Ensemble-based Efficient Anomaly Detection for Smart Building Control Systems

Nur Imtiazul Haque*, Mohammad Ashiqur Rahman*, and Hossain Shahriar†
*Analytics for Cyber Defense (ACyD) Lab, Florida International University, USA
†Department of Information Technology, Kennesaw State University, USA
*{nhaqu004, mngou002@fiu.edu, marahman}@fiu.edu, †hshahria@kennesaw.edu

***Abstract*—Modern building control systems integrate the internet of things (IoT) for real-time monitoring of the building's demand and manage the heating, ventilation, and air conditioning (HVAC) cost-efficiently and reliably. However, adversarial alterations of the sensor data can disrupt the occupants' comfort or increase energy consumption. Several intrusion detection systems (IDSs) are proposed to detect the tempering of the sensor measurements. However, these approaches either demonstrate a high false alarm rate or fail to detect anomalies, putting the HVAC control or the building occupants in a vulnerable condition. This paper proposes a novel intrusion detection technique amalgamating two unsupervised machine learning techniques, namely autoencoder(AE) and one-class support vector machine (OCSVM), for identifying abnormality in smart building sensor measurements. Our experimental analysis shows that the AE model-based anomaly detector demonstrates satisfactory performance for lowering false alarms but fails to detect a number of anomalous samples. In contrast, the OCSVM-based anomaly detection model performs significantly well for anomaly detection while raises a lot of false alarms. Our proposed ensembled AE-OCSVM model combines both models' benefits, resulting in significant reductions of false positive and false negative rates compared to the existing smart building IDSs. We evaluate the proposed intrusion detection system on the commercial occupancy dataset (COD) and find that the proposed IDS model can achieve a 99.6% F1-score.**

***Index Terms*—Machine learning, unsupervised learning, intrusion detection**

I. Introduction

Internet of things-enabled smart building control systems contributes towards utility cost reduction, effectiveness improvement, better prediction-based maintenance, and resource utilization for a wide range of control loops like Heating, Ventilation, and Cooling (HVAC) control, smart lighting or window control, audio or visual control, etc. [1]. The HVAC control system accounts for 40% to 70% energy of the smart building operational cost [2]. Ensuring comfort for the building occupants by maintaining the indoor air quality (IAQ) and keeping the temperature in the satisfactory range is the smart building HVAC control system's key responsibility.

Cyberattacks are now more common and in smart building control systems. At the beginning of 2019, Kaspersky analyzed around 40 thousand buildings and reported that more than one-third of the computers associated with smart building control and automation system is infected with malware [3]. This malware is mainly used to steal information from the building. Using that information, adversaries can inject calculated measurements in the HVAC control sensor devices to compromise occupants' comfort or increase building energy consumption expenditure. Current researches have found that almost 8000 HVAC control devices are vulnerable to various cyberattacks [4]. Novel zero-day attacks are being launched due to the computational advances and rise of wireless network-connected IoT devices [5]. For instance, Zhu et al. demonstrated the feasibility of a passive attack for stealing occupancy information in a specific zone of a building when there is at least one IoT device in each zone [6]. A detail analysis of security, privacy and threats have been analyzed and explored by recent researches [7], [8]. Current researches propose different IDS models for detecting attacks in the control system. But the main limitations of the approaches include a high false alarm rate and massive computational requirement, making it infeasible for real-time implementation.

In our work, we propose an unsupervised machine learning-based IDS that ensembles autoencoder (AE) and one-class SVM (OCSVM). For evaluating our work, we train our IDS with benign samples collected from the commercial occupancy dataset (COD) [9]. We generate several attack samples by injecting false data in the sensor measurements and compared our IDS performance based on correctly identifying benign and anomalous samples. For generating the attack vectors, we mathematically model an HVAC control system based on American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standards, mass balance, and energy balance equations [10], [11]. The attacks were developed considering cost increment and comfort disruption goals. The evaluation result shows that our proposed IDS can detect almost all attacks, although it was not trained on any of them. The ensembled OCSVM and AE model shows a lot less false positive and false negative predictions than the individual learners. To the best of our knowledge, this the only research attempt to ensemble AE and OCSVM models for constructing IDS. In summary, the contribution of this paper is as follows:

- We mathematically model smart building HVAC control system and generate an attack dataset considering stealthy data injection attacks.
- We propose an IDS based on the ensemble of OCSVM and AE techniques for efficient detection of corrupted or tampered data in smart building HVAC control systems.
- We extensively evaluate the performance of our proposed

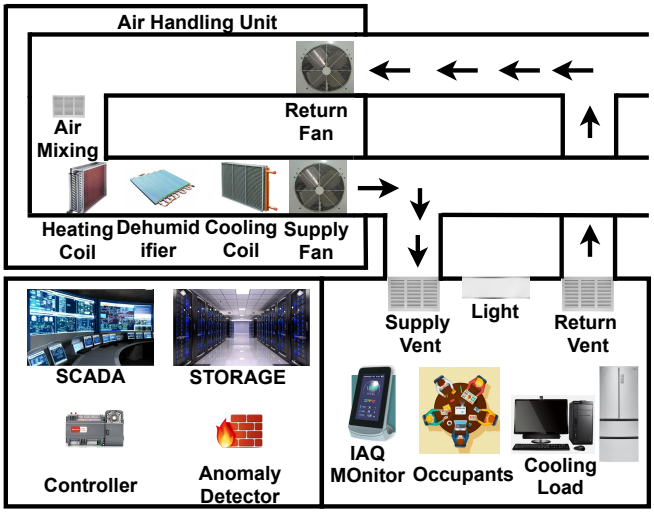

Fig. 1. A schematic diagram of a smart building HVAC control system.

IDS and compare its performance with non-ensemble-based IDSs.

The rest of the paper is organized as follows: we provide an overview and mathematical modeling of the smart building HVAC control system in Section III. For evaluating the system, we generate a set of attack data. The attack model is described in Section IV. We present the technical details of the proposed IDS in Section V. Two example case studies are discussed in Section VI for better understand the proposed IDS model. Then, we evaluate our proposed IDS model by running experiments on a state-of-the-art dataset and present the results in Section VII. Finally, we conclude the paper in Section VIII.

II. Related Works

For detecting the sensor measurement manipulation, several researchers proposed different schemes for intrusion detection systems (IDS) to find anomalies in the smart building control system. For example, Pan et al. proposed a context-aware building automation and control network (BACnet) data structure based IDS, which does not perform well for reducing false-positive alarms [12]. Some research approaches generated rule-based IDS. Luo et al. came up with a lightweight rule-based IDS framework of smart building control system [13]. The drawback of the IDS is that it is dependent on attack samples and is susceptible to misclassifying zero-day attacks.

