Enterprise Security with Adaptive Ensemble Learning on Cooperation and Interaction Patterns

Kyle Quintal[1], Burak Kantarci[1], Melike Erol-Kantarci[1], Andrew Malton[2] and Andrew Walenstein[2]
[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada
E-mails: {kquin039, burak.kantarci, melike.erolkantarci}@uottawa.ca
[2]BlackBerry Limited, Canada
E-mails: {amalton, awalenstein}@blackberry.com

***Abstract*—Social networking research has primarily focused on public social networking services and applications, while rich social interactions in an enterprise setting and their related context has received less attention. In this paper, we focus on using the enterprise social context to augment traditional authentication tools. This is motivated by the emergence of smart mobile devices which introduce ease of remote access to work from almost anywhere and anytime, adding spatio-temporal dimension to the social context. However, it remains a challenge to efficiently manage access-controlled events by using different contextual properties. This paper analyzes specific actions under specific access-control rules to extract context-aware machine learning predictions. Such analysis includes the introduction of three contextual metrics: document shareability, valuation, and user cooperation. Furthermore, these socially-dependent metrics are combined with our Smart Enterprise Access Control (SEAC) technique to achieve authenticity precision of 99% while improving the corresponding efficiency trade-off associated with high and strict security.**

I. Introduction

With the widespread use of smart devices, it is increasingly effortless to interact with co-workers in almost any context. Most large scale enterprises manage system interactions using some derivation of a Role-Based access control (RBAC) system [1], which limits controls based on permissions and roles given to particular individuals. However, these systems may not consider the context in which access is requested, which also includes (but not limited to) the shared interactions of users on particular digital assets in differing locations [2].

In simplistic RBAC systems, users with correct access rights would obtain access to their corresponding allowed assets. However, most systems do not consider the context in which these access rights are requested, which then differs the access rights themselves (working remotely for instance). Perhaps an additional layer of security should be required before granting access, or potentially access should always be prohibited in specific contexts. This complex challenge then requires dynamically adapting access rights in different contexts, all the while considering important metrics such as employee efficiency and security vulnerabilities.

Many hidden patterns exist within a large enterprise settings [3]. With large amounts of daily user interactions, patterns emerge when looking at when actions are performed, on which asset(s), in collaboration with how many other users, and the amount of performed changes within a specific period.

Such detailed and dynamic workflows require adaptive solutions [4] in need of many insights regarding user context. One of these insights is the formulation of user behaviors or behaviometrics. These social interactions can be used for continuous authentication purposes [5] and even combined with biometrics [6] to ensure proper user authenticity is obtained before approved requested access.

When considering enterprise security, a common solution is to enforce a Virtual Private Network (VPN) connection to allow or restrict certain assets to be accessed. Despite having its pros and cons [7], VPN's have limitations on how efficiently an employee can do their work since priority is always given to security over efficiency. This trade-off between user work efficiency and user access security is of particular interest, as highly robust security measures in RBAC systems are desirable, but often at the cost of blocking certain work-related tasks. We address this challenging trade-off by using a variety of machine learning techniques.

This work contains two contributions: first, to introduce three new context-based metrics on interactions within RBAC systems. These new metrics (Shareability, Valuation, and Co-operation) are then used in our second contribution, a new Smart Enterprise Access Control (SEAC) technique, i.e. an ensemble supervised learning technique. The SEAC technique employs different machine learning techniques to obtain high-security ratings of 99.60% and 97.50% while maintaining reasonable efficiency rates of 96.8% and 74.01% on two distinct datasets.

The subsequent sections are organized as follows: Section II discusses important contributions within the access-control domain; Section III introduces newly introduced contextual features and SEAC technique; Section IV discusses performance evaluation of machine learning results; Section V provides insights on the results, important challenges and additional considerations.

II. Related Work

Intelligent inferences are possible when analyzing user behavior; the degree of trust [8] assigned to users based on their social interactions is a meaningful representation. Knowing the degree at which users are being social [5] represents well user behavior and can be used in difficult authentication decisions. These specific metrics of sociability [9] show great promise in

978-1-7281-3893-0/20/$31.00 ©2020 IEEE

continuously authenticating users, especially when combined with biometrics [6]. On top of everything, improving security by increasing authenticity can be further extended by looking at the context of our daily interactions [10].

Traditional cybersecurity methods can be augmented by allowing the decision making focus to shift towards user-centric context. As presented in [11], using machine learning to aide an employee, and not the system, in making optimal decisions is an area under investigation. More specifically, when monitoring information in event management systems, machine learning algorithms can be used to assist analysts with correctly flagging malicious and suspicious events. In [11], authors show that using machine learning have reduced missed flags in their study.

