Regression Based Anomaly Detection in Electric Vehicle State of Charge Fluctuations Through Analysis of EVCI Data

Sagar Babu Mitikiri^a, Yash Tiwari^b, Vedantham Lakshmi Srininvas^c, Mayukha Pal*^d

Abstract—With the increase in the number of electric vehicles (EV), there is a need for the development of the EV charging infrastructure (EVCI) to facilitate fast charging, thereby mitigating the EV congestion at charging stations. The role of the public charging station depot is to charge the vehicle, prioritizing the achievement of the desired state of charge (SoC) value for the EV battery or charging till the departure of the EV, whichever occurs first. The integration of cyber and physical components within EVCI defines it as a cyber physical power system (CPPS), increasing its vulnerability to diverse cyber attacks. When an EV interfaces with the EVCI, mutual exchange of data takes place via various communication protocols like the Open Charge Point Protocol (OCPP), and IEC 61850. Unauthorized access to this data by intruders leads to cyber attacks, potentially resulting in consequences like energy theft, and revenue loss. These scenarios may cause the EVCI to incur higher charges than the actual energy consumed or the EV owners to remit payments that do not correspond adequately to the amount of energy they have consumed. This article proposes an EVCI architecture connected to the utility grid and uses the EVCI data to identify the anomalies or outliers present in the EV transmitted data, particularly focusing on SoC irregularities. The proposed methodology involves utilizing a ridge regression based machine learning (ML) model for predicting changes in the SoC. The adversaries have the capability of spoofing these change in SoC values, consequently making the EVCI incapable of achieving the desired task. Three distinct spoofing techniques namely, decimal shifting, incremental array spoofing, and random spoofing are implemented on the data and subsequently tested with the proposed methodology. The results show that the proposed methodology detects the anomaly accurately and also classifies the type of spoofing that causes the anomaly.

Index Terms—EVCI, SoC, fast charging, spoofing, anomaly detection, Ridge regression,

I. INTRODUCTION

The global electric vehicle (EV) market sales have exhibited robust growth reaching 2 million units sold in the first quarter of 2022, which is a 75% increase in sales compared to 2021 [1]. Due to this rising demand, the EV charging

(Corresponding author: *Mayukha Pal)

infrastructure (EVCI) needs to be developed at the same rate. To reduce the EV charging stations (EVCS) complications in penetration with the grid [2] and in charging such as charging speed, compatibility, range anxiety, grid power availability, and congestion of EVs at charging stations, etc., These charging stations are located at parking areas of corporate offices, commercial malls, and residential apartments, where there is probability for the vehicle staying in these locations, for the dwell times [3]. Due to the intermittent power available from the utility grid, charging stations are integrated with energy storage systems like battery energy storage systems (BESS), or renewable energy systems (RES) to meet it's power demand. In addition to their primary function of EV charging, these stations offer auxiliary grid-related services, including the stabilization of frequency and voltage levels, provision of reactive power support, etc., These services require bidirectional power flow that enables the flow of electricity from the grid-to-vehicle (G2V mode) and also from the vehicleto-grid (V2G mode), particularly during the instances when the demand for the loads in the grid exceeds the generation [4]. Furthermore, the charging infrastructure has exhibited consistent augmentation in charging capacity, such as ultrafast direct current (DC) [5] and extreme fast charging (XFC) [6] stations equipped with multiple modules. These high-speed charging stations could deliver up to 350 kW of rated power, enabling rapid charging cycles within 10 minutes timeframe, thereby establishing a comparable refueling experience to that of internal combustion engine vehicles (ICEVs).

As a consequence of rapid fast charging, the grid perceives the EVCI aggregated with EVs as a bulk power system load (during G2V mode) with inherent energy storage capacity such as BESS. When all EVs are simultaneously charged, the grid experiences various voltage and frequency stress, particularly during peak demand periods. These conditions necessitate load shedding actions, thereby compromising the overall reliability of the grid [7]. To sustain the balance between load and generation, the EVCI actively participates in demand response (DR), entailing the charging of the EVs during the offpeak periods, and the charged EVs along with the BESS function as sources during the peak periods [8], [9]. The integration of the EVCI with functions other than charging, such as load balancing, demand response, frequency response, etc., requires a communication network with internet-of-things (IoT) between various operators like the distribution system operator (DSO), EVCI administrators, and the EV users. These communication requirements across different elements of the

^aMitikiri Sagar Babu is a Data Science Research Intern at ABB Ability Innovation Center, Hyderabad 500084, India and also a Ph.D. Research Scholar at the Department of Electrical Engineering, IIT (ISM) Dhanbad, Dhanbad, 826004, India.

^bYash Tiwari is with ABB Ability Innovation Center, Bangalore-560048, IN, working as a Data Scientist

^cDr. Vedantham Lakshmi Srinivas is an Asst. Professor in the Department of Electrical Engineering, IIT (ISM) Dhanbad, Dhanbad, 826004, India.

^dDr. Mayukha Pal is with ABB Ability Innovation Center, Hyderabad-500084, IN, working as Global R&D Leader – Cloud & Analytics (e-mail: mayukha.pal@in.abb.com).

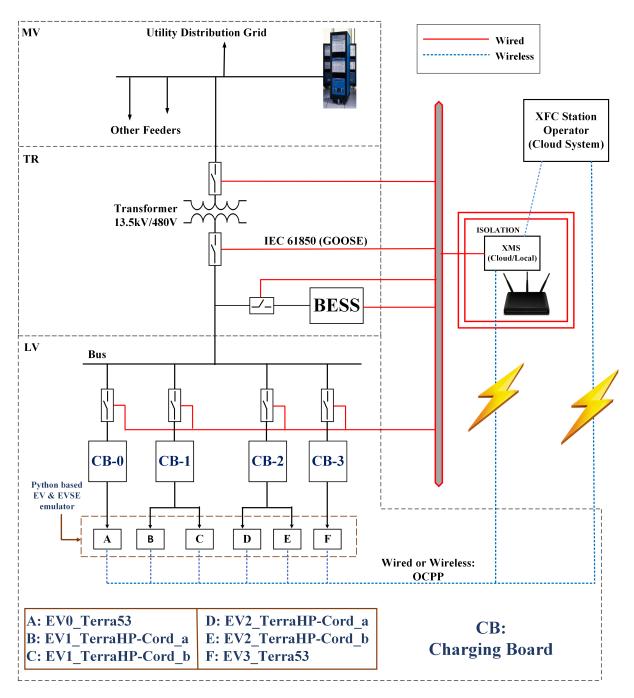


Fig. 1. Proposed architecture of EVCI

EVCI play a significant role in defining it as a cyber physical power system (CPPS) [10]. The CPPS is a physically interconnected system merged with cyber components encompassing computing, control, and communication functions allowing bidirectional flow of power and information thereby enabling the implementation of advanced smart grid technologies [11].

