Palomares Carrascosa, I., Kalutarage, H., Huang, Y., McCausland, R., Miller, P., & McWilliams, G. (2017). A fuzzy multicriteria aggregation method for data analytics: application to insider threat monitoring. In *2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS 2017): Proceedings of a meeting held 27-30 June 2017, Otsu, Japan* (pp. 329-334). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/IFSA-SCIS.2017.8023360

Peer reviewed version

Link to published version (if available):
10.1109/IFSA-SCIS.2017.8023360

A Fuzzy Multicriteria Aggregation method for Data Analytics: application to Insider Threat Monitoring

Iván Palomares
Department of Computer Science
University of Bristol. United Kingdom
i.palomares@bristol.ac.uk

Harsha Kalutarage, Yan Huang, Paul Miller
Robert McCausland, Gavin McWilliams
Centre for Secure Information Technologies
Queen's University Belfast. United Kingdom
{h.kalutarage,y.huang,p.miller,r.mccausland,g.mcwilliams}@qub.ac.uk

***Abstract*—With the increase in volume, heterogeneity and uncertainty in data, conventional analytics approaches for monitoring users behavior in organisations are no longer sufficient for the effective and reliable detection of malicious activities. This motivates the need for introducing additional analysis techniques. This paper introduces an intelligent fusion method based on fuzzy aggregation functions typically utilized in multi-criteria decision making. The proposed method, which can be integrated with analytics systems, undertakes temporal and multi-criteria fusion processes on pre-analyzed data, to enhance effective monitoring and decision-making. An application to a prominent area of research in the cyber-security domain, the insider threat problem, is shown to validate the usefulness of our method.**

I. Introduction

The insider threat problem is one of the most daunting challenges to handle in computer security, indeed in all aspects of real-world security [1]. The term *insider* refers to "a current or former employee, contractor or business partner who has authorised access to an organisation's network, system or data, and intentionally exceeds or misuses that access in an manner that adversely affects the confidentiality, integrity or availability of the organisation's information system" [2]. Recent industry surveys and academic literature revealed unequivocal evidence supporting the significance and prevalence of this threat [3]. Notable infamous cases such as Robert Hanssen, Bradley Manning, and Edward Snowden illustrate the scale of damage an insider can pose to an organisation [4].

Security monitoring tools designed to tackle external attacks, e.g. Intrusion Detection Systems (IDS) and firewalls, are not sufficient to monitor internally initiated attacks. Many of these tools are ineffective against privileged insiders who already have been granted legitimate access to sensitive data and systems, thus adding a significant degree of uncertainty to the monitoring process. Any tool that seeks modeling and reasoning about insiders behavior must accept this ground truth and deal with incompleteness (compensate for lack of knowledge), inconsistencies (resolve ambiguities and contradictions) and change (update the knowledge base over time). With the increasing volume, variety, velocity and value of Big Data [5], standalone classical approaches such as rule-based techniques are no longer sufficient to handle these complexities and overcome the difficulties in making the correct decisions.

Various machine learning and analytics techniques (e.g. Support Vector Machines (SVM) [6], Hidden Markov Models (HMM) [7] and agent-based [8]) have been tested against the insider problem with limited success, as the behavioral signatures of insider activities are often reported as normal activity by classical solutions. We argue that combining traditional user behavior analytics tools with intelligent fusion approaches based on fuzzy Multi-Criteria Decision Making (MCDM) [9], [10], may help accurately detecting malicious behaviors, analyzing data under multiple perspectives (criteria) separately, and fusing it into meaningful conclusions.

This contribution presents an intelligent fusion approach for analyzing multiple perspectives of users behavior data across the time, and inferring suspicious behavior potentially driven by malicious insider activities. Our approach enhances user monitoring systems based on data analytics to provide reliable information for decision support. The proposed solution is characterized by using popular fuzzy aggregation techniques in MCDM with the purpose of: (i) intelligently combining temporal information about the user behavior (e.g. predicated on streaming data), in terms of multiple behavioral aspects (criteria), and (ii) providing more comprehensive information about the user behavior in terms of the different activities undertaken by her/him, as means for more informed decision support. The main contributions of our work are threefold:

1) We present a novel intelligent fusion framework based on fuzzy MCDM aggregation methods, that can be integrated with a data analytics-driven monitoring system. In particular, we apply our framework to monitor insider threats predicated on streaming data describing users behavior.
2) We adopt the aggregation methods from [11] and the dynamic MCDM principles from [12] to define a temporal, multicriteria fusion scheme of user behavior information.
3) We define an improved dynamic re-weighting method within the fusion approach, making it more adaptive to the current information related to a specific user.