Machine learning also shows great promise in enterprise security, from improving DNS vulnerabilities by Big Data using semi-supervised learning [12] to personal assistants aiding employees in identifying cybersecurity vulnerabilities [13] [11]. Several studies have found deep learning promising in enterprise cybersecurity [14], where the authors demonstrate positive performance on algorithms focused on malware, intrusion and spam detection. Further work focuses on intrusion detection [15] by using machine learning on open source frameworks. These, among many, are new learning methods surfacing in industry security. This specific study [16] shows new derivations of Random Forests assist analysts in detecting threats within large networks with up to 99.8% detection rate.

Insider threats [17] are another reality in large industries. Such threats have to define properties that can be clustered into suspicious behavior categories, for predictive and in-depth features vulnerability analysis. Not only can machine learning be used in this regard, but also in identifying the original owner of a particular document in a large scale system as in [18]. Knowing an asset's author can be fruitful in address security concerns within an organization, especially when wrongful labeling is common. Finally, having access to all this data brings up concerns about privacy [19]. Emphasizing privacy concerns and considering access control decisions in parallel is an attractive goal in providing a more transparent decision-making system. In this paper, different than prior work, we focus on the social interactions of employees in an enterprise setting. We define metrics to quantify context-related features, then use machine learning algorithms to identify these contexts.

III. Context-based Access Control

Integration of contextual proprieties within access control decisions is both highly challenging and difficult in scope. We begin by describing the standard system model, which contains users performing actions on documents with restrictive permissions. Failure to comply with the given rules results in action rejections. The interest of this study is to adopt such rules by integrating the context of the performed action in the decision process. For instance, members of a top-security meeting may not all have access to a particular asset under discussion, but their mere presence that meeting should automatically grant them access, even if only temporarily. As such, including spatial-temporal interactions of individuals, alongside their contextual properties and additional extracted features (defined in the subsequent section) in an intelligent decision process is the focus of this work. This is accomplished using machine learning techniques to analyze which contexts cause certain actions to either be accepted or rejected. When a rejection is predicted, the user is prompted with an authenticate method, as a means to reduce rejections caused by insufficient authenticity. With an appropriate level of authenticity, the receiving RBAC system may then take more meaningful decisions, having both context-aware features a user's authenticity as parameters. Our dataset was generated based on behavioral data collected in an authentic enterprise setting in a probabilistic manner. The context of this data includes seven features, two which are optional. These are described in TABLE I, and the above decision-making process is illustrated in Fig. 1.

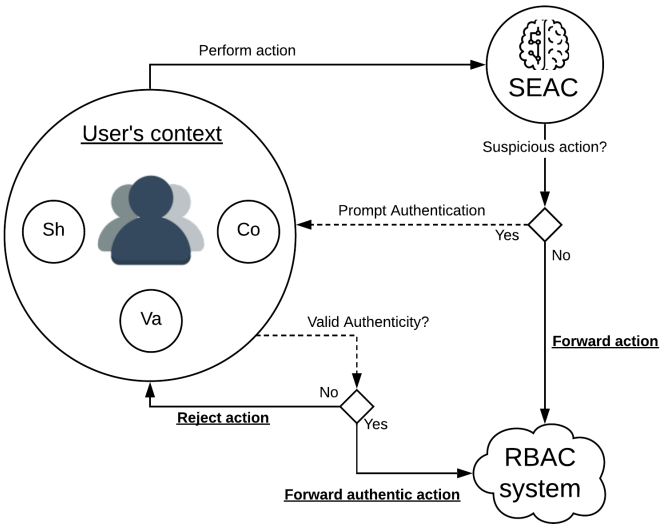

Fig. 1. SEAC technique inserted into a typical system

After investigating the contributions to the principal data components (Principal Component Analysis), it was discovered the most contributing features were the *loc* and *time* of the actions. However, the *user* who performed the action and the *action* occurred did not contribute as anticipated.

A. Contextual Analysis

In this section, we present formulations that extract three new factors defining user interactions within an enterprise setting. Precisely, the amount of value a particular document (**Va**), access frequency and how much a document is shared (**Sh**), and how often users cooperate (**Co**) are the extracted factors of context. These factors (Va, Sh, and Co) provide additional details on the given context of users and documents. All the presented factors are considered over timeframes, represented as Δ_k. In each subsequent timeframe, all three factors are recalculated by giving exponentially less weight (α) to previous values, as seen in Eqs. 4,9 and 14. Table 2 defines

TABLE I Feature definitions

Features	Description
loc	Where the action was taken
user	Who performed the action
time	When the action was performed
action	What was performed
document	The affected asset
op1	Optional parameter 1
op2	Optional parameter 2
Va	Document Valuation
Sh	Document Shareability
Co	User Cooperation

all variables definitions used in the formulation of these three factors.