Current researches focus on unsupervised learning-based IDS models to deal with novel zero-day attacks. Liu et al. proposed a long short-term memory (LSTM) encoder-decoder for detecting context and point anomaly in industrial climate control [14]. The false-positive rate of the proposed model is very high (13%). Gardner et al. applied an OCSVM-based novelty detection framework for detecting Intracranial EEG seizures. They haven't got good results for false-positive cases as well [15]. Jaikumar et al. presented an unsupervised learning strategy for multi-modal sensor data anomaly detection [16]. But the LSTM model-based IDS does not show satisfactory results in the case of fewer training samples or lack of time-series correlation of training features. Araya et al. proposed a generic collective contextual anomaly detection (CCAD) framework using a sliding window approach and constructed an AE-based IDS model to learn normal consumption patterns [17]. But they also obtained an abysmal performance in the case of false-positive rate.

Some researchers adopted ensemble learning strategies for IDS development in supervised and unsupervised learning settings to obtain improved performance [18]. Perdisci et al. [19] constructed an OCSVM-based ensembled classifier for detecting anomalies in payload-based system. Garg et al. proposed a meta-heuristic clustering approach, an ensemble of Artificial Bee colony-based IDS for multi-class dataset anomaly detection [20]. But their approach is dependent on the availability of attack data. Zhang et al.proposed a two-level unsupervised ensemble learning strategy, where the first level is used to reduce the loss of information, and the second level is used to improve the generalization ability [21]. Some other researches also came up with unique IDS solutions. Newaz et al. proposed a novel IDS on personalized medical device communication using n-gram approach [22]. Shahriar et al. proposed a generative adversarial network (GAN)-based approach to generate a synthetic attack dataset from the existing attack data [23]. But their IDS was dependent on learning known attack pattern. Chan et al. proposed an neural network-based ensemble technique for novelty detection [24].

Although these approaches helped design more accurate learners for anomaly/ intrusion detection, they recognize many benign samples as an anomaly. Hence, we are interested in ensembling different learning models together to combine their positive features.

III. Smart Building HVAC Control System

This section provides a brief description of the smart building HVAC control architecture with mathematical model representation. Fig. 1 shows the smart building HVAC control schematic diagram with two rooms, where the left room is responsible for operation and maintenance, the right one is a normal office room.

A. Control Algorithms

Maintaining good indoor air quality (IAQ) by lowering the amount of CO_2 and other volatile organic gas is one of the key responsibilities of HVAC control systems [25]. Fresh air from outside needs to be injected inside the building to keep IAQ in an acceptable range. Again, the HVAC control system also accounts for maintaining the indoor temperature in occupants' comfort range. In real life, temperature and CO_2 dynamics are intricate, as there are many variants. For maintaining good IAQ and temperature, modern smart building HVAC control system adopts model predictive control-based strategy. But mass balance equation for CO_2 control and energy balance equation for temperature control shows good performance in

HVAC control when building properties and the exact number of occupants are known [11].

At a particular timeslot in a particular zone, volumetric airflow from the supply fan can be calculated using the following equation:

$$\frac{\mathcal{M}^{occ} \times Occ^{CO_2} \times \Delta t}{Zone^{Vol}} = set^{CO_2} - \left(1 - \frac{Air^{Vol} \times \Delta t}{Zone^{Vol}}\right) \mathcal{M}^{CO_2} - \frac{Air^{Vol} \times \Delta t}{Zone^{Vol}} Mixed^{CO_2} \tag{1}$$

where,
$\mathcal{M}^{occ}$ = Occupant sensor measurements (person)
Occ^{CO_2} = CO_2 emission per occupant in the considered zone ($ft^3 min^{-1}$)
$\mathcal{M}^{CO_2}$ = CO_2 sensor measurements (ppm)
set^{CO_2} = CO_2 setpoint (ppm)
Air^{Vol} = Volumetric airflow of mixed air inn the considered zone ($ft^3 min^{-1}$)
$Zone^{Vol}$ = Volume of the considered zone (ft^3)
Δt = Difference between two timeslots (min)
$Mixed^{CO_2}$ = CO_2 concentration of mixed air (ppm).

Temperature dynamics follows the energy balance equation as follows:

$$Air^{Mass} \times Mixed^{SH}(set^{Temp} - supply^{Temp}) = Load^{Energy} + \mathcal{M}^{occ} Occ^{Energy} \tag{2}$$

where,
Air^{Mass}= Mass airflow in the zone (kgs^{-1})
$Mixed^{SH}$= Specific heat of mixed air ($Jkg^{-1}K^{-1}$)
set^{Temp}= Temperature setpoint of supply air (F)
$supply^{Temp}$= Temperature of supply air (F)
$Load^{Temp}$= Cooling or heating energy radiation or absorption from loads (kW)
Occ^{Temp} =Cooling or heating energy radiation or absorption from occupants (kW)

Supply air temperature varies based on the weather condition as demonstrated in the equation below.

$$55\ (Cooling) \leq supply^{Temp} \leq 90\ (Heating) \tag{3}$$

B. Cost Calculation

The cooling or heating cost calculation relies on the mixing of fresh outdoor air and indoor recirculated air. The deviation of the temperature of mixed air and supply air introduces cost in the HVAC control system. Mixed air needs to chill or be heated to reach the desired supply air temperature, which is performed by the high-speed flow of water in the coils. After cooling or heating, the temperature of coil water gets changed, which again needs to chill or be heated.

For determining the psychrometric values, six co-efficient values are needed according to ASHRAE standard (σ_0 = -5800.22, σ_1= 1.39, σ_2 = -0.049, σ_3 = 4.17 $\times 10^{-5}$, σ_4 = -1.44 $\times 10^{-8}$, σ_5 = 6.54). The psypy library is used to calculate and extract necessary psychrometric parameters from temperature and relative humidity [26].

Calculating Partial pressure of water is needed to determine specific heat, specific volume, and enthalpy of the supply and mixed air.

$$Mixed^{PP} = exp(\sum_{i=0}^{4} (\sigma_i Mixed^{Temp})^{(i-1)} + \sigma_5 log_e(Mixed^{Hum}))Mixed^{SH} \tag{4}$$

where,
$Mixed^{PP}$ = Partial Pressure of water for mixed (Pa)
$Mixed^{Hum}$ = Humidity of mixed air (%)
$Mixed^{SH}$ = Specific heat for mixed air ($Jkg^{-1}K^{-1}$). The above equation can be used to determine the partial pressure of supply air, $Supply^{PP}$. Specific heat of the water for mixed air can be calculated using the following equation.

$$Mixed^{SH} = \frac{0.621945 \times Mixed^{PP}}{P - Mixed^{PP}} \tag{5}$$

where,
P = Atmospheric pressure (Pa).

