The Mean-Variance Estimator Technique in Monitoring Applications using Mobile Agents over Wireless Sensor Networks

Marco Pugliese, Fortunato Santucci
Center of Excellence DEWS
University of L'Aquila
L'Aquila, Italy
marco.pugliese@ieee.org, fortunato.santucci@univaq.it

Abstract—**We propose a reliable technique to detect behavior anomalies in monitoring critical infrastructures through resource constrained devices, for instance wireless sensor networks (WSNs). The study is specifically targeted to monitoring and alerting functionalities for homeland security, that typically enforce severe requirements to the detection process. Assuming the behavior of the characteristic operation indicators in a potentially large and complex infrastructure (such as buildings, bridges, nuclear power plants, aircrafts, etc.) to be bounded by design constraints, we can introduce a novel non-parametric detection technique that we denote as "MV-estimator-based" (where MV stands for sample mean and variance): the sample mean and the sample variance are computed from observations and behavior classification is performed by defining regions in the MV-estimator space instead of the observations space. It will be shown that the novel detection technique is able to provide better performance with respect to other approaches over resource constrained platforms such as WSN, and this will be substantiated by numerical results as well as by a detailed cost analysis. Moreover MVET operations into a clustered WSN are presented where MVET distributed functions are implemented by using mobile agents.**

Keywords; Stochastic process, statistic estimator, monitoring observable, wireless sensor network, security, intrusion detection, mobile agent middleware.

I. INTRODUCTION

Homeland security and critical infrastructures monitoring (such as buildings, bridges, nuclear power plants, aircrafts, etc.) represent challenging application domains for modern networking technologies. Many daily operations currently rely on services that are provided by systems generally denoted as critical infrastructures [11]: in this context an emerging common feature of these infrastructures is their reliance on the widespread use of distributed information, communication and control functionalities, that are intended to provide more efficient and innovative services while meeting more challenging user requirements and expectations. In order to manage, supervise and control such complex, highly non-linear, and geographically distributed systems, targeted control systems named SCADA (Supervisory Control And Data Acquisition) are currently used. A SCADA system is composed of a single central room, where system intelligence is concentrated, and a number of RTUs (Remote Terminal Units) equipped with limited computational resources (e.g. wireless sensor units). RTUs communicate with the central room by sending to and receiving from it short control messages. Information associated to these control messages strongly depends on the processing capability in the single RTU. This paper proposes the adoption of Wireless Sensor Networks (WSNs) as a "network of RTUs" that are able to exploit their ad-hoc properties and to embed pervasive monitoring, networking and processing functionalities. Indeed, recent literature has addressed the perspectives of WSNs for monitoring structural and functional health of industrial plants (e.g. [1], [12]): nevertheless, we can observe that the dominating paradigm is to exploit WSNs features in terms of a "network of small sensors", while almost unexplored is the more advanced paradigm of "networked smart sensors" and the underlying opportunity to actually support autonomous (anomaly) detection processes. Motivated by the above remarks and taking into account computation constraints induced by explicit limitations in resources of WSN nodes, we have explored and investigated a novel detection logic for behavior anomalies, that is able to meet the following requirements: i) reliability – i.e. we aim to approach zero false negatives and a limited rate of false positives – and ii) light computation architecture. The underlying principle is based on the following remarks: the behavior of any "engineering manufacture", e.g. a critical infrastructure [11], can be modeled indeed as an "engineering manufacture", during its normal operation mode the behavior can be modeled as a sequence of values assumed by its indicators of interest (e.g. temperature, pressure) bounded by the system design constraints within a predefined tolerance [16]. We denote this model as the Engineering Manufacture Model (EMM) for a critical infrastructure. Any other behavior (induced by the overall effect from system anomalies or external attacks) is regarded as a deviation (either weak or strong) from this model. To estimate such deviations we will introduce the Mean Variance Estimator-based Technique (hereinafter simply MVET). We will show that it can outperform other traditional techniques when implemented over resource constrained devices like WSNs [4] [23], while still meeting the defined requirements. The remainder of this paper is organized as follows: in Sec. II the state-of-art about anomaly detection techniques applied to WSN in monitoring applications is briefly provided, along with pending issues and the rationales of our proposal; Sec. III and

MOBILWARE 2013, November 11-12, Bologna, Italy

DOI 10.4108/icst.mobilware.2013.254274

Sec. IV present the detection and estimation problems; Sec. V states the applicability conditions for MVET; Sec. VI describes the proposed technique as well as the security analysis results; Sec. VII introduces the finite memory MV-estimators; Sec. VIII estimates the bound rates for false negatives (FNR) and false positives (FPR) and states the appropriate conditions to use MVET; Sec. IX reports numerical results considering a practical case; Sec. X deals with computational and memory cost evaluations; Sec. XI reports a detailed comparison between MVET and other detection techniques widely exploited over WSN; Sec. XII described the MVET-based anomaly detection functional architecture to be mapped into implementation modules; Sec. XIII deals with MVET distributed operations into a clustered WSN and, lastly, Sec. XIV deals with future works and our next steps in developing the proposed research items.

II. Background and Motivations

Anomaly detection refers to the problem of finding patterns in data that do not conform to an expected behavior (Anomaly Detection Problem in Monitoring Applications). Several anomaly detection techniques can be found in the literature but a few of them can be considered viable for application to WSNs [4] where large amount of memory and computation resource are not available. Anomaly detection in sensor networks must capture either sensor fault detection or intrusion detection or both. A single sensor network might include sensors that collect different types of data, such as binary, discrete, continuous, audio, video, etc. The data is generated in a streaming mode. Often times the environment in which the various sensors are deployed, as well as the communication channel, induces noise and missing values in the collected data. Anomaly detection in sensor networks poses a set of unique challenges. The anomaly detection techniques are required to operate in an online approach and due to severe resource constraints, they need to be lightweight. Another challenge is that data are collected in a distributed fashion, and hence a distributed data mining approach is required to analyze the data. Moreover, the presence of noise in the data collected from sensors makes anomaly detection more challenging, since it has to distinguish between interesting anomalies and unwanted noise/missing values. Nevertheless techniques based on statistical models are promising candidates and non-parametric approaches are interesting as they do not rely on any a-priori knowledge about the underlying data distribution model. As stated in Sec. I, we introduce the MV-Estimator-based detection technique or MVET: the main difference with respect to other traditional non-parametric approaches (e.g. histogram-based [18]) is that detection analysis is performed in the MV-estimator space (or MVET space) instead then in the observations space. In other words defining regions for behavior classification in the MVET space appears more reliable than arguing directly in the observations space as sample mean and sample variance are subject to smaller oscillations. When compared to the histogram-based approach, pros for MVET are as follows: no need of counting frequency (therefore memory is saved), no need of bins (wrong bin sizes can greatly reduce histograms effectiveness). The main cons consist in the remark that introduction of new indicators would increase resource costs: nevertheless we can observe that a monitoring process is more similar to a classification problem rather than a point estimation problem, with the consequence that expressions for MVET estimators [16] can be simplified and reduced to computationally low cost algebraic forms.