TABLE II Environment Definitions

Variable	Description
D	set of Documents
i, j	instances $\in$ D
U	set of Users
u, v	instance $\in$ U
τ_k	*k*th instance of timeframe τ
Δ_k	length of each timeframe τ_k
α	weight given to past calculations
A:{R,W,X,S,C}	set of Actions: Read, Write, Delete(**X**), Set and Clear permission
$\gamma_{u,i}$	Action matrix of *u* on *i*
$\delta_{u,i}$	Interaction weight between *u* and *i*
$\omega_{u,i}$	Valued action weight of *u* performing *i*
$\Omega_{u,i}$	Average valued action on *i*
$\lambda_{u,i}$	Highest social action by *u* on *i*
Λ_u	Action matrix of *u*
$\mu_{u,v}$	Weighted cooperation of *u* with *v*
$Sh_i(\tau_k)$	Shareability of document within τ_k
$Va_i(\tau_k)$	Valuation of document within τ_k
$Co_u(\tau_k)$	Cooperation of user within τ_k

We first define $\gamma_{u,i}$ which is the matrix containing all of the actions performed by a specific user u on a document i. As seen in Eq. 1, we divide by the timeframe, or Δ_K, to represent actions performed by users (u) on documents (i).

$$\gamma_{u,i}(\Delta_k) = [\Sigma W, \Sigma S, \Sigma C, \Sigma X, \Sigma R] \tag{1}$$

1) Document Shareability: The first extracted social and contextual factor is how much a document is shared between users, referred to as document *Shareability*, or $Sh_i(\tau_k)$. The initial step is defining the weight of specific interactions - ($\delta_{u,i}(\Delta_k)$) between users and documents, as seen in Eq. 2. A document is considered to be shared when it is either written upon or read, where writing is considered twice the "sharing" weight as reading.

$$\delta_{u,i}(\Delta_k) = \begin{cases} 2, & \gamma_{u,i}(\Delta_k)|W| > 0 \\ 1, & \gamma_{u,i}(\Delta_k)|R| > 0 \\ 0, & else \end{cases} \tag{2}$$

The second step is the normalization of $\delta_{u,i}(\Delta_k)$. This results in instant Shareability(Sh_i), which quantifies how much a document was shared by each user compared to others, again only within this timeframe (Eq. 3).

$$Sh_i(\Delta_k) = \frac{\sum_{u=0}^{U} \delta_{u,i}(\Delta_k)}{\sum_{j=0}^{D} \sum_{u=0}^{U} \delta_{u,j}(\Delta_k)} \tag{3}$$

The final step is the inclusion of remembrance. Specifically, the previous Shareability values (τ_{k-1}) are used to calculate the current value - **Shareability factor(**Sh_i**)**, where α_1 defines the importance given to past Sh_i values, as seen here:

$$Sh_i(\tau_k) = \sum_{t=1}^{k} \alpha_1^t (1 - \alpha_1) Sh_i(\Delta_t) \tag{4}$$

2) Document Valuation: The second extracted factor is the value of a document itself. Each document is deemed to change in value as the number of edits, reads and permission increase or decrease in time. This four-part process begins by defining a partial function, giving different weights to different actions based on how much "value" they provide:

$$\omega_{u,i}(\Delta_k) = \begin{cases} 4, & \gamma_{u,i}(\Delta_k)|W| > 0 \\ -1, & \gamma_{u,i}(\Delta_k)|R| > 0 \\ 1, & \gamma_{u,i}(\Delta_k)|C| > 0 \\ 0, & else \end{cases}; \tag{5}$$

In the context of our study, writing (**W**) on a document is deemed four times as beneficial as clearing permission(**C**), and reading a document(**R**) without any other contributions within this timeframe degrades document valuation. Using the valued weights brought on by each user(u), the averaged valued weight - $\Omega_i(\Delta_k)$ - is defined on each document(i):

$$\Omega_i(\Delta_k) = \frac{\sum_{u=0}^{U} \omega_{u,i}(\Delta_k)}{|U'_u(\Delta_k)|} \tag{6}$$