Similarly, we can find the specific heat of the water for supply air, $Supply^{SH}$. Enthalpy of mixed air can be determined by the following equation:

$$Mixed^{Eth} = 1.006 \times Mixed^{Temp} + Mixed^{SH} \times (2501 + 1.86 \times Mixed^{Temp}) \tag{6}$$

Similarly, enthalpy of supply air, $Supply^{Eth}$ can be measured.

The specific volume of mixed air is used to calculate the volumetric flow of air simply by dividing mass flow rate by specific volume.

$$Mixed^{SV} = 287.042 \times Mixed^{Temp} \times \frac{1 + 1.607858 \times Mixed^{SH}}{P} \tag{7}$$

The mass flow of air in the condenser can be derived from the specific heat of mixed and supply air and mixed air flow rate.

$$Cond^{Mass} = Air^{Mass} \times (Mixed^{SH} - Supply^{SH}) \tag{8}$$

The mass flow rate of condenser air can determine the heat energy required to chill the hot air mixed to supply air setpoint temperature.

$$Cost^{Coil} = Mixed^{Mass} \times (Supply^{Eth} - Mixed^{Eth}) + Cond^{Mass} \times Cond^{Eth} \tag{9}$$

But for chilling the mixed air, the temperature of the coil refrigerant rises above the required temperature, and the temperature rise can be calculated using the following mathematical representation.

$$Coil^{Temp} = Set^{Coil} + \frac{Cost^{Coil}}{Coil^{Mass} \times Water^{SH}} \tag{10}$$

Hence, the refrigerants are passed to the chiller for cooling it back to the normal temperature, which also adds cost to the

system. This cost associated with the chiller cooling accounts for a similar amount of energy that was required in the coil cooling.

$$Cost^{Chil} = Coil^{Mass} \times Water^{SH} \times (Coil^{Temp} - Set^{Coil}) \tag{11}$$

where,
$Water^{SH}$ = Specific heat of water ($Jkg^{-1}K^{-1}$).

The overall cost of HVAC control cost is dependent on the coil cost and the chiller cost for the cooling condition.

$$Cost^{HVAC} = Cost^{Coil} + Cost^{Chil} \tag{12}$$

IV. Attack Model

To evaluate the proposed IDS system, we need attack samples. We consider a stealthy false data injection attack to be carried out in the sensor measurements of the smart building control system. Attack modeling is the procedure of analyzing attack goals given the attacker's knowledge, accessibility, and capabilities to launch and attack [27]. In our attack model, we have the following assumptions:

- The building entrance is exceptionally secured through biometric sensors and the controller gets alarmed while experiencing inconsistency in occupant count.
- The total number of the occupant in a zone should not exceed the maximum zone capacity.
- The controller's security system verifies the current timeslot sensor and actuator measurements with the past sensor measurements.
- The attacker can sniff all the sensor measurements throughout the day and calculate corresponding actuation measurements using the knowledge about the HVAC control model.

A. Attacker Knowledge

We consider a knowledgeable attacker in the attack model who knows about the building properties, building topology, occupancy pattern, control, and defense mechanisms of the smart building HVAC control system. The attacker also knows about the weather pattern outside the building, which helps to imitate the exact control decisions. The attacker also knows about the time-series data analysis capability of the controller. As a result, the attacker does not alter sensor measurements drastically and can remain stealthy.

B. Attacker's Accessibility and capability

In our attack model, we consider that attacker has access to all the sensor measurements at every timeslot throughout the day. But the attacker can not perform random manipulation of sensor measurements and still gets undetected. Alteration of sensor measurements would not be stealthy if any sensor measurement got out of bounds from the feasible range. For example, in a 50 person capacity zone, the occupancy sensor measurement shows 100 person or a CO_2 sensor measurement is demonstrating 4000 ppm. Attackers are restricted from compromising the sensor measurements in a feasible range. Again, if there is no occupant present in the building, the attacker cannot launch any attack be prevent getting detected by the controller's measurement verifier.

C. Attack Goal

We consider two attack intents in our attack model.

- Overall building operational and energy cost increment.
- Maximize occupant's discomfort by deteriorating IAQ beyond comfort threshold or changing the temperature beyond temperature setpoint.

D. Attack Technique

The attacker manipulates the sensor measurements based on his/her for attaining his/her attack goal. As the attacker is knowledgeable about the control system's verifier, he/she alters the sensor measurements consistently to bypass this security. Hence, while crafting with occupancy sensor measurements in the zones, the attacker keeps the total number of occupant unchanged. We term this attack technique as **swapping occupant** as the attacker is adding people in some zone while removing a similar number of people from the other zones.

E. Adversarial Sample Generation

An attack vector is defined as the set of measurements, injecting a benign sample with an adversarial sample [28]. Algorithm 1 shows the process the generating an adversarial sample. The algorithm takes sensor measurements, adversarial intent, and thresholds to change each sensor measurements as input and produce the attack vector. If the attacker intends to increase HVAC control cost, we term that attack as **cost increment** attack. In this case, the attacker solves an optimization problem of maximizing the total cost associated with HVAC control by tempering the sensor measurements within an acceptable threshold. But attacker does not alter the sensor measurements arbitrarily. The attacker swaps the occupancy within the zones, so that number of occupants of the building is consistent. As there is a verifier in the controller, the attacker tries to alter other sensor measurements accordingly to make the attack stealthy. The **comfort disruption** attack aims at disturbing maximum occupants by altering temperature, humidity, and CO_2 concentration of the zones away from the comfort range. The constraints of this attack is similar to the cost increment attack with a different objective. To avoid getting detected, the attacker also restricts swapping the number of people in the attack vector generation. We have used python API of Z3 SMT solver for both solving and optimizing our modeling constraints [29].

V. Proposed IDS Model

Our proposed machine learning-based IDS model learns the pattern of positive sensor measurements from historical data distribution and identifies anomalous data samples based on the learned pattern deviation. In this section, We provide an overview of the proposed IDS model based on the flow of Algorithm 2. The process of intrusion detection can be divided into four stages.