Due to the substantial dependence of the CPPS on the cyber systems, their communication capabilities are highly vulnerable to malicious cyber intrusions [12]. The vulnerabilities in the EVCI [13] may be internal and external vulnerabilities such as EV-X or EVCS-X interface vulnerabilities. The former consists of CAN (control area network) bus [14], tyre

pressure monitoring system (TPMS) vulnerabilities [15]. In the latter one, the X indicates various physically accessible ports, Internet Service Protocols (ISP), vendors, roadside infrastructures, and vehicular vulnerabilities. Adversaries exploit these vulnerabilities, to infiltrate the system thereby initiating a cyber attack. These attacks result in anomalies within the data, where the data exchange occurs between the diverse physical elements by way of the cyber components. The anomalies in the data of EVCI due to the cyber attacks cause service disruptions, charging delays, financial loss, safety risks, resource drain, grid instability, reputation damage, etc., [16]–[18]. So, these anomalies need to be detected and a proper

defense mechanism has to be implemented to mitigate or to prevent the effects of cyber-attacks.

II. RELATED WORK

Many researchers have addressed the cybersecurity of power grids, but the cybersecurity of EVCI has not gained much attention. The ways of cyberattacks, targeting the vital infrastructure are changing with diversified patterns in the energy distribution sector through remote communications, virtual private network (VPN) links, and corporate networks. The physical parts of the CPPS could be damaged by breaching its information and communication technology (ICT) infrastructure and obtaining more precise access to the supervisory control and data acquisition (SCADA) system which controls and monitors the components of the power grid without the need of physical attack [19]. After the cyberattacks on the Ukrainian power grid, particularly in the years 2015 and 2016 showed the necessity of cyber-physical security for substations and SCADA systems, U.S. officials from the Department of Energy (DoE), the Department of Homeland Security (DHS), the Federal Bureau of Investigation (FBI), and the North American Electric Reliability Corporation (NERC) increased their efforts realizing it, as their opportunity to examine the strategies and tactics used by the aggressor to predict the likelihood occurrence of the cyberattacks and it, s type in future [20], [21]. In both years, the attacker's motive is to successfully compromise the industrial control system (ICS) causing potential infrastructure threats, equipment failure, and power outages. Therefore, to improve the resilience, reliability, safety, and, security of power systems, it is essential to strengthen the cybersecurity of industrial automation and control systems in the power grid.

In [13], various vulnerabilities are addressed that arise due to the growth of EVs from the power grid perspective. It characterizes the vulnerabilities susceptible to exploitation, imposing potential risks to EVs and EVCI equipment, the power grid, or a combination of both. The cyber security issues associated with electric chargers and the potential consequences of cyber attacks are examined in [22]. It extensively explores the conceivable cyber attacks on utility operations, power systems, interconnected systems, and billing procedures. Additionally, it deliberates on recommendations and optimal strategies for enhancing cybersecurity within EVCI systems. The attacks on EV charging stations and their users were simulated demonstrating practical impacts on the power grid leading to service disruptions and grid failures are presented in [23]. The findings emphasize the need for mitigation strategies and uncover concerns regarding the insufficient security measures implemented in deployed charging stations.

III. CONTRIBUTIONS AND PAPER ORGANIZATION

The primary contribution of this article is detecting the anomalies in the ΔSoC values of EVs when connected to the charging stations. Since there is collinearity between the SoC and the charging current values [24], the charging current behavior is reflected in the SoC being a relative parameter thereby, ΔSoC . A ridge regression based ML model is used

for predicting the ΔSoC values. With the help of the predicted ΔSoC values, the EVCI administrator is able to detect the type of vehicle, calculate the charging time, and also in estimating the battery capacities. This work predicts the ΔSoC values from the obtained EVCI data and does not rely on the EV data. So, if there is any communication failure or incorrect estimations in the SoC of the EVs, the model won't be affected. The proposed ML model considers the current values of utility, storage system, and charging ports as the inputs of the model and ΔSoC values as the output of the model. Therefore, predicting the ΔSoC values with the current values at various ports and sensors data and the ML models helps in finding the charging behaviors of the EV, the type of the EVs arriving at the charging station, and preventing energy theft by estimating the amount of energy transferred to the EV battery with the SoC values given the battery capacity is known.

The predicted SoC ($\Delta SoC_{predicted}$) values are compared with the EV communicated SoC values (ΔSoC_{actual}) which are assumed to be spoofed later. The absolute difference is calculated between these two data points and compared with a predefined threshold iteratively. Furthermore, the overall contributions of this article are as follows:

- A grid connected EVCI system is proposed and simulated to obtain the data.
- The proposed EVCI model consists of both cyber and physical components including both wired and wireless communications by utilizing various protocols such as OCPP and IEC 61850 (GOOSE).
- A ridge regression based ML model is implemented to predict the SoC value of the EV's battery.
- This work detects the anomalies in the ΔSoC of the EV's battery using the EVCI data.
- The anomaly is detected based on the absolute differences between the communicated and predicted ΔSoC and by observing the nature of these differences continuously.
- Various types of spoofing scenarios are simulated on the testing data for producing the anomaly data.
- In addition to anomaly detection, the proposed methodology also considers the spoofing classification.
- The proposed methodology is tested and validated for different cases.