This paper is set out as follows. Related works and preliminaries are outlined in Section II. Section III presents the intelligent multi-criteria fusion approach, and Section IV integrates it with an insider threat monitoring system, to demonstrate its validity. Finally, concluding remarks are drawn in Section V.

978-1-5090-4917-2/17$31.00

II. Background

This section provides an overview of related works on aggregation approaches for MCDM [9]. Their underlying aggregation functions are also formally introduced.

Albusac et al. presented in [11] an expert surveillance system that dynamically weighs and combines information (normality scores) categorised under multiple criteria (surveillance aspects such as trajectory and speed of tracked objects). The aggregated score is subsequently utilised to support decision-making when abnormal situations are encountered in the monitored environment. The goal in their work is twofold: (i) to reduce the occurrence of false negatives, i.e. abnormal situations that are classified as normal and thus ignored by the system, and (ii) to reduce the triggering of false positives, i.e. unnecessary alarms. The authors describe the use of several weighted aggregation functions to combine normality scores, as well as a dynamic re-weighting scheme that adapts the weights to the values of the input normality scores. Among these functions, the Sugeno Integral [13] provided particularly meaningful results. This function has the ability to reflect the interaction between criteria in a MCDM framework, by using a fuzzy measure that assigns weights not only to individual criteria, but also to combinations of them.

Definition 1. *[13] Let $A = \{a_1, a_2 \ldots, a_z\} \in [0,1]^z$ be a set of z values in the unit interval, associated to a set $C = \{c_1, \ldots, c_z\}$ of criteria. The Sugeno integral $S_\mu : [0,1]^z \rightarrow [0,1]$, associated to a fuzzy measure μ, is defined as follows:*

$$a = S_\mu(a_1, \ldots, a_z) = \bigvee_{j=1}^{z} a_{(j)} \wedge \mu\left(\{c_{(j)}, \ldots, c_{(z)}\}\right) \quad (1)$$

where $c_{(j)}$ is the criterion to which the i-th lowest value in A corresponds. Thus, the Sugeno integral arranges inputs in increasing order before fusing them. $\vee$, $\wedge$ are a t-conorm and t-norm function, respectively.

Example 1. *The Sugeno Integral typically utilizes the maximum and minimum operators, in which case Eq. (1) becomes:*

$$a = \max_j \{\min\left(a_{(j)}, \mu\left(\{c_{(j)}, \ldots, c_{(z)}\}\right)\right)\} \quad (2)$$

Classical weighted aggregation functions, such as the weighted average, weigh criteria individually. By contrast, fuzzy measures are defined on subsets of criteria, modeling the interactions between them and capturing the importance of combined criteria. A fuzzy measure $\mu(C') \in [0,1]$ satisfies:

1) $\mu(\emptyset) = 0$.
2) $\mu(C') = 1$, with C the set of criteria and $C' \subseteq C$.
3) If $C' \subset C'' \subseteq C$, then $\mu(C') \leq \mu(C'')$.
4) The *additive* property, $\mu(C' \cup C'') = \mu(C') + \mu(C'')$, which only holds for so-called additive fuzzy measures [14].

In the work by Albusac et al., the temporal dimension is also considered by determining the trapezoidal area under aggregated scores, across consecutive time instants. We instead consider the use of associative, full-reinforcing aggregation functions to analyze the temporal evolution of users' behavior. The idea is inspired by the Dynamic Multicriteria Decision-Making (DMCDM) framework presented by Campanella and Ribeiro in [12]. In their work, multiple scores are fused across the time into a single dynamic score, which reflects how scores on a specific decision element (e.g. an alternative or criterion) evolve across the time. Scores are aggregated across $k \geq 1$ multiple time periods, $T-k+1,\ T-k+2, \ldots,\ T$ to obtain the *dynamic* score at time T. In order to avoid storing *all* the past decision information from time $T-k+1$ onwards, Campanella et al. [12] proposed using an associative, full-reinforcing aggregation function $\phi : [0,1]^k \rightarrow [0,1]$ to compute a dynamic rating at T, $\tilde{s}^T$, as $\tilde{s}^T = \phi(s^{T-k+1}, \ldots, s^T)$, accomplishing:

- *Associativity*: ϕ only requires fusing the original score at T with its associated dynamic rating at $T-1$ to output a reliable indicator of the rating evolution across the time:

$$\tilde{s}^T = \phi(\tilde{s}^{T-1}, s^T) \quad (3)$$

- *Full reinforcement*: It prioritises the increasing (resp. decreasing) temporal evolution of ratings, applying upward (resp. downward) reinforcement on the aggregated rating.