Algorithm 1: Generating Adversarial Samples

Input: $\mathcal{M}$, I, Th
Output: δ
if I == *'Cost'* **then**

$$\underset{\delta}{\text{maximize}} \quad HVACCost(\mathcal{M}, \delta) \tag{13a}$$

subject to

$$\forall_{z \in \mathcal{Z}} \mathcal{M}_{tz} \coloneqq \mathcal{M}_{tz} + \delta_z, \tag{13b}$$

$$\forall_{z \in \mathcal{Z}} - Th_z \leq \delta_z \leq Th_z, \tag{13c}$$

$$\sum_{i=1}^{|\mathcal{Z}|} \delta_i = 0 \tag{13d}$$

else if I == *'Comfort'* **then**

$$\underset{\delta}{\text{maximize}} \quad deviation(\mathcal{M}^{Comf}, \delta^{Comf}) \tag{14a}$$

$$\text{subject to} \quad \forall_{z \in \mathcal{Z}} \mathcal{M}_{tz}^{Comf} \coloneqq \mathcal{M}_{tz}^{Comf} + \delta_z^{Comf}, \tag{14b}$$

$$\forall_{z \in \mathcal{Z}} - Th_z \leq \delta_z \leq Th_z \tag{14c}$$

Algorithm 2: Proposed Model

Input: $\mathbb{X}$, $\mathbb{T}$
Output: $predSample$
Train Autoencoder-based ADS model, $model^{AE}$ on $\mathbb{X}$
Train OCSVM-based ADS model, $model^{OCSVM}$ on $pca(\mathbb{X})$
$threshAE \coloneqq max(abs(\mathbb{X} - model^{AE}(\mathbb{X})))$
$threshOCSVM \coloneqq 0$
for *each sample in* $\mathbb{T}$ **do**
  $weightAE \coloneqq abs(sample - model^{AE}(sample))$
  $weightOCSVM \coloneqq model^{OCSVM}(pca(sample))$
  $normWeightAE \coloneqq$
   $abs(normalize(weightAE) -$
   $normalize(threshAE))$
  $normWeightOCSVM \coloneqq$
   $abs(normalize(weightOCSVM) -$
   $normalize(threshOCSVM))$
  if $weightAE < threshAE$ **then** $predAE$ = 1
  else $predAE$ = -1
  if $weightOCSVM < threshOCSVM$ **then**
   $predOCSVM$ = -1
  else $predOCSVM$ = 1
  $weightSample =$
   $\frac{1}{2} \times (normWeightAE \times predAE +$
   $normWeightOCSVM \times predOCSVM)$
  if $weightSample < 0$ **then** $predSample$[sample]
   = "Anomaly"
  else $predSample$[sample] = "Benign"

A. Data Collection and Preprocessing

The initial stage of intrusion detection is a benign dataset creation containing all sensor measurements and corresponding actuation measurements. The sensor and actuation measurements are the features of the dataset. After the benign dataset is generated, duplicate entries are removed from it. Then dataset features are normalized using a standard normalization process. The feature space is reduced through a principle component analysis technique, which facilitates faster OCSVM model training [30]. The AE model is trained on all features. Finally, the whole benign dataset is split into two parts, where 75% of the sample are stored for model training, and the rest of the data is set aside for testing the model's performance on positive samples. Again, an attack dataset is also prepared from the preprocessed benign dataset based on the attack technique discussed in Section IV.

B. Model Training

The preprocessed data are used for model training. The training section involves training two separate models.

1) Autoencoder: The autoencoder-based ML is a neural network (NN) model that regenerates the features of the model [31], [32]. NN is a machine model that can learn the non-linear pattern from a large set of feature relationships. The input of the nodes of the NN model is the sum of the product of the weight and output of the previous nodes and a bias value. The output of the nodes is passed through an activation function for adding non-linearity in the model. The activation function is very important for capturing non-linear boundaries in the feature space for non-linear mapping between the input features and target. The weights and biases are initially assigned with random values. These model parameters are then tuned using a backpropagation process for reducing the error between the model prediction and target. Again, the regularization process is also applied in the training process to generalize the model and prevent over-fitting with the training samples.

The main difference between the Autoencoder model and a general NN model is that there is no strict restriction in the NN model for specifying the number of nodes in the hidden layer. But in the case of the AE model, the number of nodes in the hidden layer in the encoder portion should be less than the number of nodes in the previous layer, and the opposite is for the decoder portion.

2) One-class SVM: OCSVM model is a variation from support vector machine (SVM) model for detecting novel patterns [33], [34]. Support vector machine is a supervised learning technique in which specialized techniques are used to separate different classes by drawing hyper-planes. But SVM model requires a label for the training samples. As our data of interest are unlabeled, we leverage an unsupervised OCSVM model, which separates the trained patterns from the origin using a decision boundary. The model's decision boundary can be modified by tuning two hyperparameters γ and ν.

In our model training, the AE model, $model^{AE}$ gets trained on the historical samples, while the OCSVM model,

TABLE I
ZONE PROPERTIES OF COD DATASET

Zones	Volume (ft^3)	Occ^{CO_2} (CFM)	Occ^{Energy} (kW)	$Load^{Energy}$ (kW)	Capacity (Person)
Entrance	12570	0.022	0.108	0.72	50
Clemente	11688	0.021	0.107	0.30	10
Warhol	10911	0.018	0.092	0.45	25
Laboratory	17937	0.025	0.120	0.83	40

$model^{OCSVM}$ gets trained on the PCA components of training samples. After the models are trained, they are applied to the benign training dataset to calculate the benign and anomalous sample separation threshold.

C. Threshold Calculation

Threshold calculation is an essential part in the case of anomaly detection. Again, the weight of a model prediction signifies the distance of the predicted value from the model's decision threshold. In OCSVM model, a decision function returns a score in between -1 to $+1$ [33]. The score is a continuous value, and this value can be directly used as a weight of the OCSVM model for a particular sample. The threshold for the OCSVM model is 0. AE-based learning approach returns prediction for feature regeneration. The deviation between the actual features and the predicted features is said to be an error. Algorithm 2 shows that the AE model threshold, $threshAE$ is the maximum error between the training predictions and corresponding features. The thresholds of the models are on a different scale. The threshold values are normalized to use in anomaly detection.

D. Anomaly Detection

Finally, for detecting whether a sample is anomalous or not, our proposed IDS uses the trained models, $model^{AE}$ and $model^{OCSVM}$ along with their thresholds, $threshAE$ and $threshOCSVM$. After that, the AE and the OCSVM models assign score, $weightAE$ and $weightOCSVM$ respectively for the sample based on the distance from $threshAE$ and $threshOCSVM$. The direction of distances determine the prediction of the models, $predAE$ and $predOCSVM$. If both models' prediction differs, then the normalized weights of the models, $normWeightAE$ and $normWeightOCSVM$ are multiplied by the prediction to evaluate the ensembled outcome.

VI. EXAMPLE CASE STUDIES

This section provides two example scenarios from the COD dataset with numeric data to clarify how the proposed IDS works and to illustrate the need for the proposed ensemble model. Table I shows the properties of the considered four zones of the COD dataset. A detailed explanation of the dataset can be found in Section VII-A. The heat radiation and CO_2 emission of occupants are generated based on the regular metabolic rate of human [35] and other cooling/ heating loads are estimated based on the properties of the zones [36]. Table II shows important sensor and actuation measurements and corresponding principal components. Only two principal components are used as the sum of explained variance ratio of the components is greater than 0.99.

