s_1-t_2-i_3,$$
$$k-1-s_2-t_2-i_4) \quad (11)$$

Further approximations for $P(S_{m\times m} \geq k)$ are discussed in great detail in [11].

The key point is that while the exceedance probability for one location of the window of (2) is trivial, and similarly for many non-overlapping windows it is easy, to obtain a most powerful GLRT we need to "drag" the window as with the Zamboni. The exceedance probability for (2) is not at all obvious for overlapping windows, but that is exactly what the scan statistics literature provides. The contribution and importance of scan statistics is their ability to do that, to provide an answer to the question: "How many counts constitute an abnormality?"

II. TARGET DETECTION USING TWO-DIMENSIONAL SCAN STATISTICS

II-A. Model

Let us consider a scenario, where a grid of M sensors is deployed in the sensor field, which we consider to be a square of area b^2. A mobile rover (*Zamboni* machine) travels across the sensor network, see Figure 1, collecting the local decision from sensors which lie inside its field of view, which we consider to be a square of size F_{ov}. Assuming we know the total number M of sensors, we divide the square into M small subsquares. The location of the sensor inside each small subsquare (*cell*) is known[1]. Let us denote with (x_s, y_s) with $s = 1, \ldots, M$ the coordinates of sensor s.

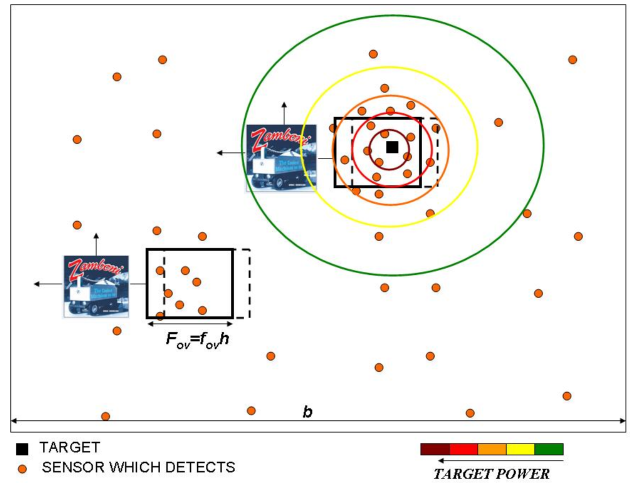

Fig. 1. The MA is depicted as a Zamboni machine and its field of view as a square of size F_{ov}. Only the sensors which detect the target are depicted. The different level curves correspond to a different target power. The Zamboni machine travelling across the sensor network simply counts how many sensors are inside its field of view. We assume that the target is uniformly distributed within the sensor network.

Noises at local sensors are i.i.d. and follow the standard Gaussian distribution with zero mean and variance σ_w^2:

$$w_s = \mathcal{N}(0, \sigma_w^2) \qquad s = 1, \ldots, M \quad (12)$$

We assume that sensors make their local decisions independently without collaborating with other sensors. We design each sensor s to decide between the following (composite) hypotheses

$$H_0 : r_s = w_s \quad (13)$$
$$H_1 : r_s = \frac{a_s}{d_s} + w_s$$

where r_s is the received signal at sensor s, a_s are i.i.d. GRV (Gaussian random variables) with zero-mean and variance σ^2 (σ^2 represents the power of the signal that is emitted by the target) and d_s is the distance between the target and local sensor s:

$$d_s = \sqrt{(x_s - x_t)^2 + (y_s - y_t)^2} \quad (14)$$

[1]Our results hold more generally: we do not need to know the exact location of the sensor inside the cell. We can also assume that the location of the sensor follows a uniform distribution in the cell.

and (x_t, y_t) are the unknown coordinates of the target. We further assume that the location of the target also follows a uniform distribution within the sensor field.

Assuming all the local sensors use the same threshold τ to make decision and according to the Neyman-Pearson lemma [14], we have the local sensor-level false alarm rate and probability of detection given by:

$$\begin{aligned} p_{fa} &= 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2}}\right) \\ p_{d_s} &= 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2 + \frac{\sigma^2}{d_s^2}}}\right) \end{aligned} \quad (15)$$

where $Q(\cdot)$ is the unit Gaussian exceedance function. We denote the binary data from local sensor s as $I_s = \{0, 1\}$ $(s = 1, \ldots, M)$. I_s takes the value 1 when there is a detection; otherwise, it takes the value 0.