Where $U'_u(\Delta_k)$ defines all users who have performed an action on i in this Δ_k:

$$U'_u(\Delta_k) : u \in U_u(\Delta_k) | \omega_{u,i} \neq 0 \tag{7}$$

Using $\Omega_i(\Delta_k)$, intervals are defined to represent importance levels. Intervals are needed since true quantification on a document's value requires more than simply action analysis. A total of five different levels were needed in our study. Here, the instant document value($Va_i(\Delta_k)$) is represented as:

$$Va_i(\Delta_k) = \begin{cases} 1, & \Omega i(\Delta_k) > 3 \\ 0.75, & \Omega i(\Delta_k) > 2 \\ 0.5, & \Omega i(\Delta_k) > 1 \\ 0.25, & \Omega i(\Delta_k) > 0 \\ 0, & else \end{cases}; \quad (8)$$

When a document is only read in sequencing timeframes, $Va_i(\Delta_k)$ would diverge to 0, as negative contributions would diminish its overall value over time (at a rate of α_2). In other words, once a document has been, say published, and read by everyone, it's value degrades to nothing in the system. This is represented with with Valuation factor($Va_i(\tau_k)$), seen here:

$$Va_i(\tau_k) = \sum_{t=1}^{k} \alpha_2^t (1-\alpha_2) Va_i(\Delta_t) \quad (9)$$

3) User Cooperation: Cooperation represents users interaction and is interpreted using actions performed on common documents by two different users. The starting point in the Cooperation factor is defining the highest collaborative action - $\lambda_{u,i}(\Delta_k)$ - similarly to our previous $\gamma_{u,i}$ definition. The difference is that $\lambda_{u,i}(\Delta_k)$ only contains a single value depicting the highest evaluated action on a specific document(i) by that user(u) (Eq. 10). x represents each index of the row or each possible action as defined in A:

$$\lambda_{u,i}(\Delta_k)[x] = \begin{cases} 1, & \gamma_{u,i}(\Delta_k)[x] > 0 \\ & \wedge (\exists! y | \gamma_{u,i}(\Delta_k)[y] > 0 \wedge (y < x)) \\ 0, & else \end{cases} \quad (10)$$

The priority given to actions is identical to their weights as seen in Eq. 12. Furthermore, all the highest social actions are combined in a matrix that represents all actions performed by each user (u) on each document (i), Eq. 11. A single row then represents the highest social action on a specific i by u, and columns represent the set D. An empty row then signifies no actions were performed on i by u within that Δ_k:

$$\Lambda_u(\Delta_k) = [\lambda_{u,i}(\Delta_k)], \forall i \in D \quad (11)$$

As cooperation requires two different users, the weighted cooperation - ($\mu_{u,v}$) - compares the $\Lambda_u(\Delta_k)$ of two users and assigns a cooperative weight, specific to this our study once again, seen here:

$$\mu_{u,v}(\Delta_k) = \begin{cases} 5, & \exists i(\Lambda_u(\Delta_k)[i, |W|] = 1 \wedge C1) \\ 4, & \exists i(\Lambda_u(\Delta_k)[i, |S|] = 1 \wedge C1) \\ 3, & \exists i(\Lambda_u(\Delta_k)[i, |C|] = 1 \wedge C1) \\ 2, & \exists i(\Lambda_u(\Delta_k)[i, |X|] = 1 \wedge C1) \\ 1, & \exists i(\Lambda_u(\Delta_k)[i, |R|] = 1 \wedge C1) \\ 0, & else \end{cases}; C2 \quad (12)$$

$C1 = \exists a | \Lambda_v(\Delta_k)[i, a] = 1, a \in A$
$C2 = \forall u, \forall v \in U, u \neq v$

It should be noted, $\mu_{u,v}$ may not be equal to $\mu_{v,u}$, as actions may not be symmetric between users. After each weighted cooperation is defined ($\mu_{u,v}(\Delta_k)$), each cooperation within the current timeframe is normalized:

$$Co_u(\Delta_k) = \frac{\sum_{v=0}^{U} \mu_{u,v}(\Delta_k)}{\sum_{u=0}^{U} \sum_{v \neq u}^{U} \mu_{u,v}(\Delta_k)} \quad (13)$$

As with other factors, it's previous values are considered using a time-degrading weight(α_3) to determine the current Cooperation factor($Co_u(\tau_k)$):

$$Co_u(\tau_k) = \sum_{t=1}^{k} \alpha_3^t (1-\alpha_2) Co_i(\Delta_t) \quad (14)$$