A. Case Study 1

Case study 1 is performed on a benign sample. The OCSVM model assigned $weightOCSVM$ to be -0.038 and counted the sample as an anomalous sample as described in Algorithm 2. On the other hand, the $weightAE$ was calculated as 0.41, which is less than the $threshAE$ = (3.38). Hence AE model labeled the sample as benign. The normalized OCSVM threshold, $normalize(threshOCSVM))$ and score, $normalize(weightOCSVM)$ are calculated to be 0.84 and 0.28 respectively. Therefore, using δ = 0.5, the measured $normWeightOCSVM$ is 0.28. Similarly, $normWeightAE$ is calculated to be 0.46. As, value of $normWeightAE$ is greater than $normWeightOCSVM$, the ensembled decision follows prediction of AE. As a result, the sample is identified as benign by the proposed model, although the OCSVM model is labeled an anomaly. Thus the proposed model helps to lower the false anomaly rate.

B. Case Study 2

Case study 2 is performed on two attack samples. The attack samples are obtained from a benign sample considering two different cost increment attacks. The first attack is performed assuming that there is no restriction on the number of people swapping among the zones. The second attack is carried out considering that the attacker swaps at most one occupant to remain stealthy.

In the non-restricted attack, The OCSVM model assigned $weightOCSVM$ to be -0.06 and counted the sample as an anomalous sample. On the other hand, the $weightAE$ was calculated as 2.57, which is less than the $threshAE$. Hence AE model labeled the sample as benign. Using δ = 0.5, the measured $normWeightOCSVM$ is 0.42. Similarly, $normWeightAE$ is calculated to be 0.07. As, value of $normWeightAE$ is way less than $normWeightOCSVM$, the ensembled decision follows prediction of OCSVM. As a result, the sample is identified as an anomaly by the proposed model, although the AE model labeled is as benign.

In the non-restricted attack, The OCSVM model assigned $weightOCSVM$ to be -0.06 and counted the sample as an anomalous sample. On the other hand, the $weightAE$ was calculated as 1.23, which is less than the $threshAE$. Hence AE model labeled the sample as benign. Using δ = 0.5, the measured $normWeightOCSVM$ is 0.42. Similarly, $normWeightAE$ is calculated to be 0.39. As, value of $normWeightAE$ is marginally than $normWeightOCSVM$, the ensembled decision follows prediction of OCSVM. As a result, the sample is identified as an anomaly by the proposed model, although the AE model labeled it as benign. Thus, the proposed model reduces the false benign rate of both restricted and non-restricted cost increment attacks. In the case of comfort disruption attack, similar benefits are served by the proposed model while providing equal weight to both models.

TABLE II
EXAMPLE CASE STUDIES

Case Study	Sample Type	Sensor Measurements								Principal Component	
		$\mathcal{M}^{Occ}_{1,t}$	$\mathcal{M}^{Occ}_{2,t}$	$\mathcal{M}^{Occ}_{3,t}$	$\mathcal{M}^{Occ}_{4,t}$	$Air^{Vol}_{1,t}$	$Air^{Vol}_{2,t}$	$Air^{Vol}_{3,t}$	$Air^{Vol}_{4,t}$	PC_1	PC_2
1	Benign	28	2	1	21	147	22	39	127	-40.72	28.98
2	Benign	30	4	11	22	515	22	91	423	433.12	57.35
	Attacked (No Restriction)	35	0	0	32	593	17	33	528	761.78	708.38
	Attacked (1 Person Restriction)	30	4	10	23	565	34	112	472	613.49	253.78

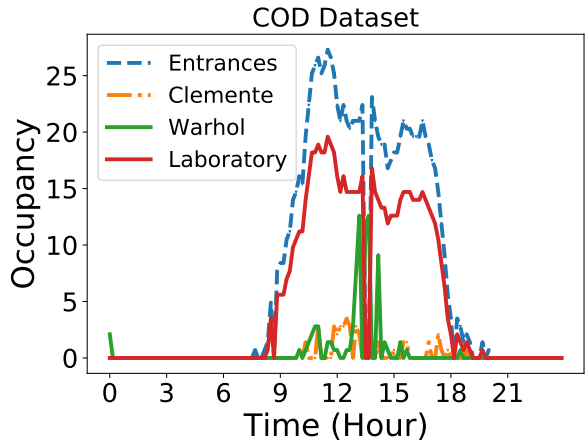

Fig. 2. COD dataset average occupant frequency at different time of the day.

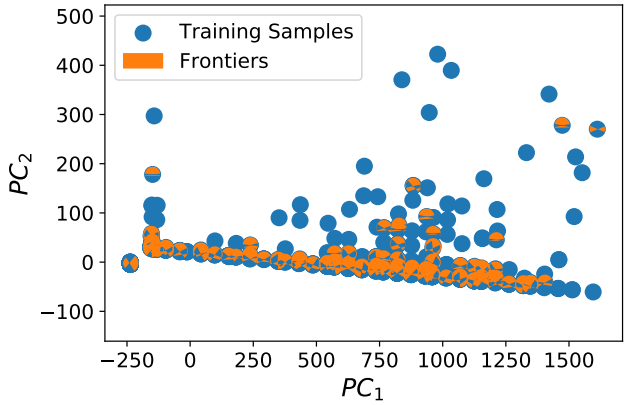

Fig. 3. Learnt Frontier of OCSVM model.

VII. EVALUATION

We extensively evaluate our proposed IDS's performance in detecting anomalies/intrusions, along with its scalability in terms of execution time.

A. Dataset Description

The COD dataset is generated from a commercial office building located in Pittsburgh, Pennsylvania [9]. The preprocessed COD dataset has almost 12357 time-series samples and 22 features collected throughout the year (2015-2016), and among them, 3666 samples were attacked using the attack technique discussed in section IV-D. The sensor measurements of the rest of the samples could not be attacked due to the absence of occupants in all the zones. The occupancy pattern of the processed COD dataset is normally distributed, and the average occupancy frequency of each zone at different time intervals is illustrated in Fig. 2. The average occupancy (95% confidence interval) of the entrance, Clemente, Warhol, and laboratory zone are (7.9 - 12.3), (0.2 - 0.5), (0.3 - 1.3), and (6.1 - 9.7) respectively. The zones' average indoor temperature ranges between (53.3 - 56.9) °F, and relative humidity (63.39 - 65.96)% in the case of 95% confidence interval. The outdoor climate data is collected from the Pennsylvania state climatologist website [37].