III. The Anomaly Detection Problem

Let $\tilde{Q}$ and Q be stochastic processes and $\tilde{q}^k = \{\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_i, ..., \tilde{q}_k\}$ be the k-point data set in terms of predicted or expected values for the indicator q (or observable) and $q^k = \{q_1, q_2, ..., q_i, ..., q_k\}$ be the corresponding k-point data set in terms of actual observed values for that indicator. In general real observations differ from theoretical expectations, i.e. $q^k \neq \tilde{q}^k$. If process $\tilde{Q}$ represents the expected behavior of a certain system during its normal operation mode and process Q the observed behavior, we can assume to model this difference (which embodies all disturbs to process $\tilde{Q}$ from the environment such as anomalies or attacks) as an additive noise n such that $q^k = \tilde{q}^k + n^k$ for $\forall k$. The data set $n^k = \{n_1, n_2, ..., n_i, ..., n_k\}$ defines the stochastic process associated to that noise, say N. Correlation between processes $\tilde{Q}$ and N will be not zero to take into account a possible malicious attack engaged specifically against $\tilde{Q}$. From basic probability theory, we get $\mu_q = \mu_{\tilde{q}} + \mu_n$ and $\sigma_q^2 = \sigma_{\tilde{q}}^2 + \sigma_n^2 + 2\mathrm{cov}[\tilde{q}, n]$ for an additive noise where $\mu_{\tilde{q}}$ and $\sigma_{\tilde{q}}^2$ correspond to the expected mean and expected variance while μ_q and σ_q^2 to the observed mean and observed variance associated to the processes $\tilde{Q}$ and Q respectively [16]. Moreover the expected observable range defines the interval $[\tilde{q}_{min}, \tilde{q}_{max}]$ the data set $\tilde{q}^k = \{\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_i, ..., \tilde{q}_k\}$ is expected to lie into and the observed range defines the interval $[q_{min}, q_{max}]$ given by the boundaries for the observed data set $q^k = \{q_1, q_2, ..., q_i, ..., q_k\}$. According to this, we will solve the Anomaly Detection Problem as follows: let us sample the observables of interest, then link anomalies to a stochastic additive noise applied to samples and finally classify system behavior according to the strength of this additive noise. However, it will be shown that the point estimation of noise is not necessary if MVET is exploited.

IV. The Estimation Problem

An *estimator* is a measurable function used to infer the value of an unknown parameter in a statistical model. The construction and comparison of estimators are the subject of the estimation theory [16]: here we can just remind that the estimator at k-th observation step of a sample parameter is defined as the expected value of the parameter at the k-th observation step. From [16] the estimators for the mean and variance of process Q at k-th observation step are given by the sample mean μ_k and sample variance σ_k^2 at k-th step whose expressions are (computed iteratively from μ_{k-1} and σ_{k-1}^2 with $\mu_0 = 0$ and $\sigma_1^2 = 0$):

$$\begin{cases} \mu_k = \dfrac{(k-1)\mu_{k-1} + q_k}{k} \\ \sigma_k^2 = \dfrac{(k-2)\sigma_{k-1}^2 + (q_k - \mu_k)^2}{k-1} + (\mu_{k-1} - \mu_k)^2 \end{cases} \quad (1)$$

V. MVET Applicability Conditions

Before starting any description, the conditions for MVET applicability must be stated. MVET is a technique targeted to detect anomalies in an artificial system or, as we say in the following sections, an *engineering manufacture* during its normal operation mode where the involved variables can be likely known and managed and, therefore, much simpler to be modeled.

Engineering Manufacture Model (EMM). Any *engineering manufacture* during its normal operation mode, or *normal behavior*, can be externally modeled as a sequence of measurable *observables* (e.g. temperature, pressure, electric power, chemical agent concentration, and so on) fully representative of the system structure and function, the stochastic distribution of the observables through their *expected mean*, *expected variance*, the *expected observable range* and some predefined tolerance: these quantities and indicators constitute the EMM and identify the behavior of the system [16]. Any other behavior (induced by internal system anomalies or external attacks) is regarded as a deviation (weak or strong) from EMM. The following test verifies if MVET can be applied to monitor an engineering manufacture.

MVET Applicability Test. Let us denote the *expected mean* $\mu_{\tilde{q}}$ and *expected variance* $\sigma_{\tilde{q}}^2$ of the observable q derived from design constraints, let us set non negative tolerances for mean and variance M and V such that $M < |\mu_{\tilde{q}}|$ and $V < \sigma_{\tilde{q}}^2$ and let us set the *observed mean* μ_q to lie into $\left]\mu_{\tilde{q}} - M, \mu_{\tilde{q}} + M\right[$ and the *observed variance* σ_q^2 to lie into $\left]0, \sigma_{\tilde{q}}^2 + V\right[$. Given the *expected observable range* $[\tilde{q}_{min}, \tilde{q}_{max}]$, if both constraints $\tilde{q}_{min} \geq \mu_{\tilde{q}} - \min[M, D]$ and $\tilde{q}_{max} \leq \mu_{\tilde{q}} + \min[M, D]$ with $D \equiv \sqrt{\sigma_{\tilde{q}}^2 + V}$ are satisfied, then the monitored system conforms to EMM and MVET can be successfully applied. This result can be mathematically derived from statistical theory [16] but here can be intuitively explained as follows. Suppose the tolerance for mean M to be zero: this implies the observable should be $q_k \equiv \mu_{\tilde{q}}$ for any k during normal operations, or $[\tilde{q}_{min}, \tilde{q}_{max}] = \left]\mu_{\tilde{q}}, \mu_{\tilde{q}}\right[$; otherwise even for a single observation out of this range, e.g. $q_k = \mu_{\tilde{q}} \pm \varepsilon$ for some $\varepsilon > 0$, the *observed mean* μ_q would result different (indeed slightly) from the *expected mean* $\mu_{\tilde{q}}$, and therefore MVET would not be applicable. Now suppose $M = \sqrt{\sigma_{\tilde{q}}^2}$ and tolerance for variance V to be zero, hence $D = \sqrt{\sigma_{\tilde{q}}^2}$: in this case it is $[\tilde{q}_{min}, \tilde{q}_{max}] = \left[\mu_{\tilde{q}} - \sqrt{\sigma_{\tilde{q}}^2}, \mu_{\tilde{q}} + \sqrt{\sigma_{\tilde{q}}^2}\right]$ and even when observables were $q_k \equiv \mu_{\tilde{q}} + \sqrt{\sigma_{\tilde{q}}^2}$ for any k, the *observed mean* μ_q and the *observed variance* σ_q^2 would result $\mu_q = \mu_{\tilde{q}} + \sqrt{\sigma_{\tilde{q}}^2}$, hence $\mu_q = \mu_{\tilde{q}} + M$, and $\sigma_q^2 = \sigma_{\tilde{q}}^2$ and therefore MVET would be applicable; otherwise even for a single observation out of the range $[\tilde{q}_{min}, \tilde{q}_{max}]$, e. g. $q_k = \mu_{\tilde{q}} + M + \varepsilon$ or $q_k = \mu_{\tilde{q}} - M - \varepsilon$ for some $\varepsilon > 0$ and MVET would not be applicable.

VI. MVET Application

The procedure for the application of the MVET can be operatively described as follows:

- Determine the regions over MVET estimator space corresponding to the possible system behaviors (the Behavior Classification Theorem, BCT) and classify eventual anomalies;
- Define and apply the appropriate hazardousness weight (score) to each observation.

