RQGPR: Rational Quadratic Gaussian Process Regression for Attack Detection in the SCADA Networks

Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma [†], Jae Min Lee, Dong-Seong Kim
IT Convergence Engineering, [†] ICT Convergence Research Center,
Kumoh National Institute of Technology Gumi, South Korea
loveahakonye, cosmas.ifeanyi, ljmpaul, dskim@kumoh.ac.kr

2022 27th Asia Pacific Conference on Communications (APCC) | 978-1-6654-9927-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/APCC55198.2022.9943564

***Abstract*—The constant development and deployment of the supervisory control and data acquisition (SCADA) in the industrial internet of things (IIoT) have enabled vast communication leading to the generation of large volumes of sensor data. This phenomenon has increased SCADA's susceptibility to vulnerability and attacks which calls for attack detection mechanisms. Existing systems only aim at detection accuracy without considering the effect of false alarm rates in large sensor data. To resolve this issue, we propose a Rational Quadratic Gaussian Process Regression (RQGPR) for the effective reduction of false alarm rate and improved prediction precision. In this algorithm, a Gaussian process regression model is trained with recourse to kernel functions to precisely predict attacks and reduce false alarms. The RQGPR outperforms all other kernels in the reduction of false alarm rates. Through simulations, we show that the proposed model reduces the false alarm rate up to 71.73% higher than other kernels. This result was validated by evaluating the CIRA-CIC-DoHBrw-2020 datasets, which also had a reduction rate of 67.61%. In addition, it also showed superior performance when compared with other state-of-the-art models.**

***Index Terms*—Intrusion/attack Detection, Gaussian Kernels, Machine Learning, Network Communication, Rational Quadratic, Reliability, SCADA, Smart Factory Communication**

I. Introduction

The supervisory control and data acquisition (SCADA) systems are control systems in the industrial environment, such as electrical power grids, oil and gas pipelines, water distribution, e.t.c. This technology allow automated control and remote monitoring of industrial processes [1]. Components that enable this functions include computer workstations, human-machine interface (HMI), programmable logic controllers (PLCs), sensors, and actuators [2]. Initially, these systems were standalone with dedicated networks. However, due to the widespread adoption of remote management, the internet is now used for SCADA system communication [3], [4]. It exposed SCADA systems to cyberspace, making them vulnerable to cyber-attacks. Fig. 1 depicts the SCADA equipment communications network.

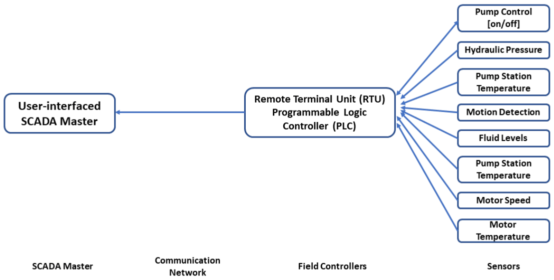

Fig. 1. Diagram showing the components of the SCADA

The SCADA network communication involves diverse elements that necessitate safeguarding against vulnerability and attacks. These incorporate the control relays, sensors, remote terminal units, master units, and the entire network [2], as depicted in Fig. 2. Security is crucial in the development of a SCADA network. Most SCADA systems have shown to be deficient in this aspect, leaving systems vulnerable to intrusion and attacks [3]. Security should be capable of limiting intrusion/attack to systems and network devices, requiring at minimum approved access to vital structures. Furthermore, a cohesive security framework must meet the authorization, accountability, dependability, and authentication requirements. It is also crucial to consider the security provided by a sturdy framework like attack detection systems.

Considering the improved user experience in the industrial internet of things (IIoT), the application of SCADA systems in the Smart factory has accelerated [5], leading to a vast generation of heterogeneous sensor data. This development has also exposed the SCADA network to vulnerability and attacks following the unrelenting efforts of intruders. Machine learning (ML) approaches have found use in SCADA IDS. Most of these approaches aim to maximize detection accuracy with little attention to false alarm rates. The Gaussian regression, Gaussian ratios, and Gaussian process (GP) have enabled the control of intrusion and vulnerability problems [6], [7]. Recent studies centered on probability density functions, covariance functions, predicting target data rate [8], and significance of common complicated Gaussian ratios viz mathematical intricacy [9]. Regardless of the growing research interests in the SCADA vulnerability and IDS, there is still minor awareness in studies targeting Gaussian kernels to minimize false alarm