Uninorms are a popular family of aggregation functions fulfilling these two properties [15]. They were introduced as a generalization of t-norms and t-conorms, with a neutral element $e \in [0,1]$ lying anywhere in the unit interval.

Definition 2. *[15] A uninorm is a mapping, $\mathcal{U} : [0,1]^2 \rightarrow [0,1]$, having the following properties for all $a, b, c, d \in [0,1]$:*

i) Commutativity*: $\mathcal{U}(a,b) = \mathcal{U}(b,a)$.*
ii) Monotonicity*: $\mathcal{U}(a,b) \geq \mathcal{U}(c,d)$ if $a \geq c$ and $b \geq d$.*
iii) Associativity*: $\mathcal{U}(a, \mathcal{U}(b,c)) = \mathcal{U}(\mathcal{U}(a,b), c)$.*
iv) Neutral element*: $\exists e \in [0,1] : \mathcal{U}(a,e) = a$.*

Example 2. *The following is an example of uninorm function with neutral element e, based on the general family of uninorms introduced by Fodor et al. in [16]:*

$$\mathcal{U}(a,b) = \begin{cases} \dfrac{ab}{e} & \text{if } 0 \leq a, b \leq e, \\ \dfrac{a+b+ab-e}{1-e} & \text{if } e \leq a, b \leq 1, \\ M_{\mathcal{U}}(a,b) & \text{otherwise.} \end{cases} \quad (4)$$

with $M_{\mathcal{U}}(a,b)$ an averaging function.

Uninorms present a pessimistic, t-norm behavior in $[0,e]^2$ (downward reinforcement) when the two aggregation inputs are *low*. Conversely, they present an optimistic, t-conorm behavior in $[e,1]^2$ (upward reinforcement) when such inputs are *high*. Finally, when one of the inputs is above e and the other is below e, uninorms exhibit an averaging behavior.

III. Intelligent multicriteria Aggregation Method

This section presents an intelligent data fusion approach that analyzes and combines multiple aspects of users' behavior, based on fuzzy aggregation functions for MCDM. Our approach is conceived for its use alongside data analytics tools,

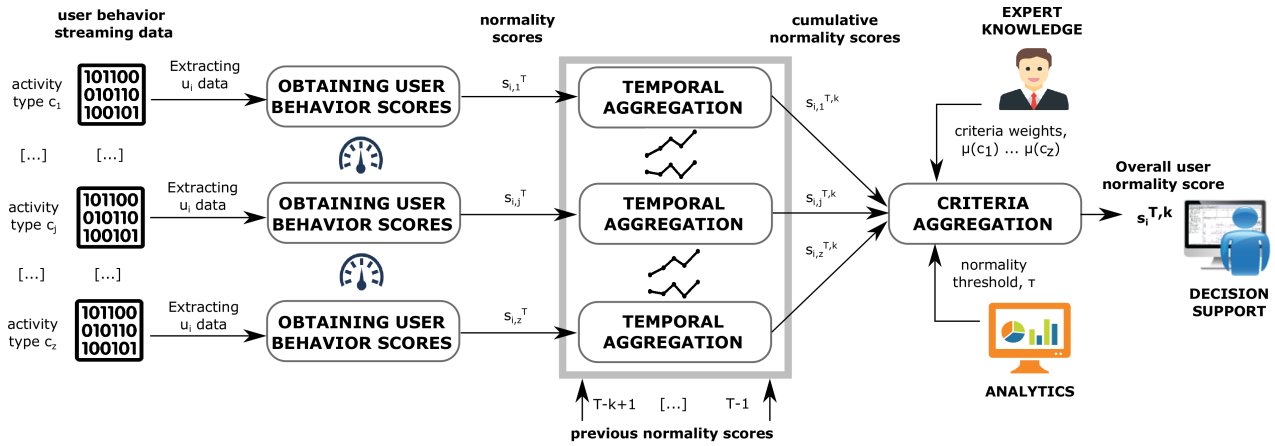

Fig. 1. Scheme of the intelligent multicriteria aggregation framework

to enhance situation-aware decision support. We extend the intelligent fusion method from [11], by: (i) introducing an improved and more adaptive weighting scheme based on each user's current data, and (ii) incorporating associative, full-reinforcement functions in the temporal aggregation process, based on [12]. In Section IV we further integrate this method with a data analytics framework for insider threat monitoring. Figure 1 shows an overview of the fusion methodology, consisting of three phases: obtaining user behavior scores, temporal aggregation, and criteria aggregation.