A. Behavior Classification

Definition. Low and Uncorrelated Noise Approximation (LUNA). As stated in Sec. III, we assume that the overall effect from anomalies and attacks to the monitored system is modeled as an additive noise n. If the conditions $|\mu_n| << |\mu_{\tilde{q}}|$, $\sigma_n^2 << \sigma_{\tilde{q}}^2$ and $|2\,\mathrm{cov}[\tilde{q}, n]| << \sigma_{\tilde{q}}^2 + \sigma_n^2$ hold true, then $|\mu_{\tilde{q}} - \mu_q| \approx 0$ and $|\sigma_{\tilde{q}}^2 - \sigma_q^2| \approx 0$ [16], i.e. the *observed mean* and *observed variance* can be confused with the *expected mean* and *expected variance* respectively. We denote this condition as *Low and Uncorrelated Noise Approximation* (LUNA). Using the estimators sample mean and sample variance for the process Q at the k-th observation step, the LUNA condition is verified if $|\mu_{\tilde{q}} - \mu_k| \approx 0$ and $|\sigma_{\tilde{q}}^2 - \sigma_k^2| \approx 0$.

The following theorem, we denote with *Behavior Classification Theorem* (BCT), determines the regions over the MV-estimator space corresponding to possible system behaviors:

BCT Part 1. Given the *expected mean* $\mu_{\tilde{q}}$ and *expected variance* $\sigma_{\tilde{q}}^2$ derived from design constraints, and set non negative tolerances for mean and variance M and V and assuming MVET applicability, if the observable q_k verifies the constraint:

$$\tilde{q}_{min} \leq q_k \leq \tilde{q}_{max} \quad (2)$$

then the corresponding sample mean and variance μ_k and σ_k^2 verify (the converse is not necessarily true) the condition:

$$\begin{cases} \left|\mu_{\tilde{q}} - \mu_k\right| \le M \\ 0 \le \sigma_k^2 \le D^2 \end{cases} \quad (3)$$

otherwise an anomaly is detected and an alarm must be generated. Moreover is $\left|q_k - q_{k-1}\right| \le 2\min[M, D]$. We indicate the *expected observable range* or the *Normal Range* for observables the interval $NR \equiv [\tilde{q}_{min}, \tilde{q}_{max}]$.

BCT Part 2. If (3) is true then the condition LUNA is verified (the converse is not necessarily true), otherwise an anomaly is detected and an alarm must be generated.

Proof Part 1. From (2) and MVET applicability, then $\begin{cases} q_k \ge \tilde{q}_{min} \ge \mu_{\tilde{q}} - \min[M, D] \\ q_k \le \tilde{q}_{max} \le \mu_{\tilde{q}} + \min[M, D] \end{cases}$ or $\begin{cases} \mu_{\tilde{q}} - M \le q_k \le \mu_{\tilde{q}} + M \\ \mu_{\tilde{q}} - D \le q_k \le \mu_{\tilde{q}} + D \end{cases}$. It is straightforward to see that if $\mu_{\tilde{q}} - M \le q_k \le \mu_{\tilde{q}} + M$ then $\mu_{\tilde{q}} - M \le \mu_k \le \mu_{\tilde{q}} + M$ (the converse is not necessary true), thus we can write $\begin{cases} -M \le \mu_k - \mu_{\tilde{q}} \le M \\ 0 \le (q_k - \mu_{\tilde{q}})^2 \le D^2 \end{cases}$ which coincides with (3) when using $D \equiv \sqrt{\sigma_{\tilde{q}}^2 + V}$. Moreover given $q_{k-1}, q_k \in NR$ we obtain $\max\left[\left|q_k - q_{k-1}\right|\right] = \tilde{q}_{max} - \tilde{q}_{min}$ when $q_{k-1} = \tilde{q}_{min}$ and $q_k = \tilde{q}_{max}$ or when $q_{k-1} = \tilde{q}_{max}$ and $q_k = \tilde{q}_{min}$; therefore $\tilde{q}_{max} - \tilde{q}_{min} \le 2\min[M, D]$. Q.E.D.

Proof Part 2. Starting from (3), we set $\begin{cases} \max\left[\left|\mu_k - \mu_{\tilde{q}}\right|\right] = M \\ \max\left[\left|\sigma_k^2 - \sigma_{\tilde{q}}^2\right|\right] = V \end{cases}$ and, by replacing the sample mean and variance with the *observed mean* μ_q and *observed variance* σ_q^2, obtain $\begin{cases} \max\left[\left|\mu_n\right|\right] = M \\ \max\left[\left|2\,\mathrm{cov}[\tilde{q}, n]\right|\right] = \left|V - \sigma_n^2\right| \end{cases}$. By definition is $M < \left|\mu_{\tilde{q}}\right|$ and if $\left|\mu_n\right| \le M$, then $\left|\mu_n\right| << \left|\mu_{\tilde{q}}\right|$; by definition is $V < \sigma_{\tilde{q}}^2$ and if $\left|2\,\mathrm{cov}[\tilde{q}, n]\right| \le \left|V - \sigma_n^2\right|$, as $\left|V - \sigma_n^2\right| < \left|\sigma_{\tilde{q}}^2 - \sigma_n^2\right| << \sigma_{\tilde{q}}^2 + \sigma_n^2$, then $\left|2\,\mathrm{cov}[\tilde{q}, n]\right| << \sigma_{\tilde{q}}^2 + \sigma_n^2$: both converse conditions are not necessary true. Inequalities $\left|\mu_n\right| << \left|\mu_{\tilde{q}}\right|$ and $\left|2\,\mathrm{cov}[\tilde{q}, n]\right| << \sigma_{\tilde{q}}^2 + \sigma_n^2$ verify the LUNA condition. Q.E.D.

Observation 1. BCT states that (BCT Part 1) if (2) is not verified for some observable, then (3) can be still satisfied and the LUNA condition is still verified and the system results only weakly altered: such alteration can be indication of just some malfunction during operations of the monitored system. If (3) is not verified (BCT Part 2) for some observable, then LUNA condition gets not verified and the additive noise which models anomalies and attacks cannot be considered low and uncorrelated anymore and the monitored system results strongly altered. A possible correlation between observations and something outside the system can be a serious indication of a malicious attack being engaged by an intruder.

Observation 2. The converse of BCT Part 2 becomes true, i.e. if LUNA condition is true then (3) is satisfied, when both $M \approx 0$ and $V \approx 0$, that is $M << \left|\mu_{\tilde{q}}\right|$ and $V << \sigma_{\tilde{q}}^2$: in this case, from truly of LUNA condition we can write $\begin{cases} \left|\mu_k - \mu_{\tilde{q}}\right| \approx 0 \\ \left|\sigma_k^2 - \sigma_{\tilde{q}}^2\right| \approx 0 \end{cases}$ hence $\begin{cases} \left|\mu_k - \mu_{\tilde{q}}\right| < \varepsilon' \\ \left|\sigma_k^2 - \sigma_{\tilde{q}}^2\right| < \varepsilon'' \end{cases}$ with ε', ε'' arbitrarily small positives. If we set $M \equiv \varepsilon'$ and $D = \varepsilon''$, the proof is shown. In other terms, more intuitively: if tolerances *M* and *V* are not small, i.e. $M \approx \left|\mu_{\tilde{q}}\right|$ and $V \approx \sigma_{\tilde{q}}^2$, from (3) their contribution is counted as additive noise so that system behavior could be erroneously estimated as anomalous and the corresponding observations as false positives: this occurrence will be important to compute the False Positive Rate is Sec. VIII. According to BCT and Observation 1, we can introduce the following classification for the different system behaviors:

- NORMAL (*NORM*) if both BCT Part 1 and Part 2 are satisfied;
- WEAK ALTERED (*W_ALT*) if BCT Part 1 is not satisfied and BCT Part 2 is satisfied;
- STRONG ALTERED (*S_ALT*) if both BCT Part 1 and Part 2 are not satisfied.