B. Smart Enterprise Access Control (SEAC)

Extracting contextual features from social interactions in an enterprise setting is the first step in providing more intelligent system decisions. In this work, besides extracting features, we introduce an access control scheme, namely Smart Enterprise Access Control (SEAC), that uses an ensemble learning approach. SEAC is based on previous work related to adaptive supervised learning [20], which uses an adaptive weighted-voting technique which attributes varying weights to classifiers depending on their True Positive Rates (TPR), or Security rates in our domain. Individual classifiers (called SEAC members) provide confidence scores instead of simple binary predictions as inputs to this technique. The combination of predictions and assigned Trust Scores (TS) aim to improve the trade-off found in standard machine learning models. This specific trade-off is between security rates (precision) and efficiency rates (recall), as explained further in section 4.

TPRs defines the weight given to SEAC members and is denoted as:

$$TPR_{i,t} = \frac{TP_{i,t}}{TP_{i,t} + FN_{i,t}} \quad (15)$$

Here, true positives (TP) are occurrences of suspicious actions being correctly predicted as rejected, thus receiving authentication. False negatives (FN) are occurrences of suspicious actions being incorrectly predicted as accepted, thus receiving no authentication. The attractive metric of focus for SEAC members is their TPRs, and this represents the security rating of each individual member. This TPR is what defines the weight given to each member, which represents that member's influence on the final SEAC prediction. These weights are assigned based on the highest TPR within each timeframe, as seen in Eq. 15. Then the Performance factor (Pf, Eq. 16) defines the performances of members over time, using α_4 as the threshold of change over time.

$$Pf_{i,t} = \alpha_4 TPR_{i,t-1} + (1 - \alpha_4) TPR_{i,t} \tag{16}$$

TABLE III Notation for SEAC ensemble model

Notation	Description
N	Set of i classifiers
T	Set of t time-frames
TP	True Positive
FN	False Negative
$\text{TPR}_{i,t}$	True Positive rate of i during t
α_4	Weight of previous TPR
$\text{Pf}_{i,t}$	Performance Factor of i during t
$\text{TS}_{i,t}$	Trust Score of i during t
$\text{P}_{Auth}(i)$	Predicted Rejection Confidences of i
$\text{P}_{Allow}(i)$	Predicted Acceptance Confidences of i

Next, Trust Score (TS) represents the trust given to each member, then relating to the weight given to their predictions. The TS is essentially normalized $Pf_{i,t}$ across all members within the ensemble, as defined by Eq. 17. In certain cases, the TF of underperforming members may converge to 0, if constantly outperformed by other members, which only improves results, as it represents members providing no valuable predictions.

$$TS_{i,t} = \frac{Pf_{i,t}}{\sum_{j=0}^{N} Pf_{j,t}} \tag{17}$$

Each trust score is combined with the members corresponding predictions. The cumulative weighted predictions of all members are defined with P_{Auth} and P_{Allow}. More specifically, they represent the two final predictions for the entire SEAC technique, where the prediction with the highest confidence is rendered as the final product, as seen by Eqs. 18-20.

$$P^{Auth}(t) = \sum_{i=0}^{N} (TS_{i,t} \cdot P_i^{Auth}), \quad \forall t \in T \tag{18}$$

$$P^{Allow}(t) = \sum_{i=0}^{N} (TS_{i,t} \cdot P_i^{Allow}), \quad \forall t \in T \tag{19}$$

$$Prediction(t) = \begin{cases} Authenticate, & P^{Auth}(t) > P^{Allow}(t) \\ Allow, & else \end{cases} \tag{20}$$

As the SEAC ensemble technique requires base classifiers at its core, a total of three classifiers were selected for comparison purposes. Random Forest (RF) represents logical classifiers, and already contains proven success in multiple areas of authentication [21] [22] by using the power of ensemble binary trees. Next is Naive Bayes (NB), a suitable candidate for statistical learning, as the most contributing features are almost independent of each other. There exist many geometric classifiers, but only a select few have proven successful in authentication studies, one being Support Vector Machines (SVM), as shown in [23], [9] and [10].

IV. Performance Evaluation

Two datasets, namely shareability-influenced dataset (SD) and valuation-influence dataset (VD), were generated from historical user behavior within an enterprise environment. Most performed actions are located within a large radius, or a single large cluster (e.g. head office) and the remaining actions are scattered into multiple smaller clusters (e.g. off-site workplaces). These clusters of actions are labelled with the resulting action decision (accepted or rejected), which is the subject of our classification predictions. It becomes evident that higher percentages of actions within the large cluster would be accepted compared to their smaller cluster counterparts, as authenticity would be less in question when actions are performed in well defined location (common workplace). As 70% of the data was used for training, the remaining portion was divided into 30 equally-sized timeframes, used by the SEAC technique for predictions.