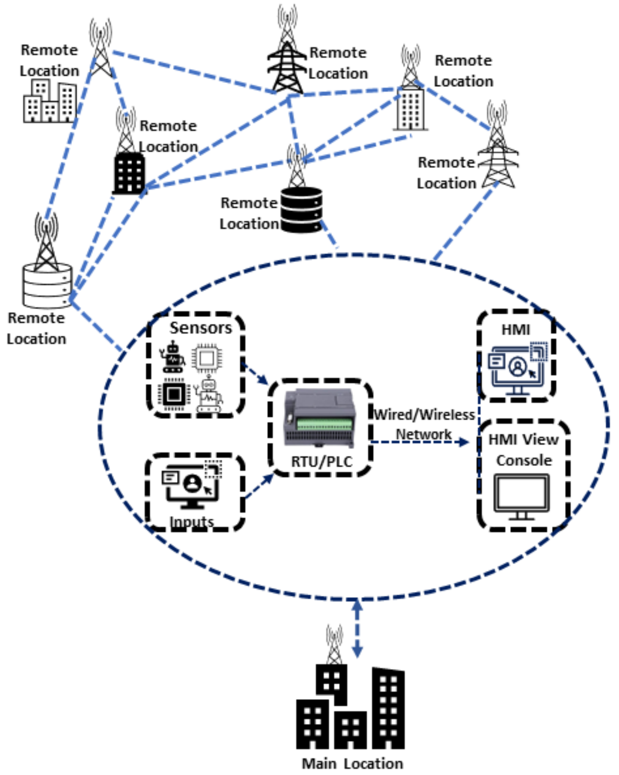

Fig. 2. Diagram showing a SCADA Network Communication Architecture

rates with improved prediction accuracy.

The problems of high false alarm rates and poor model performance have intensified the interest in GP research with recourse to kernel choice [10], [11] as GP kernels are concerned with several properties relevant to model fitting. Authors [12] proposed a boosted Gaussian fusion estimation scheme to determine the constraints of the function of the video monitoring system that requires improved speed for effective framework design. According to [11], they proposed a framework of a Gaussian neural network for online evaluation and restructuring of cyber-attacks launched on a networked system. These current studies did not accentuate the option of kernels. Hence, Gaussian kernels lack attention as being vital to the control of false alarm rates in SCADA network communication with large heterogeneous data. This study asserts that the choice of kernel is crucial in employing GP in attack detection since it influences the model performance, particularly prediction accuracy and false alarm reduction.

Due to limited studies in SCADA security problems with kernel choice, this study focus on the crucial function of kernels and their influence on the reduction of false alarm rates and improved prediction accuracy. In the context of cybersecurity data, GPR employs the layered Gaussian distribution for prediction in dynamic data that is difficult to generate using multivariate or linear regression [6]. GPR techniques typically seek to use what are known as covariance (kernels) activities to tailor a GP to specific traits. The various kernel is concerned with diverse attributes applicable to model fitting. In a scenario of implementing kernel function, there is the probability of arithmetical summation of kernel functions. The option of a suitable kernel function in a modeling process based on data features is a decisive fundamental to enable the control of the misleading effect of false alarm rates in large IIoT data. As well as improving the overall model efficiency.

This paper has the following goals:

1) To deduce and propose an efficient GPR kernel in terms of the mean absolute error (MAE) value for reducing false alarm rates.
2) Using ML (MATLAB toolbox) to determine the most appropriate Gaussian process kernels for attack detection in SCADA network.
3) As a result of its resiliency, choosing kernel is a challenge in GPR design. The constraint structure of the kernel must be chosen with skill. This study strives to aid in resolving this obstacle by putting forward the rational quadratic kernel.

The rest of the study is organized thus: Section II summarizes existing work on attack detection and alternative GP techniques, as well as identifying research gaps. In Section III, problem formulation was explored, which included a brief description of GPR and the role of kernels. The performance evaluation is described in Section IV with the efficacy of the various kernels and a comparison of the response graph and the effectiveness measure of the various kernels, followed by the conclusion in Section V.

II. Related Works

A. SCADA Attack Detection applying Gaussian Process Regression

Several studies on artificial intelligence have promoted varying applications; and improved achievements in IDS development. Substantiated by the implementation of ML/DL approaches [13], requiring focus in the target domain [14], [15]. Additionally, GPR has also found utilization for control in evaluating random linear arrays in IoT network intrusion [16]. GPR will enable efficient attack prediction in the SCADA network. Amidst the application of deep neural network for IDS is its application in the classification of anomalies based on the features of network traffic and nature of elements of the network [15].