B. Hazardousness Scores

To quantify anomaly hazardousness, we introduce a weight, the *Hazardousness Score* HS_k at k-th observation step: when normal (NORM) behavior, by definition HS_k is null; otherwise when weak or strong altered, the corresponding weights $HS\_W\_ALT_k$ and $HS\_S\_ALT_k$ are defined as:

$$HS\_W\_ALT_k \equiv \begin{cases} \mathrm{int}\left[\dfrac{\left|q_k - \tilde{q}_{min}\right|}{\left|NR\right|}\right] & q_k > \tilde{q}_{max} \\ \mathrm{int}\left[\dfrac{\left|\tilde{q}_{max} - q_k\right|}{\left|NR\right|}\right] & q_k < \tilde{q}_{min} \end{cases} \quad (4)$$

$$HS\_S\_ALT_k = \mathrm{int}\left[\frac{\left|\mu_{\tilde{q}} - \mu_k\right|}{M}\right] + \mathrm{int}\left[\frac{\left|\sigma_{\tilde{q}}^2 - \sigma_k^2\right|}{V}\right] \quad (5)$$

From the above definitions it is clear that the higher the score is, the higher the estimated strength of the detected abnormality is: a score $HS\_W\_ALT_k = s$ indicates that the observable results s times out of its normal range; a score $HS\_S\_ALT_k = s$ indicates that at least an estimator is s times out its normal range, which is M for the sample mean and V for the sample variance.

C. Hazardousness Scores

For the sake of generality, we apply a normalization by the factor $Q \equiv \max\left[\left|\tilde{q}_{min}\right|,\left|\tilde{q}_{max}\right|\right]$: the *normalised observation* becomes $\bar{q} \equiv \tilde{q}/Q$, and therefore $\overline{NR} \equiv \left[\tilde{q}_{min}/Q, \tilde{q}_{max}/Q\right]$, $\bar{\mu}_{\tilde{q}} = \mu_{\tilde{q}}/Q$, $\overline{\sigma^2_{\tilde{q}}} = \sigma^2_{\tilde{q}}/Q^2$, $\overline{M} = M/Q$ and $\overline{V} = V/Q^2$. A compact representation for MVET data set is given by the 5-pla $\left\langle \overline{NR}, \bar{\mu}_{\tilde{q}}, \overline{\sigma^2_{\tilde{q}}}, \overline{M}, \overline{V} \right\rangle$.

VII. Finite Memory MVET

According to (1) MVET estimators at k-th step are functionally dependent on the past k observations, i.e. $\bar{\mu}_k = f(\bar{q}_1, \bar{q}_2, \cdots \bar{q}_{k-1}; \bar{q}_k)$ and $\overline{\sigma^2_k} = f'(\bar{q}_1, \bar{q}_2, \cdots \bar{q}_{k-1}; \bar{q}_k)$, thus the k-th estimation could still depend on past observations even if loosely correlated (in other words the temporal distance between observations becomes longer then the temporal correlation). We denote (1) as *infinite-memory MVET estimators*. However in (1) the $1/k$ factor for large k, i.e. very long observation times, tends to saturate estimators which get less reliable leading to possible false negatives. The examples in Sec. VIII will show quantitatively this occurrence. To cope with this problem, we introduce the *normalized finite-memory MVET estimators* $\bar{\mu}_{Tk}$ and $\overline{\sigma^2_{Tk}}$ (6) which depend only on last T observations: T defines the *MVET estimators memory length* (or *memory window*). Optimal values for T should not exceed the maximum temporal correlation between observations: therefore we define the *maximum estimator memory length* T_{max} the upper bound for T before MVET estimators get saturated, i.e. until transitions induced by a single observable could be still detected from *NORM* to *S_ALT* behavior (or vice-versa). The value T_{max} is found by solving (6) for T with the following constraints: $\bar{q}_k = \mu_{\tilde{q}}/Q$ for $1 \le k \le T_{max} - 1$ and $\bar{q}_k = 1$ for $k = T_{max}$ with the constraint $HS\_S\_ALT_{T_{max}} = 1$.

$$\begin{cases} \bar{\mu}_{k>T} = \dfrac{1}{T} \displaystyle\sum_{i=k-T+1}^{k} \bar{q}_i \\ \overline{\sigma^2_{k>T}} = \dfrac{1}{T} \displaystyle\sum_{i=k-T+1}^{k} (\bar{q}_i)^2 - \bar{\mu}^2_{k>T} \end{cases} \tag{6}$$

As known from literature, finite memory estimators result to be more reactive respect to infinite memory ones as long memory tails tend to get them into saturation and unable to work properly. Therefore in our numerical tests we will focus only on finite memory tools using memory values lower than T_{max}.

VIII. Evaluation of False Negatives and False Positives

False negatives and false positives rates depend not only on MVET capability but also, for instance, on a proper tuning of the memory window (events slower than memory could be not detected) or a good setting of the sensor unit: in other words, MVET should jointly operate with a diagnostic console for what concerns sensor equipment monitoring, and an intrusion detection module for what concerns WSN security surveillance. In [20] a specific IDS is proposed and it is tailored for network attacks against WSN, in [21] and [22] this IDS has been implemented to provide an application platform over WSN with security services.

False Negatives Rate (FNR). BCT Part 1 shows that *no alterations at all* are estimated until observations lie into constraints (2) and (3) or, in other words, an event is classified as anomalous only when constraint (2) or (3) gets not verified. However, this does *not* infer that *no attacks at all* have possibly been engaged against the system but, if this is the case, the intensity of these attacks has been resulted to be too weak to induce an abnormal behavior in the system. Moreover FNR also depends on the accuracy in the stochastic description of the observables and on a proper setting of the reference stochastic parameters, so that even weak attacks could be detected. Therefore MVET can lead to FNR values arbitrarily close to zero.

False Positives Rate (FPR). As noted above, the converse of BCT Part 2 can be true only for small M and V otherwise a situation potentially leading to false positives can occur: easy calculations from (3) and taking into account Observation 2, show that MVET assures an upper bound for FPR to be $\approx \max\left[M/\left|\mu_{\tilde{q}}\right|, V/\sigma^2_{\tilde{q}}\right]$, which is a manageable quantity by tuning the tolerance parameters M and V. Obviously returns that FPR gets negligible only if $M << \left|\mu_{\tilde{q}}\right|$ and $V << \sigma^2_{\tilde{q}}$.

IX. Numerical Results

Let us consider some observation sequences $\tilde{q}^k = \{\tilde{q}_1, \tilde{q}_2, ..., \tilde{q}_i, ..., \tilde{q}_k\}$ from a certain observable (uni-variate case) to monitor the behavior of a system EMM-compliant. Numerical tests related to MVET performance have been carried on adopting the following methodology: first we prove reactivity in detection to be better with finite memory estimators rather than with infinite memory ones, then we measure MVET performance for some typical behavior profiles using test observable sequences. Suppose the *Normal Range* to be $NR = [18,22]$ units, the *expected mean* to be $\mu_{\tilde{q}} = 20$ units, the *expected variance* to be $\sigma^2_{\tilde{q}} = 4$ units^2 and, from the design constraints, the tolerances *M* and V to be $0.1\mu_{\tilde{q}}$ ($10\%\,\mu_{\tilde{q}}$) units and $0.05\sigma^2_{\tilde{q}}$ ($5\%\,\sigma^2_{\tilde{q}}$) units^2 respectively: it is easy to check MVET applicability from Sec. VIII. We will exploit the normalized form of MVET with finite-memory. The 5-pla $\left\langle \overline{NR}, \bar{\mu}_{\tilde{q}}, \overline{\sigma^2_{\tilde{q}}}, \overline{M}, \overline{V} \right\rangle$ is set as follows: $\overline{NR} \approx [0.81, 1]$, $\bar{\mu}_{\tilde{q}} \approx 0.91$, $\overline{\sigma^2_{\tilde{q}}} \approx 0.008$, $\overline{M} \approx 0.09$ and $\overline{V} \approx 0.0001$. From Sec. VIII the upper bound for FPR is 10%. From (7) the *maximum estimator memory length* results to be $T_{max} = 22$ observation steps, the *MVET estimators memory length* $T = 20$ will be set. Red horizontal lines in any picture represent the boundary values for indicators in that picture.