This article structure commences with an introduction to the CPPS, EVCI infrastructure types, and their basic operation in Section I. The comprehensive review of literature on the cybersecurity aspects of EVCI when interfaced with the utility grid is presented in Section II. Then Section III describes the author's contribution to the work. The subsequent Section IV outlines the modeling framework for the proposed EVCI system. Section V describes the parameters of the data acquired for both the training and testing data. It also provides the preprocessing of the data for using it as an input to the ML model. Within Section VI, the methodology and ML model employed for the prediction of ΔSoC are explained. The metrics used for evaluating the model performance are explained in Section VII. Additionally, various case studies are also provided to validate the proposed anomaly detection framework and spoofing classification. Finally, Section VIII concludes with the achievements of this study while mentioning the potential areas for future research.

IV. MODELLING OF THE SYSTEM

The modeling of the proposed EVCI system is categorized into two parts: 1. Physical system modeling (representing the physical components like the charging board (CB), utility grid, BESS, etc.,) and 2. Cyber system modeling (consisting of communications, and protocols like OCPP). Fig. 1 illustrates the comprehensive architecture of the EVCI comprising these two systems.

A. Physical System Modelling

The physical system modeling of an EVCI comprises multiple units pertaining to the power flow and energy storage. The key components in the physical system of the EVCI include the point of common coupling (PCC) linking the EVCI to the utility grid through a transformer, BESS, CBs, and charging ports. Fig. 1 illustrates the architectural components within the physical system of EVCI, responsible for storing the root mean square (RMS) values of the various measured electrical parameters such as active power (P), reactive power (Q), voltage (V), and current (I).

The proposed system utilizes a step-down type transformer of rating 13.5kV/480V placed between the PCC and the EVCI as shown in Fig. 1. The main function of this transformer is to step down the grid level medium voltage (MV) to the low voltage (LV). The BESS and CBs are connected to the PCC through this transformer. This article considers four CBs namely CB-0, CB-1, CB-2, and CB-3. To facilitate EV charging, the charging ports are connected to the CBs. The detailed representation of these charging ports and their current notations are provided in Table I. Two types of charging ports are considered in this work namely, Terra HP charger that supports from 175-350 kW fast charging and Terra 53 charger supporting up to 50 kW power [25].

TABLE I CHARGER CURRENTS REPRESENTATION

Charging board	Charging ports	Notation
CB-0	EV0_Terra53	I_{EV0}
CB-1	EV1_TerraHP-Cord_a EV1_TerraHP-Cord_b	$egin{array}{l} { m I}_{EV1a} \ { m I}_{EV1b} \end{array}$
CB-2	EV2_TerraHP-Cord_a EV2_TerraHP-Cord_b	\mathbf{I}_{EV2a} \mathbf{I}_{EV2b}
CB-3	EV3_TerraHP	I_{EV3}

Since the main aim of this work is detecting the anomalies, the modeling of EVs is done considering only two types of vehicles whose battery types are BEV 300 and BEV 150 where BEV stands for battery electric vehicle. The BEV 300 battery type EVs are charged with a Terra HP charger that can charge up to 300 kW and the BEV 150 battery type EVs are connected to Terra 53 chargers for charging the power up to 300 kW. For all the other types of EVs, the nominal charging power is considered 50 kW and connected with Terra 53 chargers.

It is crucial to note that any vehicle upon connecting to the EVCS transmits some amount of data. This data encompasses details such as charging protocol, EV type, arriving time, departure time, initial SoC, and target SoC [26]. The target SoC signifies the user-demanded SoC, representing the desired charge level the EV should attain before the scheduled departure time. The EV also continuously transmits the SoC values to the EVCI through chargers. The EVCI charges the EVs by supplying the grid power or the BESS power depending on availability.

1) BESS modeling: A simple model of the XFC station depot connected with a BESS is represented in Fig. 1. The parameters and the specifications of the BESS used in this work are provided in Table II. This paper focuses on the modeling of public EVCS, with the assumption of uncontrollable arrival rates of EVs and unpredictable power flow to the EVSE. Since the grid power is also not controllable, Therefore the only power (or current) in the EVCI is the BESS.

TABLE II
BESS PARAMETERS AND SPECIFICATIONS

Parameter	Value
Maximum charging power (kW)	500
Maximum Discharging power (kW)	500
Energy capacity (kWh)	250
Charging efficiency (%)	95
Discharging efficiency (%)	95
Maximum SoC (%)	90
Minimum SoC (%)	20
Initial SoC (%)	50
Maximum power import (kW)	1000
Maximum power export (kW)	1000

B. Cyber System Modelling

The cyber system consists of XMS (extreme fast charging managing system) operator and XMS cloud. Both the wired and wireless communications are employed in this system. The EVSE, BESS, and transformer parameter measurement circuits are communicated to the XMS cloud in a wired mode through the IEC 61850 messaging protocol, and the charging ports and the EVs communicate to the XMS operator and the XMS cloud system in a wireless mode through OCPP (Open Charge Point Protocol). Fig. 1 shows different modes of communication, connecting the various parts of the proposed system.

The IEC 61850 is a standard messaging protocol used for multicast messages such as generic object-oriented substation events (GOOSE). It recommends only message integrity and authenticity. Generally, in substation automation systems, GOOSE messages are used to carry the breaker open or close commands [27]. Another protocol used here is OCPP, which is an open source to facilitate the communication between the charging ports and the backend systems. This was proposed by the Dutch foundation ElaadNL and should be standardized and certified by OCA (Open Charge Alliance) to ensure

uniformity and cross-vendor compatibility [28]. The electrical parameters data is transmitted from the physical parts of the system through these protocols. Besides, the charging ports also transmit another parameter known as charge status (CS) which is a binary that returns 1 when the EV is charging and 0 otherwise. Both the physical and cyber systems simulations are performed in real-time on the OPAL-RT simulation platform using both MATLAB and Python environments.