A couple of studies have implemented the GP in attack detection; a study considering a stochastic kernel and utilizing the kernel variability to avoid assumptions in predictions, authors [17] proposed a variational Bayesian kernel selection technique for sparse GPR (SGPR) models. A study by [18] presented a hybrid model that blends Gaussian distribution and polynomial regression; the model was for detecting and visualizing anomalous activity in electricity consumption. Another study on the attack pattern prediction approach built on bag representation in time series using the Gaussian mixture model is by [19]. The authors in [20] considered a series of major disturbing technologies for the beyond-5-G ($B5G$) in large industrial networks. It claimed that these technologies would lead to an inventive switch in the design of $B5G$ base stations. The authors opined that to exploit the advantage of this invention in the future, a study in optimizing constrained resources and security is needed, which is where GP kernels find use.

Authors in [21] utilized four (4) GPR frameworks to demonstrate the efficiency of GPR in wireless sensors attack detection without consideration for computation cost. Specifically, the squared exponential kernel was in the three GPR models. However, the study did not establish the impact of other kernels as indicated in [7]. Authors [13] in a recent study attempted GPR kernel selection with a focus on IoT. Regardless of the contributions of these authors. There is little awareness of the use of GPR and the choice of kernels in the SCADA network communication attack detection. Thus, this study is a comparable research focused on the option of GPR kernels and their performance utilizing IIoT datasets. However, leveraging cyber-security datasets, this study investigates the impact of GPR kernels and their efficiency in high-dimensional spaces (datasets with higher features). In addition to the future direction of authors [13], which bothers on the doggedness of kernels across state-of-the-art datasets.

III. Methodology

The schematic diagram of the study process is in Fig. 3. Several GPR kernels were explored in this study for an optimal choice in SCADA network communication attack detection, leveraging the industrial control system (ICS-SCADA) cyber-security datasets [22], for an efficient kernel candidate. The dataset created from the power grid testbed consists of computations associated with disturbance, normal, control, and cyber-attack exploits obtained during electrical transmission. This dataset contains 128 features and two (2) responses (normal and attack traffic). Data preprocessing was to achieve good quality data before model training. It entails cleansing, sample selecting, standardization, one hot encoding, and transformation.

Of the 128 features of the ICS-SCADA dataset, there are negatively correlated features, zeros, and nan. These features need to be eliminated. The principal component analysis (PCA) for dimensionality reduction was adopted to achieve this elimination, reducing it from 128 to 74. The resultant data was divided into a ratio of 70% (train) and 30% (test), and 20% from the train set for validation. Models degrade with high-dimensional data. Hence to build a simple and more comprehensible model and improve performance, feature selection is vital [23].

Furthermore, the CIRA-CIC-DoHBrw-2020 cyber-security dataset [24], [25] was used to test and validate the reliability and applicability of the proposed model in a different network scenario. However, unlike the ICS-SCADA, the CIRA-CIC-DoHBrw-2020 dataset does not contain redundant features like negatively uncorrelated features, zeros, and nan, and needless to implement PCA. Finally, the resultant data was for attack detection.

A. DNS Traffic Dataset for Validating Model Reliability

The CIRA-CIC-DoHBrw-2020 is a recent cyber-security dataset with features as listed in [24], [25]. The choice of this dataset stems from its high volume, which is one of the limitations that affect the efficiency of GPR performance. The cyber-security data stream contains 226406 observations, 28 predictors, and two (2) responses as malicious and benign scenarios. Some predictors include the number of flow bytes sent, packet length, packet time, and rate of flow bytes sent.

B. Gaussian Process Regression Model

The ML design mechanism in this study is the GPR prediction employing MATLAB training toolbox. The GPR implements structural design and precisely as an output model training function. It is for finding models within the data [26]. The GPR outcome a is premised as a constitute y with data d as shown in equation (1):

$$a = y(d) + \varepsilon. \tag{1}$$

where ε denotes the level of noise expression as well as a simulation of the observed volatility. The "salient" expression $y(d)$ is assigned to a Gaussian Process (GP) and represented as (2):

$$y(d) \sim [GP(M(d), \lambda(d, d'))]. \tag{2}$$

A GP is a multi-role allotment that is represented by a covariance function and an average. The mean function $M(d)$ observes the supposed operation standard at data q in equation (3):

$$M(d) = \epsilon[y(d)]. \tag{3}$$

This is the estimated mean of all functions in the allocation at data d. In most cases, the previous average function is put into equation (4)

$$M(d) = 0. \tag{4}$$

as a means to avert exorbitant rearward calculations and only carry out conclusions through the covariance function, fixing the previous average outcome to zero is attained by deducting the preceding average across entire allotments. The covariance function k(d, d') patterns the reliance amongst the function codes at diverse data points d and d' as shown in equation (5):

$$\lambda(d, d') = \sum[(y(d) - M(d))(y(d') - (Md'))]. \tag{5}$$

The kernel of the Gaussian process is the function k [27].