Fig. 1, Fig. 2 and Fig. 3 show some curve families useful for our test: Fig. 1 represents three possible observable sequences where some "high-rate" abnormal event has occurred, Fig. 2 other three possible sequences in case of "slow-rate" events and Fig. 3 other three possible observable sequences in case of abnormal events estimated as "just" a little deviation (weak alteration) from normal behavior. Fig. 1a, Fig. 1b, Fig. 2a, Fig. 2b and Fig. 3a, Fig. 3b represent the corresponding values for the *hazardousness scores*, those for the weak altered case in diagrams labeled with a) and for the strong altered in diagrams labeled with b). Let's see Fig. 1.

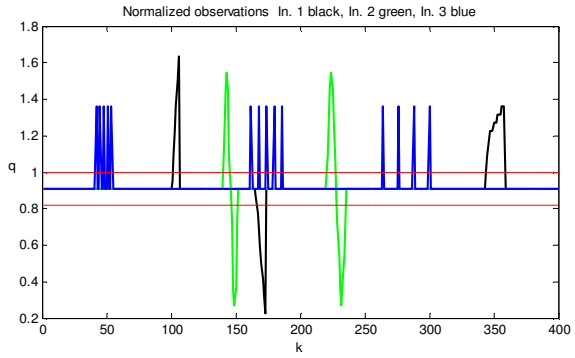

Figure 1. Normalized observation sequence from "high-rate" events

In black curve the observable values start from normal behavior and then increases with three sharp ramps: the first reaches $\approx 1.6|\tilde{q}_{max}|$ in 6 steps (≈ 3 times faster than memory window), the second $\approx 0.2|\tilde{q}_{max}|$ in 8 steps (≈ 2.5 times faster) and the third one $\approx 1.3|\tilde{q}_{max}|$ in 18 observation steps (comparable to T). The green curve is shaped like two "sinusoids" with "period" equal to 8 and 16 observation steps respectively. The blue curve is shaped as a sequence of spikes (i.e. a single abnormal value at $\approx 1.4|\tilde{q}_{max}|$) with repetition rates equal to 2, 3 and 6 observation steps respectively: these spikes can emulate an outlier event. MVET succeeds in detecting these "high-rate" events by sweeping its estimators (6) at any rate from 1 (i.e. the memory-less case) to $1/T_{max}$ and taking as reference value the maximum assumed in this rate range: black curves in Fig. 1a and Fig. 1b follow the same profile as those in Fig. 1.

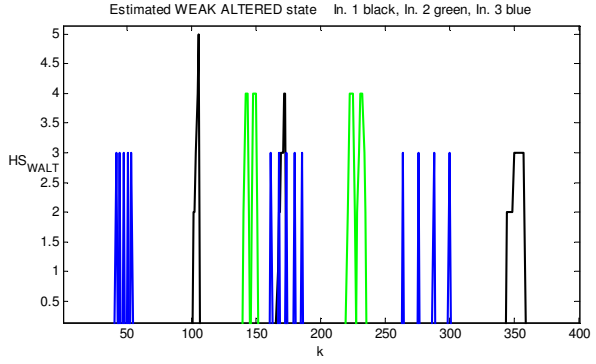

Figure 1a. Estimated scores corresponding to Weak Altered states

In particular, for the case of the spikes, it is remarkable that the corresponding scores result shaped as spikes too without delay with just a small spreading of ≈ 3 observation steps. This shows that MVET can meet the robustness requirement.

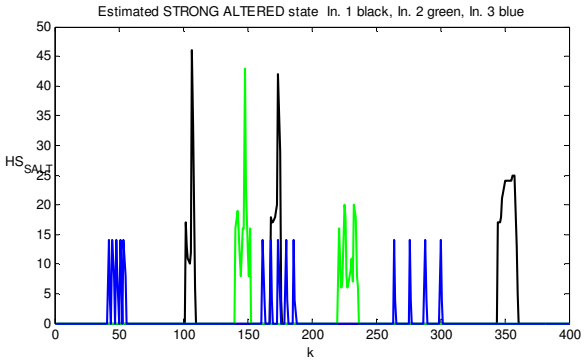

Figure 1b. Estimated scores corresponding to Strong Altered states

In Fig. 2 three observable sequences generated by "slow rate" events are shown.

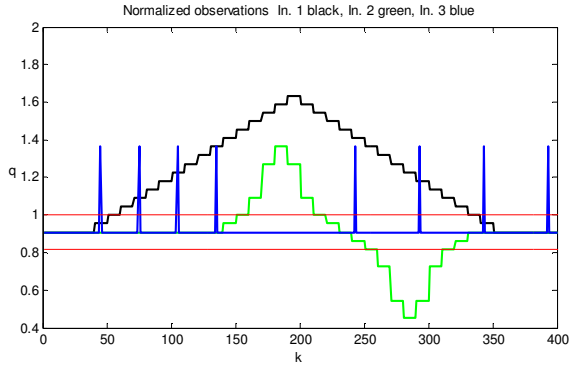

Figure 2. Normalized observation sequence from "low-rate" events

These curves can successfully test MVET capability in detecting events spanning a number of observation steps much longer than the memory window (remember we set $T = 20$), as shown by the corresponding scores represented in Fig. 2a and Fig. 2b. The black curve in Fig. 2 is shaped like a big "triangle" about $\approx 1.6|\tilde{q}_{max}|$ high spanning about 300 observation steps which represents the case of a low continuous linear increase and decrease for observable values: this can successfully test MVET capability in tracking without delay even this kind of events, as shown by the corresponding "triangle" shaping for the scores depicted in black in Fig. 2a and Fig. 2b. The green curve in Fig. 2 is shaped like a slow "sinusoid" with "period" equal to 190 observation steps (about 8 times slower than *T*) and "amplitude" $\approx 1.3|\tilde{q}_{max}|$, and the scores in Fig. 2a and Fig. 2b result marked correspondingly to the lobes out of the Normal Range. The blue curve is shaped as a sequence of spikes at $\approx 1.4|\tilde{q}_{max}|$ with repetition rates equal to 30 and 50 observation steps: also in this case scores follow these spikes without delay (blue curves in Fig. 2a and Fig. 2b).