First, security and efficiency trade-off is studied where the former denotes precision and the latter denotes recall. The reason why considered as an indicator of efficiency is to minimize the number of re-authentication attempts.

The three members of the SEAC technique are first compared using the original data, seen in the upper section of Table III-A. The predictions of each member (RF, NB, and SVM) are presented in Table IV, where the security rate of NB is seen as the highest. However, NB does also has the lowest efficiently rates. As for the other two members, both SVM and RF performed similarly, but SVM required additional training resources.

TABLE IV Optimal results with original features

	Security (%)		Efficiency (%)	
Models	SD	VD	SD	VD
RF	82.17	86.13	64.31	90.94
NB	98.35	95.70	47.57	58.7
SVM	81.56	85.80	63.44	92.52

Using the SEAC adaptive ensemble voting technique, we optimized security at a much more efficient trade-off compared to all other base models (Table V). This resolution was possible by NB and it's precise security awareness, which almost always had the highest TS during validation. SEAC and NB are also compared using the Area Under the ROC curve (AUC). This comparison (Fig.2) shows how security ratings are maintained successfully, while improving efficiency rates, resulting in a more improved trade-off.

In the rest of the results, we evaluate the impact of tunable parameters on the performance of SEAC. For comparisons purposes, we displayed their influence on the results, considering both security and efficiency rates (Fig.3, Fig.4, Fig.5).

TABLE V Performance considering *Sh*, *Va* and *Co*

	Security (%)		Efficiency (%)	
Models	SD	VD	SD	VD
RF	89.90	95.37	72.46	95.79
NB	99.93	97.50	64.74	90.13
SVM	71.6	94.68	71.33	94.79
SEAC	99.60	97.50	74.01	96.80

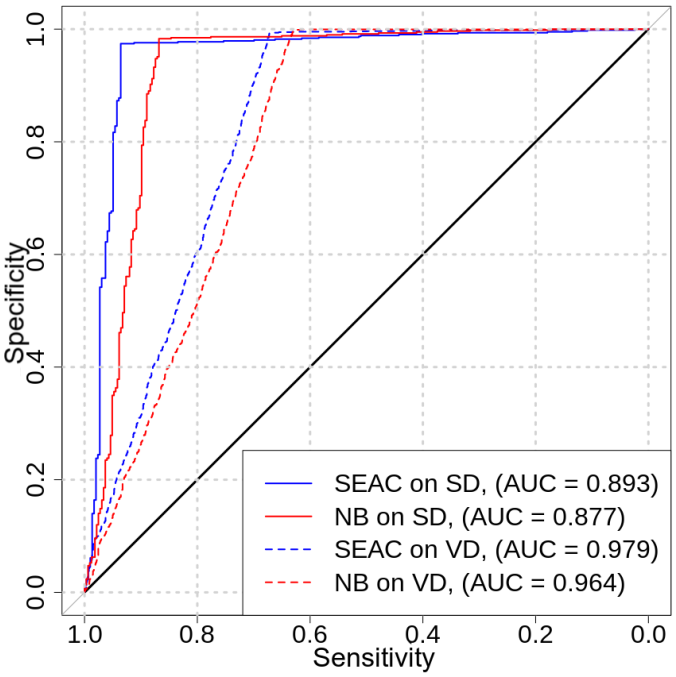

Fig. 2. Improved trade-off of SEAC

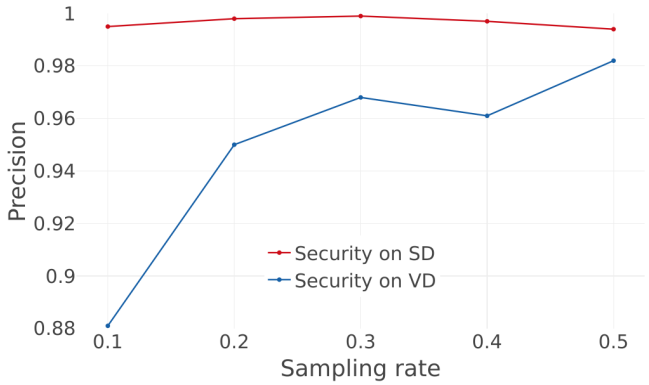

Fig. 3. SEAC undersampling rates

Sampling rates are used to balance the offset of class distribution. As seen in Fig.3, as the sampling rate increases the performance of SEAC improves. However, when sampling rates are over 0.4, there is too much loss of information, and results become unstable and misleading.