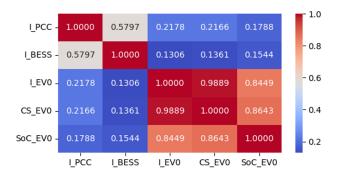


Fig. 2. Heatmap of correlation among various parameters related to charging port EV0_Terra53

V. DATA ACQUISITION AND PREPROCESSING

The data required for developing the proposed methodology is acquired and stored for both training and testing separately. The training data is collected for four days at a rate of one sample per second and the testing data is collected for one day at the same rate. The acquired data consists of variables PPC current (I_{PCC}), BESS current (I_{BESS}), charging port current (I_{EVn}), active power (P_{EVn}) and reactive power (P_{EVn}) of charging port, charge status (P_{EVn}) where P_{EVn} is the representation of the EV charging port. Since it is required for the EVCI to know the SoC of the connected EV, The SoC of the connected EV is continuously transmitted to the EVCI through OCPP as explained later. This data is also collected and stored in the variable P_{EVn} at an equivalent sampling rate.

Since the main aim of this article is detecting the anomalies in ΔSoC , it is calculated by subtracting the present instant SoC value from its previous instant. This ΔSoC calculation is performed for all the EVs connected to various charging ports and stored in variables labeled as ΔSoC_{EVn} . It is the actual or calculated SoC, termed as ΔSoC_{actual} . This data variable is the output of the proposed ML model where various types of spoofing techniques are assumed to be done in the form of a cyber attack. This study explores three distinct types of spoofing techniques: decimal shifting, incremental array spoofing, and random spoofing. Table III presents the numerical levels associated with each type of spoofing technique. These values are chosen to simulate stealthy attack scenarios [29] and implemented individually on the testing data at different instances.

Fig. 4 shows the communicated and predicted ΔSoC values where the spikes in communicated ΔSoC values signify the large variation in it that occurs due to the arrival or departure of the EVs at the respective charging port of the EVCI. Since

TABLE III
SPOOFING TECHNIQUES: VALUE RANGES AND TYPES

Noise type	Range (or) Type
Random	Uniform distribution (in range of 10^{-2})
Shifting in decimals	-0.009 to $+0.009$
Incremental array	Arithmetic progression

these data points are not considered anomalies, the spoofing techniques are simulated to the particular windows of the obtained time series data with lengths more than the length of these spike durations. The random spoofing replaces the ΔSoC_{actual} values with random values within the specified range by following a uniform distribution. Similarly, the incremental array spoofing technique also replaces a series of ΔSoC_{actual} values, with these series of values should be in arithmetic progression. The scaling and shifting in decimals type of spoofing is performing mathematical calculations to the actual output values for simulating the spoofing scenario. The actual output of the time series data is replaced with the spoofed time series output of the testing data and later compared with the output of the ML model ie., $\Delta SoC_{predicted}$ for anomaly detection.

For any ML model, the input data has to be preprocessed. Due to the spikes and high variations in ΔSoC values, the min-max normalization fails because of the irregular distribution of data. The standard scaling method is used to preprocess the data as required by the Ml model.

VI. METHODOLOGY

A. Multicollinearity

Multicollinearity is defined as the phenomenon of the existence of a strong correlation between various variables in a regression model [30]. This phenomenon arises from diverse circumstances, including the inherent physical interpretations of variables, such as metrics or measurements utilized within software defect prediction models. The interdependence among these variables could be influenced by the intrinsic nature of their meanings, contributing to the multicollinearity. Another reason for multicollinearity is the large collection of metrics. For constructing a better model we tend to collect metrics fitting to the defects as many as possible resulting in serious multicollinearity issues in datasets [31].

The multicollinearity in the acquired datasets exists between the battery current of the EV (I_{EVn}) and its estimated SoC values, where n indicates the charging port to which the EV is connected. Since coulomb counting [32] is the most adopted method, it estimates the SoC as a relative measure of charging or discharging EV current and integrates it over time and is given as follows:

$$SoC(t) = SoC(t-1) + \frac{I(t)}{Q_n} \Delta t$$
 (1)

Where SoC(t) is the SoC to be estimated at present instant, SoC(t-1) is the SoC of the battery at the previous time instant. I(t) is current at present time instant, Q_n is the capacity

of the battery pack and Δt is the discrete sampling period of the battery management systems (BMS). Fig. 2 demonstrates the correlations with the help of heat-map between the various variables showing the high collinearity between the current and SoC values.

There also exists another strong collinearity between the multiple input variables of the ML model such as I_{PCC} , I_{BESS} , and I_{EVn} (currents flowing through all the EVs). This collinearity is easily verified by Kirchoff's current law (KCL) and is given as follows:

$$I_{PCC} = I_{BESS} + \sum_{k=1}^{N} I_{CBk}$$
 (2)

Where I_{CBk} is the EVCI's CB current whose values are equal to the sum of all charging ports current connected to the particular CB. N is the total number of CBs present in the charging station.

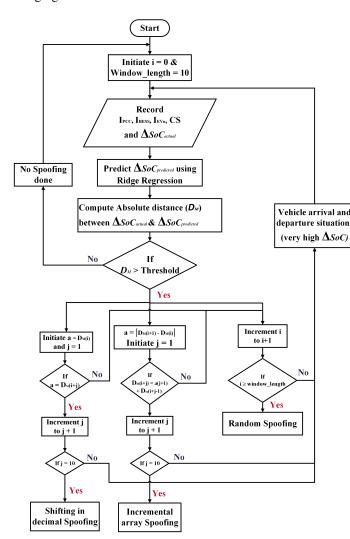


Fig. 3. Proposed Anomaly detection methodology flowchart

B. Ridge Regression

In the context of multiple linear regression applications, ridge regression is the most commonly used parameter forecasting technique particularly used to overcome the drawbacks

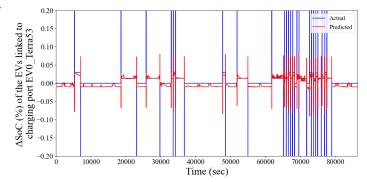


Fig. 4. Comparison of actual and predicted ΔSoC values of the EVs connected to charging port EV0_Terra53

associated with collinearity that arise frequently in other regression scenarios [33]. It helps to mitigate the negative impact of collinearity on the least squares (LS) estimator. The main feature of this ridge regression is addressing the problem of collinearity without removing any data variable. With the other regression estimators, there is a probability of coefficients becoming larger in absolute values and sometimes may also have a wrong sign. The likelihood of encountering these challenges increases the deviation in prediction vectors from orthogonality. As discussed in the previous sections, due to the collinearity present between the multiple input variables in this work, considering the multiple linear regression standard model:

$$\mathbf{y} = \mathbf{x}\beta + \varepsilon \tag{3}$$

Where $E(\varepsilon)=0$, $E(\varepsilon\varepsilon')=\sigma^2I_n$, and ${\bf x}$ is a full rank matrix of n×p dimensions. The bold letter symbols denote vectors and matrices. It is assumed that the variables are standardized to the correlation form $({\bf x}'{\bf x})$ and the response variable with each of the explanatory variables is given by the vector $\gamma\equiv {\bf x}'{\bf y}$. Let $\hat{\beta}$ be the LS estimate of the variable β and is given by