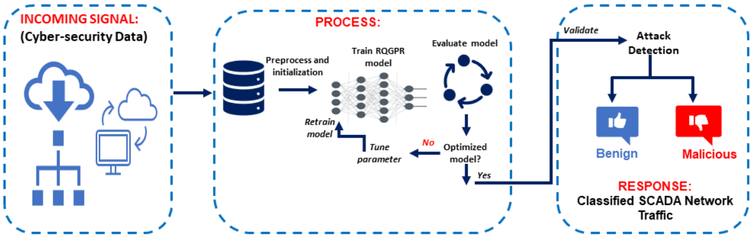

Fig. 3. The Cyber-security dataset input flow diagram, GPR modeling and application, and classification outcome in the form of Benign and Malicious

The RQGPR is an endless sum of radial basis function kernels with varying characteristic length scales. A length gives it ($l > 0$) and mixture ($\alpha > 0$) scale parameters, which allows the modeling of data at varying multiple scales [27]. The preference for the RQGPR is due to its ability to model extensive data varying at multiple scales, distinguishing it from other compared kernels. This function is given by:

$$k(a_i, a_j) = \left(1 + \frac{d(a_i, a_j)^2}{2\alpha l^2}\right)^{-\alpha}. \qquad (6)$$

where l is the length scale, α is the scale mixture parameter and $d(a_i, a_j)$ is the Euclidean distance sum of two functions (input and output).

The learning configurations of the benchmark ICS-SCADA dataset is presented in Table I. For a comprehensive discourse on GPR and application see [27]

TABLE I
DATA LEARNING CONFIGURATION WITH THE ICS-SCADA DATASET

Configuration	Input
Observations	4615 samples
No of Features Before PCA	128
No of Features After PCA	74
Response	2
Model type	Quadratic GPR
Result Presentation type	Response plot
Prediction speed	11000 obs/sec
Training time	167 sec
Kernel Sigma	Automatic
Optimize numeric parameters	True
Kernel Scale	Automatic
Use isotropic kernel	True
Standardize	True
Basis Function	Linear
Kernel Function	Rational Quadratic
Sigma	Automatic
K-fold Validation	5

IV. EXPERIMENTAL RESULT DISCUSSION

A. Investigating the Performance of Proposed Kernel

Performance evaluation of the proposed RQGPR was demonstrated on two datasets for the SCADA network attack classification using the benchmark ICS-SCADA cyber-security datasets [22], and the CIRA-CIC-DoHBrw-2020 dataset [25] respectively. Repeating the performance evaluation on the CIRA-CIC-DoHBrw-2020 is important as the dataset contains recent attack types from DNS tunnelling. The consistency in the superior performance of the proposed model further lends credibility and thus reliable. The CIRA-CIC-DoHBrw-2020 dataset parameters is in Table II.

TABLE II
PROPOSED MODEL KERNEL VALIDATION PARAMETER USING THE CIRA-CIC- DOHBRW-2020 DATASET

Parameter	Remark
Observations	226406 samples
Predictors	12
Response	2
Model type	Quadratic GPR
Result Presentation type	Response plot
Prediction speed	260 obs/sec
Training time	24360 sec
Kernel Sigma	Automatic
Optimize numeric parameters	True
Kernel Scale	Automatic
Use isotropic kernel	True
Standardize	True
Basis Function	Linear
Kernel Function	Rational Quadratic
Sigma	Automatic
K-fold Validation	5

B. Efficiency Evaluation

The performance analysis guides in this investigation are the root means squared error (RMSE), MAE, and R-Squared (R^2). It is represented as the variation in the true outcome and trained outcome and derived as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}[X_i - P_i]^2. \qquad (7)$$

where X_i denotes the exact value and P_i represents expected value. MAE determines the degree of proximity between the true and forecast values. The closer the true and anticipated values are, the lower the MAE value. MAE is a common approach to determining a model's flaw or efficiency in predicting measurable facts. There are multiple kernels in GPR, including Matern 5/2 (MGPR), rational quadratic (RQGPR), exponential (EGPR), and squared exponential (SEGPR) [13]. The viability of the kernels was verified by applying the MAE. It is a method of examining the preciseness of ML techniques in minimizing false alarm and enhanced model prediction.

Table III presents the level of accuracy of the different GPR kernels analyzed. Following the investigation, the result shows that RQGPR decreases MAE by 71.73% and surpasses other kernels like MGPR (17.02%), EGPR (14.46%), and SEGPR (0.00%) the baseline.