II-B. Results

In Section II-A we introduced the sensor network and we modeled it as a square of area b^2. Letting $h = \frac{b}{N}$, where N is such that $N^2 = M$, we divide the square of area b^2 into M cells such that in each cell of area h^2 there is only one sensor. Let us denote by $c(i, j)$ the cell $[(i-1)h, ih] \times [(j-1)h, jh]$. We define X_{ij} as the binary data from the local sensor s inside $c(i, j)$ with $1 \leq i \leq N$ and $1 \leq j \leq N$. It is easy to verify that

$$\sum_{s=1}^{M} I_s = \sum_{j=1}^{N} \sum_{i=1}^{N} X_{ij} \quad (16)$$

For $1 \leq i \leq N$ and $1 \leq j \leq N$, let us denote with $X_{ij}^0 = I_s$ $(X_{ij}^1 = I_s)$ be the number of events that have been observed in the cell $c(i, j)$, under the hypothesis H_0 (H_1). X_{ij}^l, with $l = \{0, 1\}$, represents the binary data ("1" or "0") from the sensor inside the cell $c(i, j)$. It is straigthforward that X_{ij}^0 are i.i.d. Bernoulli random variables with $P\left(X_{ij}^0 = 1\right) = p_{fa}$, whereas X_{ij}^1 are discrete independent random variables but not identically distributed with $P\left(X_{ij} = 1\right) = p_{d_s}$ [2].

Now we assume that under H_1, p_{d_s} can be approximated with p_{fa} for a sensor s whose distance to the target is such that $d_s^2 >> \frac{\sigma^2}{\sigma_w^2}$. In fact we have that

$$p_{d_s} = 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2 + \frac{\sigma^2}{d_s^2}}}\right) \approx 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2}}\right) = p_{fa} \quad (17)$$

for $d_s^2 \gg \sigma^2/\sigma_w^2$. Since the target power attenuates as function of the distance from the target, we expect that there is a cluster of sensors which are stronger (closer to the target), under the hypothesis H_1. When mh, the window size in which the change of probability of a monitored event has occurred, is known, the scan statistics $S_{m \times m}$, defined in equation (2), is a generalized likelihood ratio test statistic for testing the above hypotheses [15]. We note, of course, that for us this region is fuzzily-defined (and probably round) according to (14); the question of the best field-of-view $F_{ov} = f_{ov}h$ size (or multiple sizes) to use is addressed in [16].

[2] p_{d_s} is greater than p_{fa}, but is otherwise unknown.

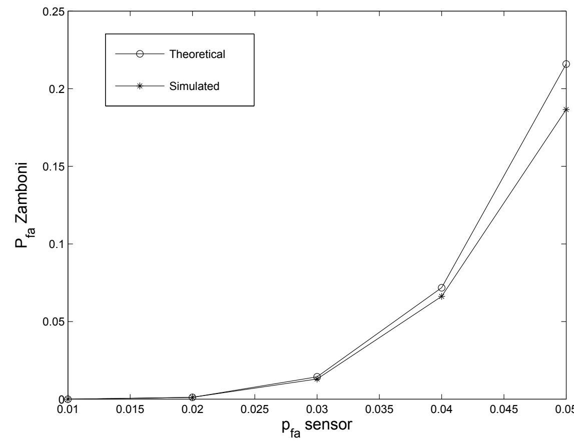

Fig. 2. Probability of false alarm for the zamboni P_{fa} versus probability of false alarm for the local sensor p_{fa}. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $M = 625$, $k = 7$, $f_{ov} = 6$ and $F_{ov} = 1.2$.

In Figure 2 we have plotted the *global* probability of false alarm P_{fa} for the Zamboni versus the *local* probability of false alarm p_{fa}. In Figure 2, the curves obtained by using the scan statistic approximations in equation (3) and those by simulations (based on 5000 Monte Carlo runs) are plotted. Figure 2 shows that the approximation in equation (3) is very accurate for small values of P_{fa} as we expected from the approximations of Scan Statistic theory.

Now we want to compare the performance of our model to that of the counting decision fusion rule, which is based on the total number of local detections from local sensors [17]. As proposed in [18], the decision is made by first counting the number of detections made by the local sensors in the surveillance region and then comparing it to a threshold T:

$$\Lambda = \sum_{s=1}^{M} I_s = \sum_{j=1}^{N} \sum_{i=1}^{N} X_{ij} \begin{cases} \geq T & \text{decide } \mathcal{H}_1, \\ < T & \text{decide } \mathcal{H}_0, \end{cases} \quad (18)$$

where $X_{ij} = \{0, 1\}$ is the local decision made by sensor s which is located in the cell $c(i, j)$.