Regarding time degradation, only α_4 is presented in results

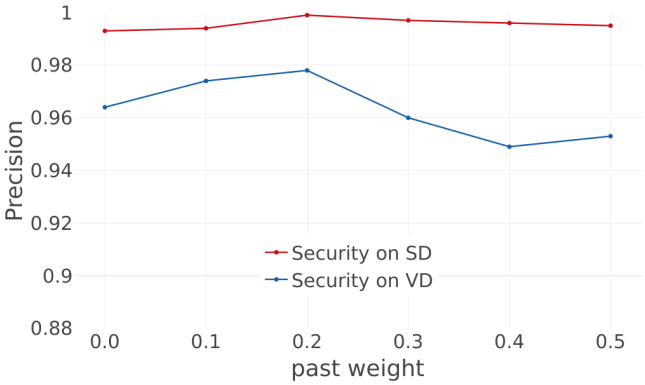

Fig. 4.

in Fig.4, although the concepts are similar for all time thresholds. The optimal value is observed as 0.2.

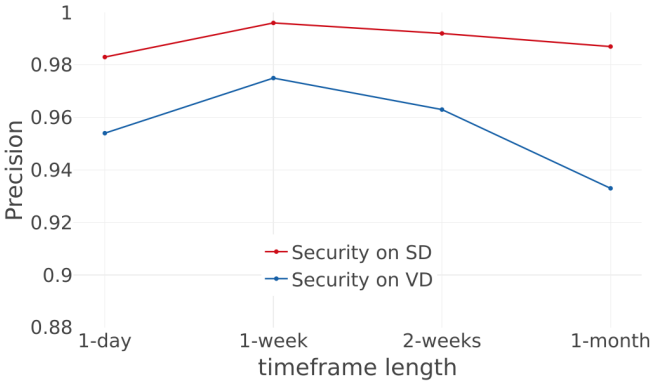

Fig. 5. Variations in SEAC timeframes

Finally, different timeframes were investigated in our results: a single day, single week, bi-weekly and monthly timeframes. Again, these time-related thresholds correlate with the defined weights that are given to degrading calculations $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, but still performed best with timeframes set at weekly intervals.

V. Conclusion and Future Work

In this work, specific contextual properties were extracted and combined with multiple machine learning classifiers in making authenticity decisions. Furthermore, an ensemble learning technique, SEAC, has been introduced to enhance overall trade-off often found in enterprise system security and user efficiency. Going over the results, each model has it's own advantages and disadvantages; RF is the most consistent of models, with the highest efficiency metric; NB provides the best security rate, but at the price of lower efficiency; SVM is also consistent, but overshadowed by RF and take an exceeding amount of training time on large datasets ($O(n^2)$). By utilizing the weighted voting technique between both RF and NB, the focus was directed on added security. However, a

shift in focus towards efficiency can be achieved by replacing eq. 1 by TNR and selecting efficiency-focused models for the composition of SEAC. Additional tuning should also be investigated (α) for optimizing how quickly weights are transferred between models, based on $\mathrm{Pf}_{i,t}$. As our data was generated based on the habitual patterns employees in a realistic scenario, these results demonstrate the feasibility of using context and authenticity for adaptive access control decisions.

As future work, datasets generated from different settings could further evaluate the method and contextual properties suggested here. This includes, but is not limited to, the evaluation of SEAC using larger and more heterogeneous data sets.Regarding the details of the ensemble classifier, there may be interest in further optimizing $\mathrm{TS}_{i,t}$. Improvements could include the considerations of different success metrics (True-Negative-Rate, Accuracy, F1 score) for more adaptive predictions. Furthermore, comparing SEAC with similar existing ensemble techniques which use dynamic weighting could provide additional insights on its performance in enterprise security.

ACKNOWLEDGEMENT

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) DISCOVERY, and ENGAGE programs

REFERENCES

[1] L. Zhou, V. Varadharajan, and M. Hitchens, "Achieving secure role-based access control on encrypted data in cloud storage," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 1947–1960, Dec 2013.

[2] B. Shafiq, J. B. D. Joshi, E. Bertino, and A. Ghafoor, "Secure interoperation in a multidomain environment employing rbac policies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1557–1577, Nov 2005.

[3] C. Feltus, E. Dubois, and M. Petit, "Alignment of remmo with rbac to manage access rights in the frame of enterprise architecture," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, May 2015, pp. 262–273.