B. Evaluation of Proposed IDS against Cost Increment and Comfort disruption Attacks

The learnt frontier of the OCSVM model with benign data points is shown in Fig. 3 with γ = 0.03 and ν = 0.003. It seems that the model has not completely learned the pattern of the distributed points. Tuning the γ and ν parameters allows capturing more training points. But the ultimate performance of the model to identify true benign samples gets deteriorated because the frontier shifts away from the dense regions. The AE model is trained on a simple NN model with a single hidden layer consisting of 10 nodes. The models' parameters are determined by tuning and observing the performance of the models on the historical benign sensor and actuator measurements.

Fig. 4(a), 4(b), 4(c), and 4(d) demonstrates performance of OCSVM model and AE model against cost increment attack for no swapping restriction, 1-person restriction, 3-person restriction, and 5-person restriction respectively. Similarly, performance of individual learner models on restricted and non-restricted comfort disruption attack are shown in Figure 5(a), 5(b), 5(c),and 5(d). From the figures, we can observe a lot of such points where both models differ in their opinions. The proposed IDS ensembles both of them and can produce correct labeling in most cases, as seen in Table III. The OCSVM model does not provide any false benign labeling, while the AE model does not generate any false anomaly labeling in all considered attacks. That's why those legends are missing from the Figures. It is clear from Table III that the proposed model is better than individual OCSVM and AE models based on accuracy, precision, recall, and F1-score performance metrics for all cost increment and comfort disruption attacks [38].

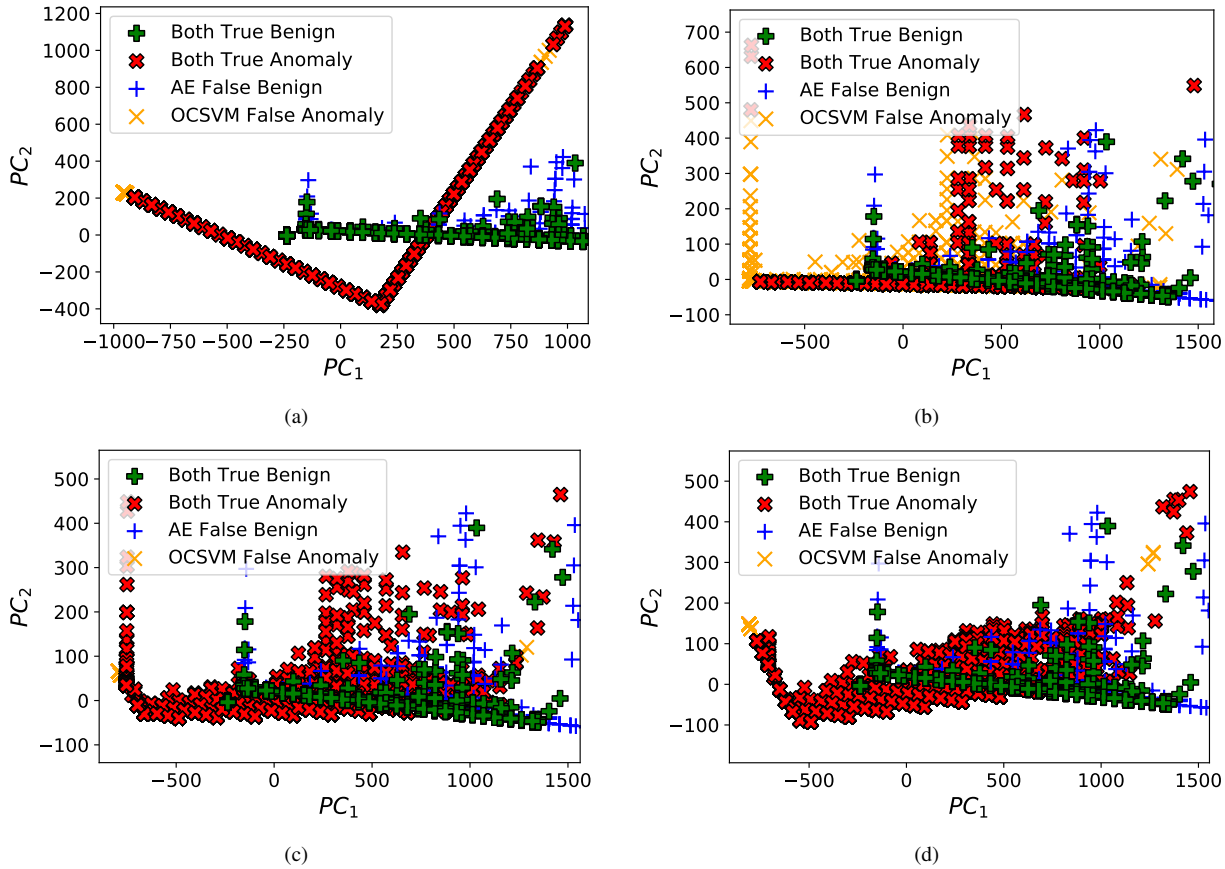

Fig. 4. OCSVM and AE-based IDS performance against cost increment attack with (a) no person restriction, (b) 1-person restriction, (c) 3-person restriction, and (d) 5-person restriction.

TABLE III
PERFORMANCE COMPARISON OF INDIVIDUAL IDS AND PROPOSED ENSEMBLE IDS FOR VARIOUS ATTACK CONDITIONS

Attack	Occupant Swapping Restriction	IDS Model	True Anomaly	True Benign	False Anomaly	False Benign	Accuracy	Precision	Recall	F1-Score
Cost Increment	No	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	3482	3666	0	184	97.5%	100.0%	95.0%	97.4%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
	1 Person	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	2343	3666	0	1323	82.0%	100.0%	63.9%	78.0%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
	3 Person	OCSVM	3613	3476	190	53	96.7%	95.0%	98.6%	96.7%
		AE	3484	3666	0	182	97.5%	100.0%	95.0%	97.5%
		Proposed	3664	3638	28	2	99.6%	99.2%	99.9%	99.6%
	5 Person	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	3432	3666	0	234	96.8%	100.0%	93.6%	96.7%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
Comfort Disruption	No	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	2712	3666	0	954	87.0%	100.0%	74.0%	85.0%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
	1 Person	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	2179	3666	0	1487	79.7%	100.0%	59.4%	74.6%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
	3 Person	OCSVM	3666	3476	190	0	97.4%	95.1%	100.0%	97.5%
		AE	3158	3666	0	508	93.1%	100.0%	86.1%	92.6%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%
	5 Person	OCSVM	3662	3476	190	4	97.4%	95.1%	99.9%	97.4%
		AE	2999	3666	0	667	90.9%	100.0%	81.8%	90.0%
		Proposed	3666	3638	28	0	99.6%	99.2%	100.0%	99.6%

The anomalous samples are counted as positive samples for evaluating the performance metrics, and benign samples are identified as negative samples. The accuracy metric provides the ratio of correctly identified samples with respect to all