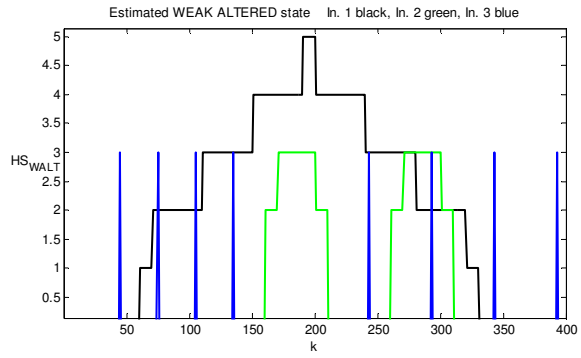

Figure 2a. Estimated scores corresponding to Weak Altered states

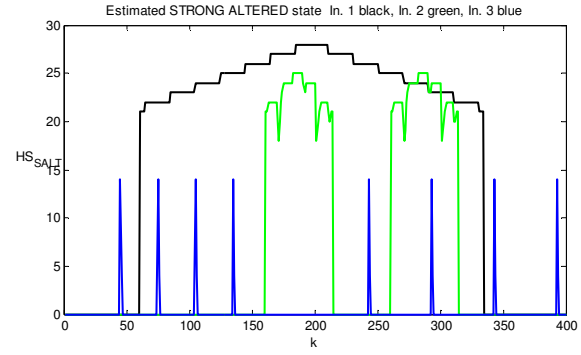

Figure 2b. Estimated scores corresponding to Strong Altered states

Fig. 3 represents the case when observable values exhibit fast oscillations with small amplitude around the expected boundaries of the Normal Range (in black) and around the expected mean (in green) which can be regarded as a form of parameter instability, maybe induced by some mistuning or malfunction in a sensor unit. It is important to note that this kind of events are estimated mainly as weak alterations (compare Fig 3a with Fig. 3b where blue is absent and green is reduced). Effectively very "slow-rate" spikes (with repetition rate longer than 50 observation steps) can indicate mostly a problem related to sensor units operations rather than some external attack engaged by an intruder.

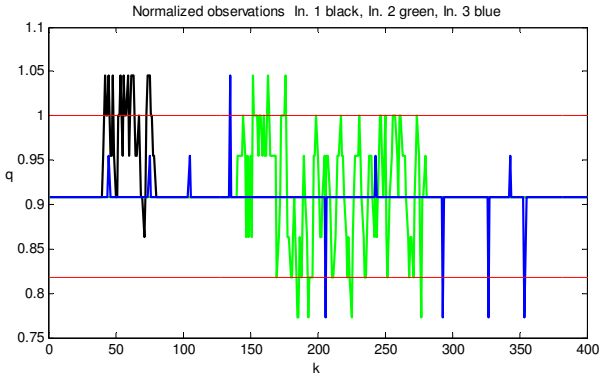

Figure 3. Normalized observation sequence generated by "fast oscillating" observables

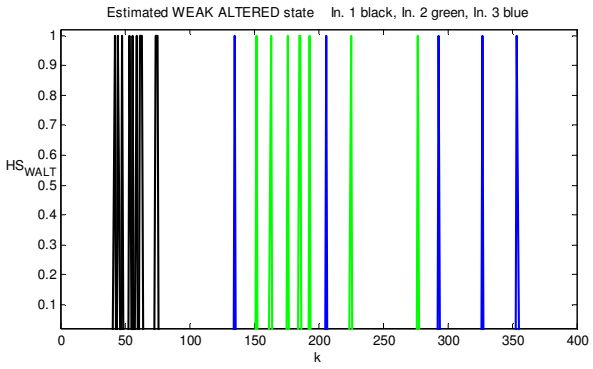

Figure 3a. Estimated scores corresponding to Weak Altered states

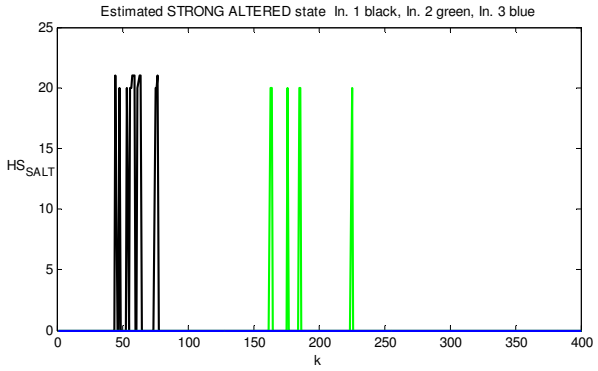

Figure 3b. Estimated scores corresponding to Strong Altered states

X. COMPUTATION AND MEMORY COSTS

We first show that computational and memory costs of MVET estimators (both infinite and finite memory) do not scale with input size: expressions (1) and (6) can be formally written as $\begin{cases}\bar{\mu}_k = f(\bar{\mu}_{k-1}, \bar{q}_k) \\ \bar{\sigma}^2_k = g(\bar{\sigma}^2_{k-1}, \bar{q}_k)\end{cases}$ and $\begin{cases}\bar{\mu}_{Tk} = f_T(\bar{\mu}_{Tk-1}, \bar{q}_k) \\ \bar{\sigma}^2_{Tk} = g_T(\bar{\sigma}^2_{Tk-1}, \bar{q}_k)\end{cases}$ respectively: both expressions at k-th observation step depend on the (k-1)-th estimation but the former depends only on the current observation sample, the latter only on the past $T \le T_{max}$ observation samples. In this case the dominant term is T_{max} and, therefore, computational and memory costs can be managed by tuning $T \le T_{max}$. Computation costs can be quantitatively measured in terms of computational time to perform arithmetic/logic operations. As it can checked from (1) and (6), the computational complexity of MVET estimators is reported in Table I.

TABLE I. MVET-ESTIMATORS COMPUTATION COST

MV-estimator	Sums	Products
$\mu_k(1)$	2	2
$\sigma^2_k(1)$	5	6
$\mu_{Tk}(6)$	T	1
$\sigma^2_{Tk}(6)$	T+1	T+2

It can be show that if MicaZ [7] motes are employed (8-bit processor ATMega128L @ 7.4 MHz), and assuming 20 clock cycles per arithmetic/logic operation, the average computation time per 32-bit operation is ~3 μs. Therefore the average computation time per observation step for the infinite-memory *MV-estimators* (1) $\mu_{\tilde{q}}$ and $\sigma^2_{\tilde{q}}$ are ~10 μs and ~40 μs, and for the *finite-memory MVET estimators* (6) $\bar{\mu}_{T_k}$ and $\overline{\sigma^2_{T_k}}$ are ~ 3T μs and ~ 25T μs respectively. If IMOTE [19] motes are employed (32-bit processor PXA271Xscale@{312, 416} MHz), and assuming 5 clock cycles per arithmetic / logic operation, the average computation time per 32-bit operation is ~0.03 μs (assuming a conservative ~300 MHz clock). Therefore the average computation time per observation step for the *infinite-memory MVET estimators* (1) $\mu_{\tilde{q}}$ and $\sigma^2_{\tilde{q}}$ are ~0.1 μs and ~0.4 μs, and for the *finite-memory MVET estimators* (6) $\bar{\mu}_{T_k}$ and $\overline{\sigma^2_{T_k}}$ are ~0.03T μs and ~ 0.25T μs respectively. Memory costs can be quantitatively measured in terms of memory usage to store variables and data. As it can checked from (1) and (6), at k-th observation step, some values for past MVET-estimators have to be stored as reported in Table II. The amount of bytes to put in memory depends on the format size for each variable and the measurement resolutions.

TABLE II. MVET-ESTIMATORS MEMORY COST

MV-estimator	Stores data
μ_k (1)	1
σ^2_k (1)	2
μ_{T_k} (6)	T
$\sigma^2_{T_k}$ (6)	T

XI. COMPARISATION WITH OTHER TECHNIQUES

The aim of this analysis is to show that the proposed MVET can have interesting performance in terms of FNR and FPR values with independency on data processing, or at least comparably but at a (much) lower computational and memory costs. Following the analysis results in [4], Table III schematically compares these performance indicators between the MVET and other widely employed detection techniques.