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}^{-1}\mathbf{y} \tag{4}$$

The average distance between the β and $\hat{\beta}$ directly contributes to the challenges in this standard estimation. It is achieved by the standard squared distance (L^2) between β and $\hat{\beta}$ with the following properties:

$$L^{2} = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \tag{5}$$

$$E(L^2) = \sigma^2 \operatorname{trace}(\mathbf{x}'\mathbf{x})^{-1} \tag{6}$$

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \operatorname{trace}(\mathbf{x}'\mathbf{x})^{-1} \tag{7}$$

Traditional linear regression methods, when dealing with high collinear data suffer from coefficient stability as the input matrix (x) becomes singular due to correlated independent variables. This paper adopted ridge regression to provide a realistic solution by penalizing excessive coefficients shrinking toward null values, by incorporating the L2 normalization penalty term into the loss function $(\lambda \|\beta\|_2^2)$. This penalty term makes the model coefficients smaller and more stable. The

parameters λ and α govern the trade-off between achieving a good fit of the model and restricting the coefficients to small values respectively. The final solution for the ridge regression is given as follows:

$$\hat{\beta}_{ridge} = \operatorname{argmin}_{\beta} \left\{ \left\| y - X\beta \right\|_{2}^{2} + \alpha \left\| \beta \right\|_{2}^{2} \right\} \tag{8}$$

Grid search and cross-validation (GS-CV) method is employed to rigorously evaluate a set of variables [34]. This involves assessing the model's performance across each fold of cross-validation. To ensure the model's generalization ability beyond the training data, the optimal configuration was selected based on the fold-averaged performance metric. The choice of specific parameters influences the extent of coefficient shrinkage subsequently impacting the stability and accuracy of the model. Through GS-CV, the value of the parameter achieves a good trade-off between variance reduction and minimum bias [35].

This basic strategy reduces multicollinearity by providing three critical benefits: reduced variance (ensuring stable estimates), improved bias-variance trade-off (favoring generalizability), and higher model stability (reducing noise sensitivity). As a result, ridge regression enables researchers to build robust and interpretable models even when confronted with tough collinear data, improving the quality of scientific discourse and encouraging more robust conclusions.

The performance of the six regression models for all the six EV charging ports is evaluated individually with suitable metrics. Table V depicts the various metrics such as mean squared error (MSE) and its values are also provided to evaluate the model.

Algorithm 1 Anomaly Detection

Initialize the value threshold and window length as max_iter.

Record the values I_{PCC} , I_{BESS} , I_{EVX} , CS, and ΔSoC_{actual} .

for $i = 1, 2, \dots$ max_iter do

Predict $\Delta SoC_{predicted}$ using ridge regression ML model.

Compute the absolute difference (D_M) between the values ΔSoC_{actual} and $\Delta SoC_{predicted}$.

```
if D_M \geq threshold then \mathbf{x} \leftarrow D_M(i), \ \mathbf{y} \leftarrow \mid D_M(i+1) - D_M(i) \mid. Go to Algorithm-2, Algorithm-3, and Algorithm-4. end if end for
```

The comparison between actual and predicted ΔSoC values of the EVs connected to the charging port EV0_Terra53 are shown in Fig. 4. Since there are large deviations in the actual ΔSoC values, these scenarios occur during the arrival and departure of the EVs to the EVCI as depicted in Fig. 4, it is due to the reason that during the EV arrival to the EVCI, SoC changes abruptly from 0 to a greater value and also during departure, the SoC drops from a higher value to 0. So, the proposed methodology considers a time window of different lengths for the different charging ports. As a result, whenever the D_M exceeds the threshold value, the methodology starts

the iteration count and checks for the larger thresholds in the consecutive data samples throughout the window. If the D_M exceeds the threshold throughout the window then the methodology detects it as an anomaly. The Algorithm 1 provides the pseudocode for the anomaly detection.

Algorithm 2 Shifting in Decimals Spoofing

```
1: a \leftarrow D_M(i)

2: for j = 1, 2, \dotsmax_iter do

3: if a = D_M(i) then

4: j \leftarrow j + 1

5: if j = \max_iiter then

6: Display Shifting in decimals Spoofing detected

7: end if

8: end if

9: end for
```

Algorithm 3 Incremental Spoofing

```
1: a \leftarrow \mid D_M(i+1) - D_M(i) \mid

2: for j = 1, 2, \dotsmax_iter do

3: if D_M(i+j) = (a*j) + D_M(i+j+1) then

4: j \leftarrow J+1

5: if j = \max_iiter then

6: Display Incremental array spoofing detected

7: end if

8: end if

9: end for
```

In addition to anomaly detection, the proposed methodology also categorizes the type of spoofing that causes anomaly in the target data i.e., ΔSoC . Since this work considers the three types of spoofing techniques namely, shifting in decimals, incremental array spoofing, and random spoofing, the proposed framework classifies in parallel to the anomaly detection. The Algorithm 2 describes the procedure for detecting the shifting in decimals type of spoofing which records the absolute difference once it reaches the threshold and compares it with the other succeeding values for the entire window duration if all the values are equal then the resulting spoofing is of shifting in decimals type.