Remark 1. *We use the terms "activity type" and "event type" (such as e-mail usage, HTTP activity, device authentication, etc.) hereinafter to refer to the same concept indistinctly, i.e. a criterion $c_j \in C$ under a MCDM perspective.*

A. Obtaining user behavior scores

We firstly introduce a method to calculate, for each existing user u_i, a number z of scores that describe her/his behavior within a time window $T = [t^-, t^+]$, according to $z \geq 2$ criteria $C = \{c_1, \dots, c_z\}$. Let us assume z data sources, containing streaming information about different types of events (criteria) separately, e.g. different types of user activity registered in a monitored environment. Such data are partitioned across consecutive, disjoint time windows $\{\dots, T-2, T-1, T, \}$ (for instance, on an hourly or daily basis). Suppose each data instance or event in the j-th source, $j \in \{1, \dots, z\}$ is a triplet $< u_{i,j}, t, B_{i,j}^t >$ with u_i a monitored user, $t \in T$ a time stamp, and $B_{i,j}^t \in \{1\!:\!normal, 0\!:\!abnormal\}$ a (binary) classification of an event triggered by u_i at t. Importantly, we assume without loss of generality that the value of $B_{i,j}^t$ stems e.g. from applying prior analytics (e.g. a SVM-classifier, as will be illustrated later on in experiments) on raw streaming sensor data of users activity. Let $E_{i,j}^T$ be the number of events of type j caused by u_i within the time window T:

$$E_{i,j}^T = \#B_{i,j}^t \ : \ t \in T \tag{5}$$

Likewise, let $N_{i,j}^T$ be the number of events of type j originated by u_i and classified as *normal* during T:

$$N_{i,j}^T = \#B_{i,j}^t \ : \ B_{i,j}^t = 1 \wedge t \in T \tag{6}$$

A *normality score* $s_{i,j}^T \in [0,1]$ associated to u_i and time interval T is calculated, for each event type or criterion c_j, as follows:

$$s_{i,j}^T = \frac{N_{i,j}^T}{E_{i,j}^T} \tag{7}$$

If $E_{i,j}^T = 0$, i.e. the user has no reported activity under c_j during T, we adopt the convention $s_{i,j}^T = s_{i,j}^{T-1}$. The normality score is an indicator of u_i's behavior during T regarding c_j: the larger its value, the more "normal" (less malicious) her/his activity is deemed. Figure 2 illustrates the process of calculating normality scores.

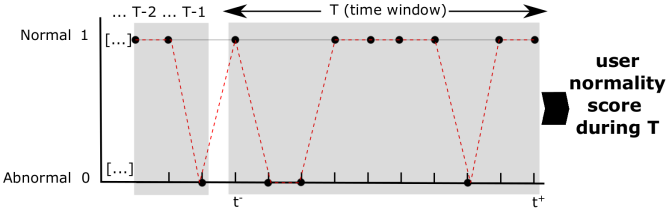

Fig. 2. Obtaining user behavior score at time window T

B. Temporal aggregation

Instead of relying solely on users behavior at the latest time window T, we are also interested in their evolution across the time. Therefore, in this phase the user normality scores for the k latest time windows are combined to obtain a *cumulative normality score* $s_{i,j}^{T,k} \in [0,1]$ spanning the last k time windows [1] ending in T. This cumulative score reflects both the current and recent trend of u_i's behavior under c_j. To do this, and inspired by the DMCDM framework in [12], we propose using a uninorm aggregation function $\mathcal{U}$ to iteratively combine the normality scores of k consecutive time windows, as follows:

$$s_{i,j}^{T,k} = \mathcal{U}(s_{i,j}^{T-1,k-1}, s_{i,j}^T) \tag{8}$$

The cumulative normality score $s_{i,j}^{T,k}$ incorporates user behavior information from k previous time windows back to T–k+1. Intuitively, $s_{i,j}^{T,1} = s_{i,j}^T$.

[1] In this work, both the window size T and the temporal aggregation size k are deemed domain specific paremeters whose values are fixed by an expert.

Remark 2. *As a consequence of the full reinforcement property of uninorms, if u_i's behavior under criterion c_j is reported as normal both lately and currently, then the evidence of normal behavior becomes stronger (a higher value is assigned to $s_{i,j}^{T,k}$). Conversely, if both recent and current behavior are rather abnormal, then the resulting $s_{i,j}^{T,k}$ is lower, thus reinforcing the hypothesis of a possibly abnormal behavior.*