In Figure 3, the ROC curves obtained by simulations (based on 1000 Monte Carlo runs) for the counting rule and the scan statistic test are shown. We can see that the ROC curves corresponding to the counting rule and those of the scan statistic test are pretty close and actually the scan statistic test seems to outperform the counting fusion rule. The ROC performances of the scan test are equivalent to those of the counting rule in the limit case when the size of

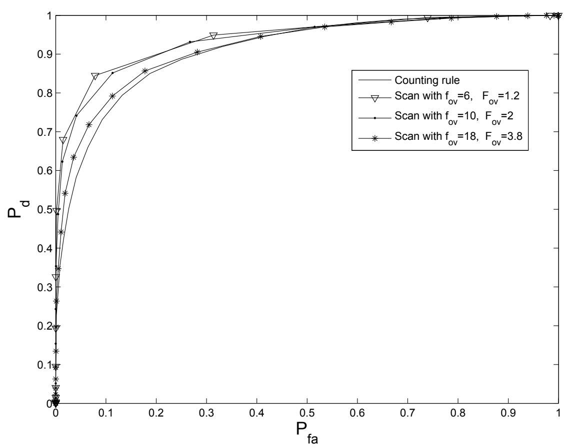

Fig. 3. ROC curves obtained by simulations for different values of the size of the field of view f_{ov}. Here we have $M = 625$, $SNR = \frac{\sigma^2}{\sigma_w^2} = \frac{1}{4}$. Simulation are based on standard Monte Carlo counting process, involving 5000 runs. .

the Zamboni field of view is equal to the size of the sensor network, that is $F_{ov} \rightarrow b$.

III. CONCLUSIONS AND FUTURE DIRECTIONS

We have proposed and studied a new sequential test, a two-dimensional scan statistics, that is based on the maximum number of detections reported by local sensors that the Zamboni observes in its field of view. We have shown how the system parameters can affect the detection performance. Numerical results, compared with the theoretical ones corroborate our analysis. An important result is that the proposed fusion rule outperforms the counting fusion rule, which requires the total number of detections in the sensor network. To the best of our knowledge we are not aware of any work on the sequential detection in two dimensions in sensor networks. Future work includes assuming that the sensors are in a Poisson field; and *Multiple Window Discrete Scan Statistics* [19] can also be used to improve the performance of the proposed detection procedure.

IV. REFERENCES

[1] Y. Sung, L. Tong, and A. Swami, "Asymptotically locally optimal detector for large scale sensor networks under the poisson regime," *IEEE Transactions on Signal Processing, XXX.*

[2] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," *Proceedings of 2003 MILCOM.*

[3] S. Marano, V. Matta, P. Willett, and L. Tong, "Cross-layer design of sequential detectors in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4105–4117, November 2006.

[4] K. P. Kleinman, A. M. Abrams, M. Kulldorff, and R. Platt, "A model adjusted space-time scan statistic with an application to syndromic surveillance," *Epidemiol. Infect.*, vol. 133, pp. 409–419, 2005.

[5] J. Hoh and J. Ott, "Scan statistics to scan markers for susceptibility of genes," *Journal of Computational and Graphical Statistics*, vol. 10, pp. 296–328, 2001.

[6] H. S. Burkom, "Biosurveillance applying scan statistic with multiple, disparate data sources," *J. Urban Health*, vol. 80, pp. 57–65, 2003.

[7] G. P. Patil and C. Taillie, "Upper level set scan statistic for detecting arbitrarily shaped hotspots," *Environmental and Ecological Statistics*, vol. 11, pp. 183–197, 2004.

[8] D. Q. Naiman and C. Priebe, "Computing scan statistic p-values using importance sampling, with applications to genetics and medical image analysis," *Proceedings National Academy Sciences USA*, vol. 97, pp. 9615–9617, 2000.

[9] J. Glaz, *Scan statistics. Encyclopedia of Statistics in Quality and Reliability*. New York, Wiley (in press), 2007.

[10] N. Balakrishnan and M. Koutras, *Runs and Scans with Applications*. New York: Wiley, 2002.

[11] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. New York: Springer-Verlag, 2001.

[12] J. Fu and W. Lou, *Distribution Theory of Runs and Patterns and its Applications*. Singapore: World Scientific Publishing, 2003.

[13] M. V. Boutsikas and M. Koutras, "Reliability approximations for markov chain imbeddable systems," *Methodology and Computing in Applied Probability*, vol. 2, pp. 393–412, 2000.

[14] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[15] J. Glaz and J. Naus, "Tight bounds and approximations for scan statistic probabilities for discrete data," *Ann. Appl. Probab.*, vol. 1, pp. 306–318, 1991.

[16] M. Guerriero, P. Willett, and J. Glaz, "Target detection in sensor network using scan statistics," *IEEE Transactions on Signal Processing*, submitted.

[17] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, no. 4, pp. 462–472, 2005.

[18] R. Niu, P. K. Varshney, M. H. Moore, and D. Klamer, "Decision fusion in a wireless sensor network with a large number of sensors," *Proc. 7th IEEE International Conference on Information Fusion (ICIF'04)*, 2004.

[19] J. Glaz and Z. Zhang, "Multiple window discrete scan statistics," *Journal of Applied Statistics*, vol. 31, no. 8, pp. 967–980, 2004.