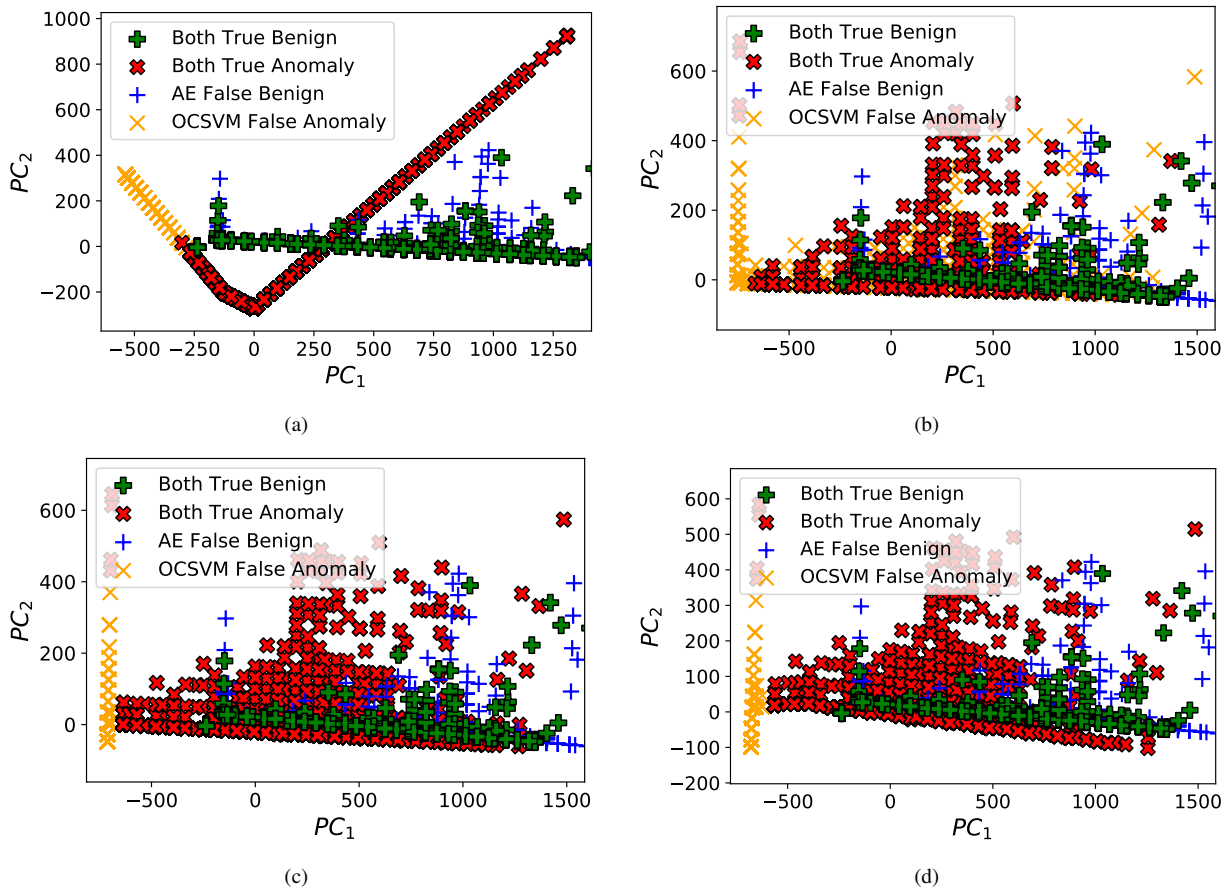

Fig. 5. IDS performance against comfort disruption with (a) no person restriction, (b) 1-person restriction, (c) 3-person restriction, and (d) 5-person restriction.

samples, while precision comes up with correctly identified anomalous samples over all IDS labeled anomalous samples. Again, the recall metric provides us the measure of correctly identified anomalous samples over all actual anomalous samples. Finally, the F1-score gives the harmonic mean of precision and recall considering both false positive and negative samples. Table III shows that although in some cases AE shows better precision than the proposed model, the F1-score is higher for all the cases [39].

C. Evaluation of the IDS's Scalability

The linear relationship between the number of features and required execution time, as shown in Fig. 6 indicates the feasibility of implementation of the proposed IDS model. The average execution time for 8-feature control systems is found to be 2.6 ms and, for 22-feature system, the execution time has risen up to 5.3 ms. The IDS is tested on Dell Precision 7920 Tower workstation with Intel Xeon Silver 4110 CPU @3.0GHz, 32 GB memory, 4 GB NVIDIA Quadro P1000 GPU.

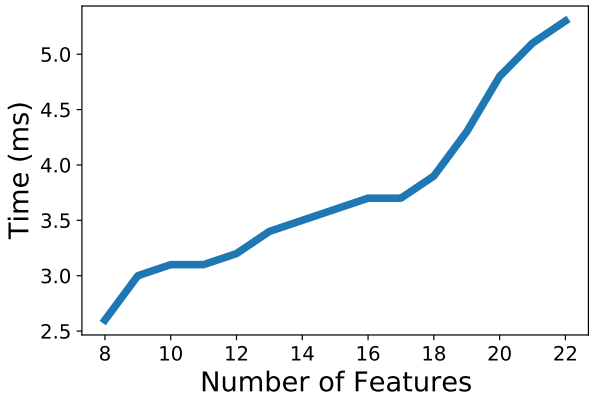

Fig. 6. Execution time of the anomaly detection with the increasing the number of features.

VIII. Conclusion

The smart building control system is becoming more vulnerable day by day due to novel cyberattacks posing health and economic concerns to the building occupants. This work proposes a novel ensemble learning-based IDS that combines the benefits of two different unsupervised ML models, AE and OCSVM. We generate attack samples for evaluating our IDS considering cost increment and comfort disruption-based stealthy false data injection attack. Our experimentation on the COD dataset shows that this ensemble technique provides a significant performance boost, providing up to 99.6% F1-score. In the future, we will explore the ensemble technique

with other unsupervised ML approaches for better novelty detection. We will also consider verifying our IDS against other attacks and control systems in our future works.

REFERENCES

[1] 5 key benefits of smart buildings. https://www.trueoccupancy.com/blog/5-key-benefits-of-smart-buildings, 2019. Accessed: 2021-01-10.

[2] Learn how smart hvac technology is designed to modernize your workplace. https://serraview.com/smart-hvac-sensor-technology-smart-buildings. Accessed: 2021-01-23.

[3] Smart buildings at high risk for cyber attacks: Study. https://www.facilitiesnet.com/buildingautomation/tip/Smart-Buildings-At-High-Risk-for-Cyber-Attacks-Study--44839, 2021. Accessed: 2021-01-11.