Example 3. *Consider a user whose normality scores for his activity related to $c_j : e-mail$, in the last five time windows are $s_{i,j}^{T-4} = 0.4$, $s_{i,j}^{T-3} = 0.36$, $s_{i,j}^{T-2} = 0.9$, $s_{i,j}^{T-1} = 0.8$, and $s_{i,j}^{T} = 0.9$. Assume a uninorm function with neutral element $e = 0.5$ is used. The five scores can be combined to obtain $s_{i,j}^{T,5}$, as follows:*

$$s_{i,j}^{T-3,2} = \mathcal{U}(s_{i,j}^{T-4}, s_{i,j}^{T-3}) = \mathcal{U}(0.4, 0.36) = 0.3$$
$$s_{i,j}^{T-2,3} = \mathcal{U}(s_{i,j}^{T-3,2}, s_{i,j}^{T-2}) = \mathcal{U}(0.3, 0.9) = 0.67$$
$$s_{i,j}^{T-1,4} = \mathcal{U}(s_{i,j}^{T-2,3}, s_{i,j}^{T-1}) = \mathcal{U}(0.67, 0.8) = 0.86$$
$$s_{i,j}^{T,5} = \mathcal{U}(s_{i,j}^{T-1,4}, s_{i,j}^{T}) = \mathcal{U}(0.86, 0.9) = 0.93$$

with $\mathcal{U}$ a uninorm function. Initially, the user follows an abnormal behavior trend (low scores), thus downward reinforcement occurs in the first aggregation step. However, his behavior drastically changes from $T-2$ onwards, therefore the cumulative scores are eventually reinforced upwards. Finally, when a low score is combined with a high one, the uninorm aggregation function shows an averaging behavior, as occurs with the computation of $s_{i,j}^{T-2,3}$. Figure 3 illustrates this effect.

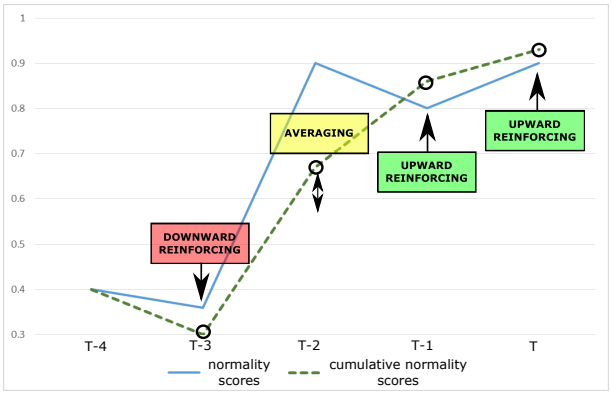

Fig. 3. Temporal aggregation example based on uninorm functions

C. Criteria aggregation

The purpose of the criteria aggregation phase is to combine cumulative normality scores over z criteria into one by using a weighted aggregation function ϕ_μ, as $s_i^{T,k} = \phi_\mu(s_{i,1}^{T,k}, \ldots, s_{i,z}^{T,k})$. This aggregated score is ultimately compared against a normality threshold $\tau \in [0, 1]$ to support decision-making (τ can be e.g. periodically updated based on the analytics model used alongside our method, see Figure 1).

This phase extends the dynamic re-weighting aggregation method from [11], which exhibits a number of desirable properties in monitoring and surveillance scenarios, including: the *adaptability* of importance weights to the current score values, the *reliability* in reducing the number of false alarms, and the *effectiveness* in detecting as many real abnormal behaviors as possible. Importantly, here we define a novel approach to dynamically adjust weights of criteria, based not only on the normality scores being aggregated, but also on the current normality threshold.

Without loss of generality, the Sugeno integral in Eq. (2) is considered hereinafter, and weights of individual criteria are gathered a priori, e.g. by a domain expert. The following two aspects are described in further detail in this phase:

i Dynamically adjustment of weights based on current aggregation inputs.

ii Defining the fuzzy measure μ, in other words, obtaining the weights of combined criteria in C.

Dynamic Weight Adjustment. The importance weight $\mu(c_z)$ assigned to an activity type c_z determines how important such activity becomes in the presence of abnormal (low) score values, in order to conclude whether the user exhibits a malicious behavior or not. The Sugeno Integral assumes such weights (resulting from the definition of a fuzzy measure) are static, i.e. once calculated/assigned they are not affected by the actual scores to be aggregated. However, in a security-oriented monitoring context, it is more sensible that the weights of criteria describing normal behavior at present *decrease* with respect to those weights of criteria on which a possibly abnormal behavior is occurring. It is therefore necessary to "adapt" the initial (single) criteria weights $\mu(c_j)$ according to each user behavior, in a manner that allows for enhancing reliable and effective aggregation results. In [11], an *abnormality score* is calculated for each normality score, as $a_{i,j}^{T,k} = 1 - s_{i,j}^{T,k}$. This implies that the greater the normality score, the larger its difference with the associated abnormality score. The method in [11] assumes that the score domain [0,1] is evenly distributed around a central and static normality threshold, $\tau = 0.5$. However, when the value of this threshold is not static, but instead it is periodically updated based on existing data and prior analytics processes, it may occur that $\tau \neq 0.5$. Assume for instance that $\tau = 0.6$ and $s_{i,j}^{T,k} = 0.8$. This normality score falls midway between τ and a "total normality" value of 1. Intuitively, the value of $a_{i,j}^{T,k}$ should in this case lie midway between 0 and τ. However, we instead would have $a_{i,j}^{T,k} = 1 - 0.8 = 0.2$, which is significantly closer zero to than to τ. We propose making the computation of the abnormality score *adaptive* to any normality threshold $\tau \in]0, 1[$, with the following formula:

$$a_{i,j}^{T,k} = \begin{cases} s_{i,j}^{T,k} & \text{if } s_{i,j}^{T,k} = \tau, \\ 1 - \dfrac{s_{i,j}^{T,k}(1-\tau)}{\tau} & \text{if } s_{i,j}^{T,k} < \tau, \\ \dfrac{\tau(1 - s_{i,j}^{T,k})}{1-\tau} & \text{if } s_{i,j}^{T,k} > \tau. \end{cases} \quad (9)$$

This formula maps $s_{i,j}^{T,k}$ from $[0, \tau]$ to $[\tau, 1]$ (or vice versa)

and then obtains $a_{i,j}^{T,k}$ as its symmetrical value with respect to the middle point of the mapped interval. Eq. (9) ensures that an accurate abnormality score is calculated, regardless of τ. The difference between abnormality and normality, $\Delta_{i,j}^{T,k} = a_{i,j}^{T,k} - s_{i,j}^{T,k}$, is then calculated. Clearly, $\Delta_{i,j}^{T,k} > 0$ when the the normality score is below τ, whereas $\Delta_{i,j}^{T,k} < 0$ when the score lies above τ. Based on $\Delta_{i,j}^{T,k}$, the weight $\mu(c_j)$ initially assigned to criterion c_j is updated to adapt it to the user u_i currently being analyzed, as shown below:

$$\mu_i^T(c_j) = \begin{cases} \mu(c_j) & \text{if } \Delta_{i,j}^{T,k} \leq 0 \\ \mu(c_j) + \Delta_{i,j}^{T,k} & \text{otherwise;} \end{cases} \quad (10)$$

Consequently, $\mu(c_j)$ is adjusted (increases) for user u_i at time T if, given c_j, the associated scores reflect a rather abnormal behavior, and it remains unchanged otherwise. Weights are finally re-normalized to ensure $\sum_j \mu_i^T(c_j) = 1$.

Obtaining combined criteria weights. Ideally, domain expert knowledge would be used to determine all importance weights for (subsets of) criteria. However, this is often not feasible in realistic scenarios, particularly when the number of criteria is large (notice that z criteria require defining a fuzzy measure consisting of 2^z weights, one for each $C' \subseteq C, C' \in 2^C$). To alleviate this problem, Sugeno proposed in [13] a formula that, based on weights of single criteria, allows to automatically determine the weights of combinations of them:

$$\mu_i^T(C' \cup C'') = \mu_i^T(C') + \mu_i^T(C'') + \lambda \mu_i^T(C') \cdot \mu_i^T(C'') \quad (11)$$

with $C', C'' \subseteq C$. The parameter λ is computed upon individual criteria weights $\mu_i^T(c_j)$, by solving the equation[2]:

$$1 + \lambda = \prod_{i=1}^{z} (1 + \lambda \mu_i^T(c_j)) \quad (12)$$

The method above is convenient when there is no available expert information on the effects of combined criteria, hence only weights of individual criteria, $\mu_i^T(c_1) \ldots \mu_i^T(c_z)$, are provided. Eqs. (12) and (11) are subsequently applied to determine λ and the combined criteria weights, respectively. There exist other methods in the literature to construct a fuzzy measure, most of which are based on optimization problems.

IV. Integration with an Insider Threat Monitoring System

In this section we integrate the fusion approach with a classification-based insider threat monitoring system, to demonstrate its validity in practice. For the purpose of validation, we use the insider threat test dataset provided by [17].

A. Dataset description

The dataset consists of both synthetic background data and data from synthetic malicious actors, describing a collection of logs from distributed host-based sensor networks within a large organisation. The underlined attack scenario is outlined as follows: a user who did not previously use removable drives or work outside working hours, begins logging in after working hours, using a removable drive, and uploading data to wikileaks.org. After this, she leaves the organization shortly. In order to produce the dataset, entries containing streaming user behavior data have been recorded over the time, for four different activity types: c_1: logging details, c_2: device usage details, c_3: email activities, and c_4: web surfing details alongside Lightweight Directory Access Protocol (LDAP). For this proof of concept demonstration, we randomly select a small sample from the above dataset which includes data for only one insider and four innocent users over 25 days.