The incremental array spoofing is detected by observing the series of absolute differences between the predicted and communicated ΔSoC for the window if the initial distance reaches the threshold. If any progression is observed the proposed methodology returns the type of progression present in the data. The flow diagram for this type of spoofing detection is provided in Algorithm 3. If the spoofing does not belong to the any of aforementioned categories, then it is termed as random spoofing. It compares the absolute differences with the threshold iteratively only if the conditions of the shifting in decimals and incremental spoofing aren't achieved as explained in Algorithm 4. The overall framework of the proposed methodology combining all the pseudocodes is depicted in Fig. 3. Furthermore, the step-wise procedure of the proposed methodology is as follows:

1 Initialize the values iterative count (i), window length, and threshold.

- 2 Predict the output of the model ($\Delta SoC_{predicted}$).
- 2 Compute $D_M = abs(\Delta SoC_{predicted} \Delta SoC_{actual})$.
- 4 if $D_M \leq$ threshold go to step 2. else, go to next step.
- $5 x = D_M(i), y = |D_M(i+1) D_M(i)| \text{ and } j = 1.$
- 6 If $D_M(i+j) = a$, increment j to j+1 else, go to step 8.
- 7 If j = 10, display shifting in decimal spoofing detected and break.
- 8 If $y = |D_M(i+1) D_M(i)|$, increment j to j+1 else, go to step 10.
- 9 If j = 10, display Incremental array spoofing detection and break.
- 10 If i < 10, increment i to i+1 else, display random spoofing detected.

Algorithm 4 Random Spoofing

- 1: **if** $D_M(i) \ge threshold$ **then**
- 2: $i \leftarrow i + 1$
- 3: **if** i = max_iter **then**
- 4: Display Random type Spoofing detected
- 5: end if
- 6: end if

VII. RESULTS & CASE STUDY

Various ML and neural network (NN) models are tested for the acquired data by preprocessing with standard scaling. The parameters and the performance metric values used in this model are shown in Table V. The MSE values of the different model shows that the ridge regression model is the best fit for predicting ΔSoC values using the EVCI data. The ride regression model was implemented using scikitlearn in Python environment, and robust model assessment was ensured by conducting the grid search with k-fold crossvalidation strategy with k = 20. The optimal hyperparameter α is determined to be 10.05. This ridge regression model is used for predicting the ΔSoC , and the performance of the model is also evaluated by metrics such as mean squared error (MSE). Table V depicts the values of the MSE metric of all the ML models used for each charging port. The testing data is used to validate the proposed methodology by spoofing it for various cases are shown with the obtained results.

TABLE IV PERFORMANCE METRICS COMPARISON OF VARIOUS MODELS FOR PREDICTING ΔSoC

Model	Parameters	MSE
Linear regression	fit_intercept = false	1.771121117
Multi-layer perceptron	Adam optimizer, Dropout = 0.2	1.77119094
Support vector regression (SVR)	C = 1.5, epsilon = 0.18	1.77747376
Random Forest (RF)	number of estimators = 230	2.01821433
Ridge regression	alpha = 10	0.000194

TABLE V
PROPOSED REGRESSION MODEL PERFORMANCE ACROSS VARIOUS
CHARGING PORTS

charging Port	MSE
EV0_Terra53	0.000194
EV1_TerraHP-Cord_a	0.000217
EV1_TerraHP-Cord_b	0.000180
EV2_TerraHP-Cord_a	0.000324
EV2_TerraHP-Cord_b	0.000129
EV3_TerraHP	0.000356

The different spoofing types considered in this work are shifting by decimals, incremental array shifting, and random spoofing. These are implemented on the charging port EV0_Terra53 and the proposed anomaly detection methodology is tested for all types of spoofing. since the results are shown for only one charging port the proposed framework, could be scaled to the other remaining charging ports also. The results obtained for detecting the anomaly caused by the different spoofing types on the charging port EV0_Terra53 are further discussed as different case studies.

A. Case Study: Sifting in decimals

This case consists of shifting the actual or communicated ΔSoC values of the connected EVs. Since this work considers shifting the decimals for a smaller window size, it does not have any impact on the EVCI. However, shifting the communicated ΔSoC values continuously for longer durations may affect the EVCI system and also the EV battery. A few sample windows of size 10 samples each are considered at random instances on the testing data for spoofing. Fig. 5(a) shows the actual communicated ΔSoC (spoofed) and the predicted ΔSoC values along with spoofed data points marked distinctly, for the charging port EV0_Terra53 of the testing data. The shifting of ΔSoC values is done by adding a constant number to the communicated ΔSoC values. If this spoofing increases the actual value then EVCI assumes that EV is charged at higher speeds and the charging process is commuted without achieving the desired task. In cases of spoofing with lower values, the charging process is not commuted even after achieving the desired task. hence, these issues are prevented by detecting this type of spoofing. A threshold is predefined for this anomaly detection based on the type of EV connected to the respective charging port. Fig. 5(b) and Fig. 5(c) depicts the anomaly detection for this type of spoofing for different time windows at time instances 190 and 81908 respectively along with the threshold where the spoofed values are the communicated ΔSoC values. The results show that the proposed anomaly detection methodology works fine and it also classifies this shifting type spoofing with an accuracy of 99.31%.

B. Case Study: Incremental array shifting

The incremental shifting of the ΔSoC consists of the shifting of the communicated ΔSoC values to a significant

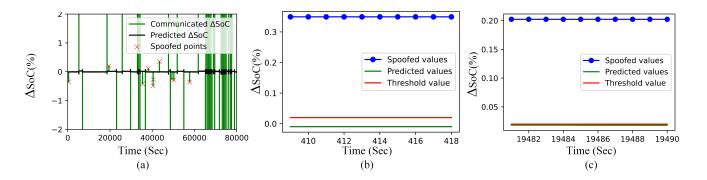


Fig. 5. Shifting in decimals case study indicating (a) communicated and predicted ΔSoC values highlighting the spoofed points, anomaly detection for the spoofed windows for the instances starting with (b) 409 and (c) 19481 each window of size 10 samples

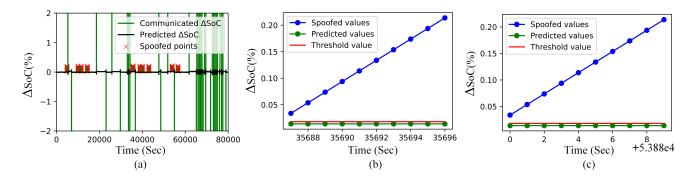


Fig. 6. Incremental array shifting case study indicating (a) communicated and predicted ΔSoC values highlighting the spoofed points, anomaly detection for the spoofed windows for the instances starting with (b) 35687 and (c) 53880 each window of size 10 samples