TABLE III. DETECTION TECHNIQUES COMPARED

Technique \ Performance	Resource Consumption	Reliability
Classification based	generally MEDIUM	Depends on label assignment policy
Clustering based	can be HIGH	Depends on the resolution of the clustering algorithm
Parametric Statistical Modeling	LOW for uni-variate data HIGH for multi-variate data	Depends on the mathematical relationship among data attributes
Information theoretic	always HIGH	HIGH in case of large number of anomalies
Nearest Neighbor-based Techniques	can be HIGH	Depends on distance measure
Spectral Techniques	always HIGH	generally HIGH
MV-estimator based	always LOW	generally HIGH: Tuneable FPR and FNR

Classification-based techniques [10] [25] are used to learn a model (classifier) from a set of labeled data instances (*training*) and then, classify a test instance into one of the classes using the learnt model (*testing*). Classification based anomaly detection techniques operate in a similar two-phase fashion. The training phase learns a classifier using the available labelled training data. The testing phase classifies a test instance as normal or anomalous using the classifier. Computational complexity depends on the classification algorithm being used: generally, training decision trees tends to be faster while techniques that involve quadratic optimization are more expensive. The testing phase of classification techniques is usually very fast since the testing phase uses a learnt model for classification. From the reliability point of view, the main disadvantage of these detection techniques is that multi-class classification relies on the availability of accurate labels for various normal classes, often not possible, which implies that both FNR and FPR strongly will depend on data processing.

Nearest Neighbor-based techniques are based on the key assumption that normal data instances occur in dense neighborhoods, while anomalies occur far from their closest neighbors: for continuous attributes, Euclidean distance is a popular choice but other measures can be used [25]; for categorical attributes, simple matching coefficient is often used but more complex distance measures can be used [3] [5]; for multivariate data instances, distance or similarity is usually computed for each attribute and then combined [25]. However main cons are that, for unsupervised techniques, if the data has normal instances that do not have enough close neighbors or if the data has anomalies that have enough close neighbors, the technique fails to label them correctly, resulting in missed anomalies which lead to high FNR; even for semi-supervised techniques, if the normal instances in test data do not have enough similar normal instances in the training data, FPR can be high; moreover the computational complexity of the testing phase is also a significant challenge since it involves computing the distance of each test instance with all instances belonging to either the test data itself, or to the training data, to compute the nearest neighbors.

Statistical anomaly detection techniques are based on the key assumption: that normal data instances occur in high probability regions of a stochastic model, while anomalies occur in the low probability regions of the stochastic model. In Sec. II has been reported that our MVET can be classified as a non-parametric statistical method which assumes the model structure not defined a priori, but determined from given data. Indeed, the key disadvantage of statistical techniques is that they rely on the assumption that the data is generated from a particular distribution. This assumption often does not hold true, especially for high dimensional real data and, moreover, even when the statistical assumption can be reasonably justified, there are several hypothesis test statistics that can be applied to detect anomalies; choosing the best statistic is often not a straightforward task. In particular, constructing hypothesis tests for complex distributions that are required to fit high dimensional data sets is nontrivial. Computational complexity depends on the nature of statistical model that is required to be fitted on the data. Fitting complex distributions (such as mixture models, Hidden Markov Models, etc.) using iterative estimation techniques such as Expectation Maximization (EM) [2], are also typically linear per iteration, though they might be slow in converging depending on the

problem and / or convergence criterion. As already considered in Sec. II, the proposed MV-Technique can deal with these issues exploiting a light and reliable detection algorithm.

Spectral techniques try to find an approximation of the data using a combination of attributes that capture the bulk of variability in the data [6]. Such techniques are based on the key assumption that data can be embedded into a lower dimensional subspace in which normal instances and anomalies appear significantly different. Several techniques use the Principal Component Analysis (PCA) [15] for projecting data into a lower dimensional space. Standard PCA based techniques are typically linear in data size but often quadratic in the number of dimensions. However spectral techniques are useful and FNR and FPR performance are good only if the normal and anomalous instances are separable in the lower dimensional embedding of the data. This occurrence can be very expensive in terms of memory and computational costs.

XII. MVET BASED FUNCTIONAL ARCHITECTURE FOR ANOMALY DETECTION OVER WSN

A prototype implementation of MVET technique over a test WSN is on-going in our labs. Fig. 4 depicts our reference functional architecture for the anomaly detection service and some different blocks are shown: the EMM module, which implements the expected stochastic indicators, the expected normal ranges for observables and the execution of the MVET Applicability Test (see Sec. V); the MVET block implements the MVET logic such as the computation of the sample indicators (6) at the k-th observation step and the generation of the Hazardousness Scores (4) (5) (eventually in an aggregated form) according to BCT. The Data Aggregation block (DAGG) is a special function that aggregate data (e.g. hazardousness scores) from different sensors that are monitoring the same set of observables from the same system.

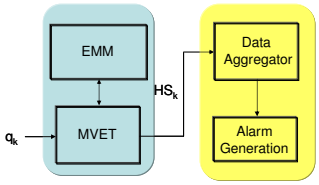

Figure 4. MVET-based AD logic over a generic sensor node

In a clustered WSN [1] [8], data flows are hierarchically structured: data patterns start from sensor members to the sensor head (or cluster head) which can be a member of another cluster with another cluster head and so on; in this way data patterns result to be converge-cast, i.e. from network boundaries to the base station. However all data from all sensors cannot be transferred to the base station as too much energy should be necessary for transmissions and too much memory should be available for their storage: therefore some aggregation function must be located at the cluster head. The Alarm Generation block (ALGEN) is another function up to a cluster head: any cluster head is in charge to issue alarms backwards to the base station according to received inputs from DAGG. In Fig. 4 functions related to member sensor nodes in a cluster are coloured in blue, while functions in charge of a cluster head are coloured in yellow (indeed a cluster head can be member node in another cluster and therefore it is also in charge of the functions in blue). It is important to note that functions in yellow do not need to be implemented anywhere in the WSN but dynamically over certain nodes.

XIII. MVET OPERATIONS OVER WSN

This is a *work in progress* issue too as we are implementing MVET in our lab. The adopted architectural design [20] is cross-layer [17] and platform-based [24]. Cross-layer (CL) results in the interplay between network layer (topology management and routing protocol) and presentation layer (mobile agent based execution environment for distributed monitoring applications): when applied to security, an important benefit of CL mechanism is the exploitation of the interplay between different security mechanisms in different layers to provide an enhanced security service to applications. Platform-based design (PBD) results in the availability of a software platform where the internal structure is composed by *interconnected* SW components, which represent abstractions of the wired hardware components [14]. Achievements of research goals are sought by taking care of the following major topics: selection of the right layers in the architectural design (a middleware layer is an essential component), application of the platform-oriented concepts for service mappings between layers, enhancement of the middleware layer with security services offered by lower layers entities and, on top, the creation of a flexible application environment by means of agents.