B. Experimental settings & results

In user behavior analytics applications, the validation based on real traces of user activity with ground truth on attack activity is often difficult, due to security, privacy and other legislation issues. A commonly adopted solution to this problem consists in employing unary classification algorithms. In practice, it is relatively easier to find "clean" (normal) data than malicious (abnormal) data, particularly for the insider problem. Hence we employ one class support vector machine (OCSVM) as the primary classifier (baseline method), and integrate our proposed fusion method on top of it for comparisons. For experiments, each day is split into two segments, with each 12-hour segment a time window T. Within T, users are profiled using raw streaming data of their activity.

In order to compare our approach (OCSVM+MCDM) against the standalone OCSVM baseline, we calculate the normality score for a given user in two different ways. Firstly, we apply the OCSVM classifier itself under each criterion c_j separately and calculate a normality score per user, as the percentage of normal events produced by that user within T. Secondly, output values from the OCSVM classifier are used as inputs to the proposed fusion method, whose aggregated normality score $s_i^{T,k}$ is deemed as the overall normality score of the user. Figure 4 presents and describes the results gathered across 43 time windows, with a temporal aggregation size $k = 7$ and the following criteria weights (before dynamic re-weighting): $w_1 = 0.3$, $w_2 = 0.1$, $w_3 = 0.3$, $w_4 = 0.3$. Dark red lines show the scores from the insider user based on the proposed method (OCSVM+MCMD). Since our method combines all criteria, these scores have been plotted in all four sub graphs to facilitate comparisons. Orange lines show insider scores applying OCSVM separately on each criterion. Green lines show normal user scores using the proposed method, and blue lines show normal user scores using OCSVM. Our proposed approach outperforms the baseline approach: OCSVM can not detect the insider by looking at http, email and device user activity separately. It might be argued that event data from different sensors should be combined first into a single source and then apply OCSVM to improve detection results. Although seemingly feasible, this implies a substantially higher computational cost in practice, hence it is not feasible to merge together raw data from different activity types without losing some discriminating features. Therefore,

[2]E.g. via *R* mathematical software suite: https://www.r-project.org/

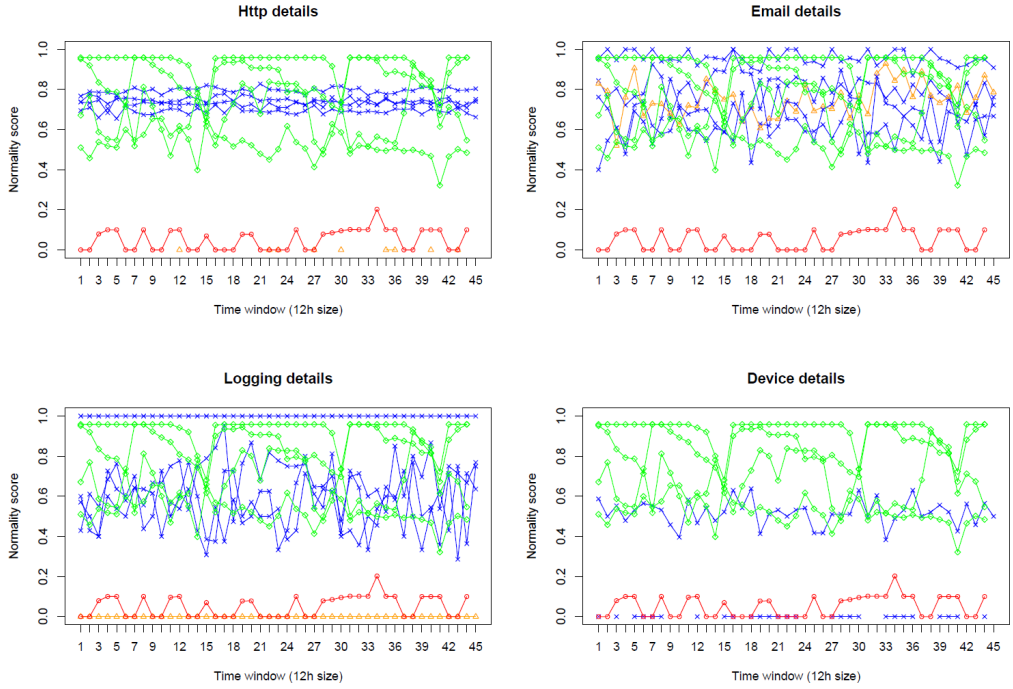

Fig. 4. Comparison: OCSVM vs OCSVM+MCMD. Dark red (circle) line shows the insider scores over the time using OCSVM+MCMD, orange (triangle) line shows insider scores using OCSVM, green (diamond) lines show innocent user scores using OCSVM+MCMD and blue (cross) lines show innocent user scores using OCSVM. Device details show only two users for OCSVM, as only two users reported device activity during the time period considered.

the proposed approach can help overcoming this cost limitation and effectively handling the uncertainty in data sources.