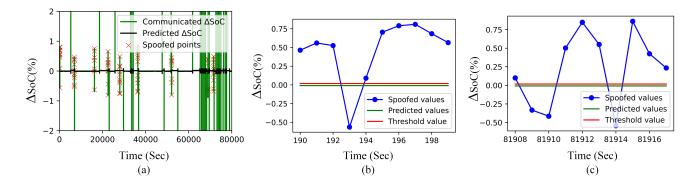


Fig. 7. Random array spoofing case study indicating (a) communicated and predicted ΔSoC values highlighting the spoofed points, anomaly detection for the spoofed windows for the instances starting with (b) 190 and (c) 81908 each window of size 10 samples

amount. The earlier shifting in decimal consists of adding or replacing the actual values with a constant value, whereas this shifting type adds or replaces the actual values with a series of progression values. This type of spoofing plays a crucial role during the EVs arrival and departure duration where the ΔSoC values are in higher magnitudes. The communicated (spoofed) values and the predicted values of the ΔSoC are shown in Fig. 6(a) with the spoofed data points marked distinctly for the EVs arriving at the charging port EV0_Terra53 of the testing data. This spoofing makes the EVCI assume that the EV is charging at faster or slower rates than desired which results in disturbing the charging and discharging rates (C-rates) of the EVs. To mitigate these issues, this type of spoofing

is detected by employing a predefined threshold. Since this progression based shifting may be incremental or decremental, the proposed methodology is verified for the incremental array shifting but it could be applied to both types of spoofing. This work considers shifting the ΔSoC values by adding a series of the arithmetic progression to the existing values thus making the SoC increase at a constant faster rate. Fig. 6(b) and Fig. 6(c) show the anomaly detection window for this type of spoofing considered at a faster rate in the testing data for the different instances. These results indicate that the proposed methodology has 90.84% accuracy in detecting the anomalies attributed to spoofing, specifically related to the array shifting of the SoC values with a sequence of arithmetic progression

values.

C. Case Study: Random Spoofing

The type of spoofing techniques that do not belong to any of the aforementioned categories are classified as random spoofing. Random spoofing consists of replacing the actual ΔSoC values with random numbers. A few sample windows of size 10 instances each are spoofed and these time windows are chosen randomly. The communicated (spoofed) and the predicted ΔSoC values are shown in Fig. 7(a) along with the spoofed points marked distinctly, for the charging port EV0_Terra53 of the testing data. The effect of random spoofing is irregular variations in ΔSoC values resulting in the EVCI not achieving the desired tasks. This work considers replacing the actual ΔSoC value with a series of random numbers within the specified range as presented in Table III. Fig. 7(b) and Fig. 7(c) show the anomaly detection where the spoofing is of random type. The results show that the proposed methodology has 93% accuracy in detecting the random spoofing and classifying it.

VIII. CONCLUSIONS AND FUTURE WORK

An EVCI consisting of four CBs and six charging ports is simulated as shown in Fig. 1. The communications of the EVCI system are also simulated with suitable protocols like OCPP and GOOSE. The required data is obtained from the simulation and segregated into training and testing data as needed. A data-driven approach is used to predict the ΔSoC , which adopts a ridge regression based ML model due to the multicollinearity between the data variables. This work assumes that anomalies are caused due to spoofing the communicated ΔSoC values. Absolute differences between the predicted ΔSoC and the communicated ΔSoC are computed and these differences are compared with the predefined threshold. The proposed methodology iteratively compares the absolute differences with the threshold to detect the anomalies. This work differentiates the anomalies with the EV's arrival and departure scenarios.

Three types of spoofing techniques are implemented on the testing data and the proposed algorithm is validated. various case studies are evaluated to validate the proposed algorithm for the spoofing classification. The spoofing techniques are classified by observing the patterns in the calculated absolute differences. Spoofing ΔSoC values results in the irregular and improper charging of the EVs leading to energy theft, and financial loss to the EVCI administrators and the EV users. The possible future works are as follows:

- Load forecasting of the EVCI as the manipulation of the SoC leads to performance degradation of batteries and causes irregular loading behavior on EVCI.
- Study of the grid instabilities due to large scale cyber attacks targeting the multiple EV's SoC impacting the power grid, leading to fluctuations and potential instabilities in the grid.
- Congestion management and handling the unplanned charging sessions that may arise due to false or manipulated SoC values.

• Estimating the revenue loss incurred by the energy theft due to the cyber attacks on the EVCI.