[4] M. Leitner, S. Rinderle-Ma, and J. Mangler, "Aw-rbac: Access control in adaptive workflow systems," in *2011 Sixth International Conference on Availability, Reliability and Security*, Aug 2011, pp. 27–34.

[5] F. Anjomshoa, M. Catalfamo, D. Hecker, N. Helgeland, A. Rasch, B. Kantarci, M. Erol-Kantarci, and S. Schuckers, "Mobile behaviometric framework for sociability assessment and identification of smartphone users," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, June 2016, pp. 1084–1089.

[6] I. Deutschmann, P. Nordström, and L. Nilsson, "Continuous authentication using behavioral biometrics," *IT Professional*, vol. 15, no. 4, pp. 12–15, July 2013.

[7] O. Adeyinka, "Analysis of problems associated with ipsec vpn technology," in *2008 Canadian Conference on Electrical and Computer Engineering*, May 2008, pp. 001 903–001 908.

[8] B. Kantarci, K. G. Carr, and C. D. Pearsall, "SONATA: Social Network Assisted Trustworthiness Assurance in Smart City Crowdsensing," *International Journal of Distributed Systems and Technologies (IJDST)*, vol. 7, no. 1, pp. 59–78, 2016.

[9] F. Anjomshoa, M. Aloqaily, B. Kantarci, M. Erol-Kantarci, and S. Schuckers, "Social behaviometrics for personalized devices in the internet of things era," *IEEE Access*, vol. 5, pp. 12 199–12 213, 2017.

[10] L. Fridman, S. Weber, R. Greenstadt, and M. Kam, "Active authentication on mobile devices via stylometry, application usage, web browsing, and gps location," *IEEE Systems Journal*, vol. 11, no. 2, pp. 513–521, June 2017.

[11] C. Feng, S. Wu, and N. Liu, "A user-centric machine learning framework for cyber security operations center," in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, July 2017, pp. 173–175.

[12] L. Watkins, S. Beck, J. Zook, A. Buczak, J. Chavis, W. H. Robinson, J. A. Morales, and S. Mishra, "Using semi-supervised machine learning to address the big data problem in dns networks," in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan 2017, pp. 1–6.

[13] C. M. Sayan, "An intelligent security assistant for cyber security operations," in *2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, Sep. 2017, pp. 375–376.

[14] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," in *2018 10th International Conference on Cyber Conflict (CyCon)*, May 2018, pp. 371–390.

[15] L. Zomlot, S. Chandran, D. Caragea, and X. Ou, "Aiding intrusion analysis using machine learning," in *2013 12th International Conference on Machine Learning and Applications*, vol. 2, Dec 2013, pp. 40–47.

[16] O. Y. Al-Jarrah, A. Siddiqui, M. Elsalamouny, P. D. Yoo, S. Muhaidat, and K. Kim, "Machine-learning-based feature selection techniques for large-scale network intrusion detection," in *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, June 2014, pp. 177–181.

[17] M. Mayhew, M. Atighetchi, A. Adler, and R. Greenstadt, "Use of machine learning in big data analytics for insider threat detection," in *MILCOM 2015 - 2015 IEEE Military Communications Conference*, Oct 2015, pp. 915–922.

[18] K. Wrona, S. Oudkerk, A. Armando, S. Ranise, R. Traverso, L. Ferrari, and R. McEvoy, "Assisted content-based labelling and classification of documents," in *2016 International Conference on Military Communications and Information Systems (ICMCIS)*, May 2016, pp. 1–7.

[19] K. Chen and D. Wang, "An aspect-oriented approach to privacy-aware access control," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 5, Aug 2007, pp. 3016–3021.

[20] S. Otoum, B. Kantarci, and H. Mouftah, "Adaptively supervised and intrusion-aware data aggregation for wireless sensor clusters in critical infrastructures," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[21] Z. Rauen, F. Anjomshoa, and B. Kantarci, "Gesture and sociability-based continuous authentication on smart mobile devices," in *16th ACM International Symposium on Mobility Management and Wireless Access (ACM MobiWac)*, October 2018 (accepted).

[22] A. El Masri, H. Wechsler, P. Likarish, C. Grayson, C. Pu, D. Al-Arayed, and B. B. Kang, "Active authentication using scrolling behaviors," in *2015 6th International Conference on Information and Communication Systems (ICICS)*, April 2015, pp. 257–262.

[23] A. Mansour, M. Sadik, E. Sabir, and M. Azmi, "A context-aware multimodal biometric authentication for cloud-empowered systems," in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct 2016, pp. 278–285.