A key characteristic of mobile agent-based middleware is that any host in the network is allowed a high degree of flexibility to possess any mixture of code, resources, and processors. Its processing capabilities can be combined with local resources. Code (in the form of mobile agents) is not tied to a single host, but it is rather available throughout the network. Moreover, the mobile agent paradigm supports data-centric applications because the implementation code can migrate towards data, no matter about node addressing. Therefore, in a mobile-agent application execution environment (MAEE), each agent implements a sub-set of application components which can be proactively aggregated through agent mobility (code mobility across the network). Among the agent-based middleware solutions available from literature, we will refer to AGILLA [13]. According to block decomposition shown in Fig. 4, the mapping between MVET-based functional architecture and SW / agent components is as follows: EMM and MVET blocks are mapped into SW components, while DAGG and ALGEN blocks are mapped into a mobile agent. This design enables optimal allocation and code distribution for those functions, such as DAGG and ALGEN, that do not need to be implemented anywhere but just on cluster heads: in a clustered WSN, DAGG/ALGEN mobile agents are hosted only on cluster heads and move and clone across the network toward a cluster head as soon as elected.

A. *Data Aggregation*

Data aggregation function can be performed as follows: at k-th observation step, hazardousness scores coming from member nodes are compared and correlated (e.g. through a voting mechanism) in order to issue a reliable alarm and to

isolate conflicting scores within a cluster. See Fig. 5: suppose a clustered WSN with a DAGG/ALGEN mobile agent initially hosted only on node #1.

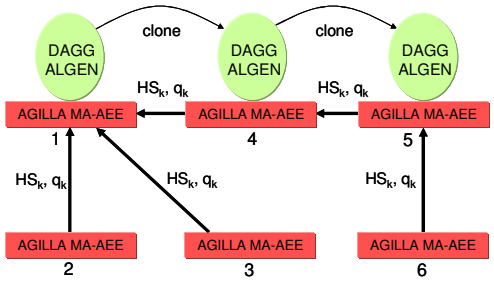

Figure 5. Mobile agent diffusion mechanism towards cluster heads

Next nodes #2, #3 and #4 attach themselves to the first node, the cluster grows and node #1 remains the cluster head. As soon as sensor node #5 joins the network at node #4, this node becomes cluster head of a new cluster formed by nodes #4 and #5 and the mobile agent DAGG/ALGEN clones itself and a copy moves to node #4 as new cluster head. In turn, the same occurs to node #5 when node #6 joins the network at node #5.

XIV. Conclusions and Future Work

In this paper we have presented our contribution to the anomaly detection problem over resource limited platforms as the wireless sensor networks. Our Mean-Variance estimation technique for the uni-variate case is being implemented in the WINSOME project at DEWS labs (WIreless sensor Network-based Secure system fOr structural integrity Monitoring and AlErting) targeted at the development of a cross-layer secure framework for monitoring and alerting applications over a MicaZ wireless sensor network. Major efforts in current activities are focused on completing this prototype implementation including the multi-variate extension. Several developments are also planned for the near future. Another important issue is to consider monitoring as a component in a control process where correlated actuations on the environment can be performed. This vision implies the integration of Hybrid System Control [9] items into the service platform.

Acknowledgment

We would like to thank Dr. Annarita Giani (formerly at UC Berkeley), co-author in some papers on the topic of intrusion detection.

References

[1] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., "A Survey on Sensor Networks," IEEE Communications Magazine, no. 8, 2002

[2] Bilmes, J. A., "A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," Technical Report TR-97-021. University of California at Berkeley, 1998

[3] Boriah, S., Chandola, V., and Kumar, V., "Similarity Measures for Categorical Data: A Comparative Evaluation," in Proceedings of the 8th SIAM International Conference on Data Mining, 2008

[4] Chandola, V., Banerjee, A., and Kumar, V. 2009 Anomaly Detection – A Survey," ACM Computing Surveys, vol. 41, no. 3, 2009

[5] Chandola, V., Boriah, S., and Kumar, V., "Understanding Categorical Similarity Measures for Outlier Detection," Technical Report 08-008. University of Minnesota, 2008

[6] Chatzigiannakis, V., Papavassiliou, S., Grammatikou, M., and Maglaris, B., "Hierarchical Anomaly Detection in Distributed Large-scale Sensor Networks," in Proceedings of the 11th IEEE Symposium on Computers and Communications, 2006

[7] CROSSBOW Technology Inc., http://www.xbow.com

[8] Culler, D., Estrin, D., and Srivastava, M., "Overview of sensor networks," IEEE Computer Magazine, no. 8, 2004

[9] Di Benedetto, M. D., Di Gennaro, S., and D'Innocenzo, A., "Discrete State Observability of Hybrid Systems," International Journal of Robust and Nonlinear Control, vol. 19, n. 14, 2009

[10] Duda, R. O., Hart, P. E., and Stork, D. G., Pattern Classification, J. Wiley & Sons, 2nd Edition, 2001

[11] Dunn, M., and Wigert, I., "An Inventory and Analysis of Protection Policies in Fourteen Countries," International CIIP (Critical Information Infrastructure Protection), Handbook 2004, edited by A. Wenger and J. Metzger, ETH Swiss Federal Institute of Technology Zurich, 2004

[12] Flammini, F., Gaglione, A., Mazzocca, N., Moscato, V., and Pragliola, C., "Wireless Sensor Data Fusion for Critical Infrastructure Security," in Proceedings of International Workshop on Computational Intelligence in Security for Information Systems, CISIS'08, 2008

[13] Fok, C. -L., Roman, G. C., and Lu, C., "Agilla: A Mobile Agent Middleware for Sensor Networks," Technical Report, Washington University in St. Louis, WUCSE-2006-16, 2006

[14] Gay, D., Levis, P., Von Behren, R., Welsh, M., Brewer, E., and Culler, D., "The nesC Language: A Holistic Approach to Networked Embedded Systems," in Proceedings of ACM SIGPLAN, 2003

[15] Jolliffe, I. T., Principal Component Analysis. Springer, 2nd Edition, 2002

[16] Kay, S. M., Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, 1993.

[17] Kliazovich, D., Devetsikiotis, M., and Granelli, F., "Formal Methods in Cross Layer Modeling and Optimization of Wireless Networks," Handbook of Research on Heterogeneous Next Generation Networking, 2009

[18] Lazarevic, A., Ozgur, A., Ertoz, L., Srivastava, J., and Kumar, V., "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection," in Proceedings of 3rd SIAM International Conference, 2003

[19] MEMSIC Corporation, http://www.memsic.com

[20] Pugliese, M., Pomante, L., and Santucci, F., "Agent-based Scalable Design of a Cross-Layer Security Framework for Wireless Sensor Networks Monitoring Applications," in Proceedings of the International Workshop on Scalable Ad Hoc and Sensor Networks, 2009

[21] Pomante L., Pugliese M., Marchesani S., and Santucci F., "WINSOME: a Middleware Platform for the Provision of Secure Monitoring Services over Wireless Sensor Networks," in Proceedings of the 9th International Wireless Communications and Mobile Computing Conference (IEEE IWCMC2013), Cagliari, 2013

[22] Pugliese, M., Pomante, L., and Santucci, F., "Secure Platform over Wireless Sensor Networks," chapter in Applied Cryptography and Network Security, ISBN 978-953-51-0218-2, INTECH Publishers, 2012

[23] Rajasegarar, S., Leckie, C., Palaniswami, M., and Bezdek, J. C., "Distributed Anomaly Detection in Wireless Sensor Networks," in Proceedings of Communication Systems, 2006

[24] Sangiovanni-Vincentelli, A., and Martin, G., "Platform-based Design and Software Design Methodology for Embedded Systems," in Proceedings of IEEE Computer Design & Test, vol. 18, n. 6, 2001

[25] Tan, P.-N., Steinbach, M., and Kumar, V., Introduction to Data Mining. Addison-Wesley, 2005