REFERENCES

- "Iea global ev outlook 2022." https://www.iea.org/reports/ global-ev-outlook-2022.
- [2] M. A. Ravindran, K. Nallathambi, P. Vishnuram, R. S. Rathore, M. Bajaj, I. Rida, and A. Alkhayyat, "A novel technological review on fast charging infrastructure for electrical vehicles: Challenges, solutions, and future research directions," *Alexandria Engineering Journal*, vol. 82, pp. 260–290, 2023.
- [3] L. Pan, E. Yao, Y. Yang, and R. Zhang, "A location model for electric vehicle (ev) public charging stations based on drivers' existing activities," *Sustainable cities and society*, vol. 59, p. 102192, 2020.
- [4] D. C. Erb, O. C. Onar, and A. Khaligh, "Bi-directional charging topologies for plug-in hybrid electric vehicles," in 2010 Twenty-Fifth Annual IEEE Applied Power Electronics Conference and Exposition (APEC). IEEE, 2010, pp. 2066–2072.
- [5] D. Aggeler, F. Canales, H. Zelaya-De La Parra, A. Coccia, N. Butcher, and O. Apeldoorn, "Ultra-fast dc-charge infrastructures for ev-mobility and future smart grids," in 2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe). IEEE, 2010, pp. 1–8.
- [6] H. Tu, H. Feng, S. Srdic, and S. Lukic, "Extreme fast charging of electric vehicles: A technology overview," *IEEE Transactions on Transportation Electrification*, vol. 5, no. 4, pp. 861–878, 2019.
 [7] X. Chen, H. Wang, F. Wu, Y. Wu, M. C. González, and J. Zhang,
- [7] X. Chen, H. Wang, F. Wu, Y. Wu, M. C. González, and J. Zhang, "Multimicrogrid load balancing through ev charging networks," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5019–5026, 2021.
- [8] H. Barati and F. Ashir, "Managing and minimizing cost of energy in virtual power plants in the presence of plug-in hybrid electric vehicles considering demand response program," *Journal of Electrical Engineering & Technology*, vol. 13, no. 2, pp. 568–579, 2018.
- [9] P. Chakraborty, M. Pal et al., "Planning of fast charging infrastructure for electric vehicles in a distribution system and prediction of dynamic price," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109502, 2024.
- [10] R. V. Yohanandhan, R. M. Elavarasan, P. Manoharan, and L. Mihet-Popa, "Cyber-physical power system (cpps): A review on modeling, simulation, and analysis with cyber security applications," *IEEE Access*, vol. 8, pp. 151019–151064, 2020.
- [11] M. Abdelmalak, V. Venkataramanan, and R. Macwan, "A survey of cyber-physical power system modeling methods for future energy systems," *IEEE Access*, 2022.
- [12] K.-D. Lu and Z.-G. Wu, "Resilient event-triggered load frequency control for cyber-physical power systems under dos attacks," *IEEE Transactions on Power Systems*, 2022.
- [13] S. Acharya, Y. Dvorkin, H. Pandžić, and R. Karri, "Cybersecurity of smart electric vehicle charging: A power grid perspective," *IEEE Access*, vol. 8, pp. 214434–214453, 2020.
- [14] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham *et al.*, "Experimental security analysis of a modern automobile," in 2010 IEEE symposium on security and privacy. IEEE, 2010, pp. 447–462.
- [15] I. Rouf, R. Miller, H. Mustafa, T. Taylor, S. Oh, W. Xu, M. Gruteser, W. Trappe, and I. Seskar, "Security and privacy vulnerabilities of {In-Car} wireless networks: A tire pressure monitoring system case study," in 19th USENIX Security Symposium (USENIX Security 10), 2010.
- [16] S. Acharya, Y. Dvorkin, and R. Karri, "Public plug-in electric vehicles+ grid data: Is a new cyberattack vector viable?" *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5099–5113, 2020.
- [17] E. U. Soykan, M. Bagriyanik, and G. Soykan, "Disrupting the power grid via ev charging: The impact of the sms phishing attacks," *Sustainable Energy, Grids and Networks*, vol. 26, p. 100477, 2021.
- [18] E. Gumrukcu, A. Arsalan, G. Muriithi, C. Joglekar, A. Aboulebdeh, M. A. Zehir, B. Papari, and A. Monti, "Impact of cyber-attacks on ev charging coordination: The case of single point of failure," in 2022 4th Global Power, Energy and Communication Conference (GPECOM). IEEE, 2022, pp. 506–511.
- [19] C.-W. Ten, C.-C. Liu, and G. Manimaran, "Vulnerability assessment of cybersecurity for scada systems," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1836–1846, 2008.
- [20] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, no. 1-29, p. 3, 2016.

- [21] G. Simons, Y. Danyk, and T. Maliarchuk, "Hybrid war and cyberattacks: creating legal and operational dilemmas," Global Change, Peace & Security, vol. 32, no. 3, pp. 337-342, 2020.
- [22] J. Johnson, T. Berg, B. Anderson, and B. Wright, "Review of electric vehicle charger cybersecurity vulnerabilities, potential impacts, and defenses," Energies, vol. 15, no. 11, p. 3931, 2022.
- [23] T. Nasr, S. Torabi, E. Bou-Harb, C. Fachkha, and C. Assi, "Power jacking your station: In-depth security analysis of electric vehicle charging station management systems," Computers & Security, vol. 112, p. 102511, 2022.
- [24] J. Lee and J. Won, "Enhanced coulomb counting method for soc and soh estimation based on coulombic efficiency," IEEE Access, vol. 11, pp. 15 449–15 459, 2023. [25] "Abb-asea brown boveri. ev charging infrastructure [online]." https://
- new.abb.com/ev-charging.
- [26] Q. Chen, F. Wang, B.-M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. Catalão, "Dynamic price vector formation model-based automatic demand response strategy for pv-assisted ev charging stations," IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 2903-2915, 2017.
- [27] S. S. Hussain, S. M. Farooq, and T. S. Ustun, "A method for achieving confidentiality and integrity in iec 61850 goose messages," IEEE transactions on Power Delivery, vol. 35, no. 5, pp. 2565-2567, 2020.
- [28] Z. Garofalaki, D. Kosmanos, S. Moschoyiannis, D. Kallergis, and C. Douligeris, "Electric vehicle charging: A survey on the security issues and challenges of the open charge point protocol (ocpp)," IEEE Communications Surveys & Tutorials, 2022.
- [29] T. Sui, Y. Mo, D. Marelli, X. Sun, and M. Fu, "The vulnerability of cyber-physical system under stealthy attacks," IEEE Transactions on Automatic Control, vol. 66, no. 2, pp. 637-650, 2020.
- [30] H. Altland, "Regression analysis: Statistical modeling of a response variable," Technometrics, vol. 41, pp. 367-368, 03 2012.
- [31] X. Yang and W. Wen, "Ridge and lasso regression models for crossversion defect prediction," IEEE Transactions on Reliability, vol. 67, no. 3, pp. 885-896, 2018.
- [32] K. S. Ng, C.-S. Moo, Y.-P. Chen, and Y.-C. Hsieh, "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," Applied energy, vol. 86, no. 9, pp. 1506-1511,
- [33] G. C. McDonald, "Ridge regression," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 1, no. 1, pp. 93-100, 2009.
- [34] "Grid search cross-validation [online]." https://scikit-learn.org/stable/ modules/generated/sklearn.model_selection.GridSearchCV.html.
- [35] M. Adnan, A. A. S. Alarood, M. I. Uddin, and I. ur Rehman, "Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models," PeerJ Computer Science, vol. 8, p. e803, 2022.