[4] Smart building automation systems vulnerable to cyber attack. https://inbuildingtech.com/smart-buildings/cyber-attack-smart-building-iot/, 2021. Accessed: 2021-01-15.

[5] Ahmed Aleroud and George Karabatis. Toward zero-day attack identification using linear data transformation techniques. In *2013 IEEE 7th International Conference on Software Security and Reliability*, pages 159–168. IEEE, 2013.

[6] Yanzi Zhu, Zhujun Xiao, Yuxin Chen, Zhijing Li, Max Liu, Ben Y Zhao, and Haitao Zheng. Et tu alexa? when commodity wifi devices turn into adversarial motion sensors. *arXiv preprint arXiv:1810.10109*, 2018.

[7] AKM Newaz, Amit Kumar Sikder, Mohammad Ashiqur Rahman, and A Selcuk Uluagac. A survey on security and privacy issues in modern healthcare systems: Attacks and defenses. *arXiv preprint arXiv:2005.07359*, 2020.

[8] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A Selcuk Uluagac. A survey on sensor-based threats and attacks to smart devices and applications. *IEEE Communications Surveys & Tutorials*, 2021.

[9] Kin Sum Liu, Elvin Vindel Pinto, Sirajum Munir, Jonathan Francis, Charles Shelton, Mario Berges, and Shan Lin. Cod: a dataset of commercial building occupancy traces. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, pages 1–2, 2017.

[10] Ansi/ashrae standard 62.1-2019: Ventilation for acceptable indoor air quality. https://www.ashrae.org/technical-resources/bookstore/standards-62-1-62-2. Accessed: 2021-01-28.

[11] Davide Cali, Peter Matthes, Kristian Huchtemann, Rita Streblow, and Dirk Müller. Co2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment*, 86:39–49, 2015.

[12] Zhiwen Pan, Salim Hariri, and Jesus Pacheco. Context aware intrusion detection for building automation systems. *Computers & Security*, 85:181–201, 2019.

[13] Hong Luo, Ruosi Wang, and Xinming Li. A rule verification and resolution framework in smart building system. In *2013 International Conference on Parallel and Distributed Systems*, pages 438–439. IEEE, 2013.

[14] Yu Liu, Zhibo Pang, Magnus Karlsson, and Shaofang Gong. Anomaly detection based on machine learning in iot-based vertical plant wall for indoor climate control. *Building and Environment*, 183:107212, 2020.

[15] Andrew B Gardner, Abba M Krieger, George Vachtsevanos, Brian Litt, and Leslie Pack Kaelbing. One-class novelty detection for seizure analysis from intracranial eeg. *Journal of Machine Learning Research*, 7(6), 2006.

[16] P. Jaikumar, A. Gacic, B. Andrews, and M. Dambier. Detection of anomalous events from unlabeled sensor data in smart building environments. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2268–2271, 2011.

[17] Daniel B Araya, Katarina Grolinger, Hany F ElYamany, Miriam AM Capretz, and G Bitsuamlak. Collective contextual anomaly detection framework for smart buildings. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 511–518. IEEE, 2016.

[18] Rasha F Kashef. Ensemble-based anomaly detetction using cooperative learning. In *KDD 2017 Workshop on Anomaly Detection in Finance*, pages 43–55. PMLR, 2018.

[19] Roberto Perdisci, Guofei Gu, and Wenke Lee. Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 488–498. IEEE, 2006.

[20] Sahil Garg, Kuljeet Kaur, Shalini Batra, Gagangeet Singh Aujla, Graham Morgan, Neeraj Kumar, Albert Y Zomaya, and Rajiv Ranjan. En-abc: An ensemble artificial bee colony based anomaly detection scheme for cloud environment. *Journal of Parallel and Distributed Computing*, 135:219–233, 2020.

[21] Jia Zhang, Zhiyong Li, Ke Nai, Yu Gu, and Ahmed Sallam. Delr: A double-level ensemble learning method for unsupervised anomaly detection. *Knowledge-Based Systems*, 181:104783, 2019.

[22] AKM Iqtidar Newaz, Amit Kumar Sikder, Leonardo Babun, and A Selcuk Uluagac. Heka: A novel intrusion detection system for attacks to personal medical devices. In *2020 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE, 2020.

[23] Md Hasan Shahriar, Nur Imtiazul Haque, Mohammad Ashiqur Rahman, and Miguel Alonso. G-ids: Generative adversarial networks assisted intrusion detection system. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 376–385. IEEE, 2020.

[24] Felix TS Chan, ZX Wang, S Patnaik, MK Tiwari, XP Wang, and JH Ruan. Ensemble-learning based neural networks for novelty detection in multi-class systems. *Applied Soft Computing*, 93:106396, 2020.

[25] Anjali Rajith, Sakurai Soki, and Mine Hiroshi. Real-time optimized hvac control system on top of an iot framework. In *2018 Third international conference on fog and mobile edge computing (FMEC)*, pages 181–186. IEEE, 2018.

[26] Psypy project description. https://pypi.org/project/psypy/. Accessed: 2020-03-19.

[27] Dietmar PF Möller. Attack models and scenarios. In *Cybersecurity in Digital Transformation*, pages 89–98. Springer, 2020.

[28] Marc Capellupo, Jimmy Liranzo, Md Zakirul Alam Bhuiyan, Thaier Hayajneh, and Guojun Wang. Security and attack vector analysis of iot devices. In *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, pages 593–606. Springer, 2017.

[29] Leonardo De Moura and Nikolaj Bjørner. Z3: An efficient smt solver. In *International conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 337–340. Springer, 2008.

[30] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[31] Sun-Chong Wang. Artificial neural network. In *Interdisciplinary computing in java programming*, pages 81–100. Springer, 2003.

[32] Wei Wang, Yan Huang, Yizhou Wang, and Liang Wang. Generalized autoencoder: A neural network framework for dimensionality reduction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 490–497, 2014.

[33] Yunqiang Chen, Xiang Sean Zhou, and Thomas S Huang. One-class svm for learning in image retrieval. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 34–37. IEEE, 2001.

[34] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[35] Andrew Persily and Lilian de Jonge. Carbon dioxide generation rates for building occupants. *Indoor air*, 27(5):868–879, 2017.

[36] Hvac load calculator. https://www.servicetitan.com/tools/hvac-load-calculator. Accessed: 2021-01-21.

[37] The pennsylvania state climatologist. http://www.climate.psu.edu/data/city_information/index.php?city=pit&page=dwa&type=big7. Accessed: 2020-06-21.

[38] Ilhame El Farissi, Mohammed Saber, Sara Chadli, Mohamed Emharraf, and Mohammed Ghaouth Belkasmi. The analysis performance of an intrusion detection systems based on neural network. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, pages 145–151. IEEE, 2016.

[39] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.