V. Concluding Remarks

This contribution has presented an intelligent multi-criteria fusion approach for monitoring user behavior data. Our approach performs both temporal and multi-criteria aggregation processes predicated on fuzzy aggregation methods, to fuse different types of users activity across the time, providing meaningful insights of their behavior to enhance monitoring and decision support. By integrating our methodology with an unary classification-based framework for insider threat monitoring, we have demonstrated its validity to effectively and efficiently detect malicious behaviors. Future directions of work aim at introducing additional dimensions of knowledge into the proposed fusion framework, such as psychometric data, and incorporating evidential reasoning techniques for its application to intrusion detection systems under uncertainty.

Acknowledgment

This work has been funded by EPSRC CSIT 2 project (Ref: EP/N508664/1).

References

[1] M. Bishop, H. M. Conboy, H. Phan, B. I. Simidchieva, G. S. Avrunin, L. A. Clarke, L. J. Osterweil, and S. Peisert, "Insider threat identification by process analysis," in *Security and Privacy Workshops (SPW)*, 2014, pp. 251–264.

[2] C. Dawn, A. Moore, R. Trzeciak, and T. Shimeall, "Common sense guide to prevention and detection of insider threats," *CERT*, 2009.

[3] J. R. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. Wright, and M. Whitty, "Understanding insider threat: A framework for characterising attacks," in *Security and Privacy Workshops (SPW) 2014*, 2014, pp. 214–228.

[4] P. A. Legg, N. Moffat, J. R. Nurse, J. Happa, I. Agrafiotis, M. Goldsmith, and S. Creese, "Towards a conceptual model and reasoning structure for insider threat detection." *JoWUA*, vol. 4, no. 4, pp. 20–37, 2013.

[5] B. Furht and F. Villanustre, *Big Data Technologies and Applications*. Springer, 2016.

[6] M. B. Salem and S. J. Stolfo, "Masquerade attack detection using a search-behavior modeling approach," *Columbia University, Computer Science Department, Technical Report CUCS-027-09*, 2009.

[7] P. Bradford and N. Hu, "A layered approach to insider threat detection and proactive forensics," in *Proceedings of the Twenty-First Annual Computer Security Applications Conference (Technology Blitz)*, 2005.

[8] J. A. Sokolowski, C. M. Banks, and T. J. Dover, "An agent-based approach to modeling insider threat," *Computational and Mathematical Organization Theory*, vol. 22, no. 3, pp. 273–287, 2016.

[9] M. Aggarwal, "Discriminative aggregation operators for multicriteria decision making," *Applied Soft Computing*, vol. 52, pp. 1058–1069, 2017.

[10] C. Kahraman (Ed.), *Fuzzy Multi-Criteria Decision Making: Theory and Applications with Recent Developments*. Springer, 2008.

[11] J. Albusac, D. Vallejo, J. Castro-Schez, C. Glez-Morcillo, and L. Jimnez, "Dynamic weighted aggregation for normality analysis in intelligent surveillance systems," *Expert Systems with Applications*, vol. 41, no. 4, Part 2, pp. 2008 – 2022, 2014.

[12] G. Campanella and R. A. Ribeiro, "A framework for dynamic multiple-criteria decision making," *Decision Support Systems*, vol. 52, no. 1, pp. 52 – 60, 2011.

[13] M. Sugeno, "Theory of fuzzy integrals and its applications (doctoral thesis, tokyo institute of technology)," 1974.

[14] M. Grabisch, *Fuzzy Measures and Integrals: Recent Developments*. Springer, 2015, pp. 125–151.

[15] R. Yager and A. Rybalov, "Uninorm aggregation operators," *Fuzzy Sets and Systems*, vol. 80, pp. 111–120, 1996.

[16] J. Fodor, R. Yager, and A. Rybalov, "Structure of uninorms," *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 5, pp. 411–427, 1997.

[17] J. Glasser and B. Lindauer, "Bridging the gap: A pragmatic approach to generating insider threat data," in *Security and Privacy Workshops (SPW)*, 2013, pp. 98–104.