TABLE III
MODEL EFFECTIVENESS ANALYSIS AND VALIDATION

Dataset	GPR Models	MAE	Reduction Rate
CIRA-CIC-DoHBrw-2020	**RQGPR**	**0.57734**	**67.61%**
	EGPR	0.63017	63.10%
	MGPR	0.76502	51.56%
	SEGPR	1.1697	0.00%
ICS-SCADA cybersecurity dataset	**RQGPR**	**0.053081**	**71.73%**
	MGPR	0.15604	17.02%
	EGPR	0.16086	14.46%
	SEGPR	0.18805	0.00%

Compared to other kernels, the MAE of the proposed RQGPR is the lowest, as evidenced in Fig. 4. It indicates that the RQGPR supports improved prediction accuracy and reduction of false alarm rate. The summarized equations (3), (4) and (5) of the covariance function is analyzed and verified that the rational quadratic kernel, when applied in prediction as proffered, is most suitable and dependable for predictions.

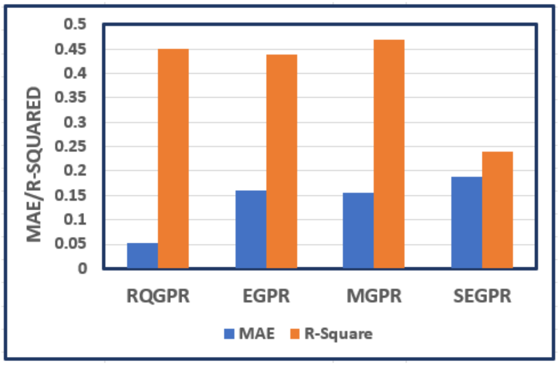

Fig. 4. This is the performance graph of the proposed RQGPR showing it having the best MAE and R-Squared result in comparison to other kernels. Note the least error rate.

C. Comparison of some ML Algorithms to the Performance of the GPR

This section compares the GPR's accuracy to other regressors, such as ML techniques. Some related articles include the k-nearest neighbor (KNN), neural network (NN), random forest (RF), fine tree (FT), coarse tree (CT), medium tree (MT) and support vector machine (SVM) methods. For more facts on GPR, toolbox, and coding model, see [27]. In addition the superior performance of the proposed RQGPR over other GP candidates, Fig. 5 further shows that it outperformed other ML algorithms for attack classification. The response plot value in Fig. 6 is the prediction observation plot illustrating the features of the True response (with higher observation) against the predicted response.

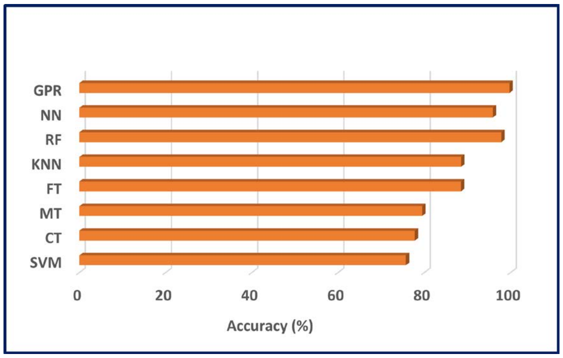

Fig. 5. Performance of Proposed RQGPR showing the highest accuracy in comparison to some ML algorithms.

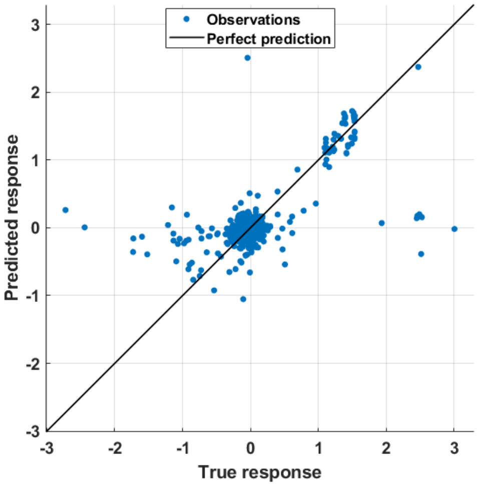

Fig. 6. Predictions vs True response of the Proposed Rational Quadratic GPR (RQGPR)

D. Proposed Model Result Validation using the CIRA-CIC-DoHBrw-2020 Datasets

Consistent with the result, Table III validates the proposed kernel as it had the least MAE. Following the investigation, the result shows that RQGPR decreases MAE by 67.61% and surpasses other kernels like MGPR (51.56%), EGPR (63.10%), and SEGPR (0.00%) the baseline. This result is in Fig. 7.

V. CONCLUSION

This study proposed RQGPR for SCADA network communication attack detection using cyber-security datasets. The findings reveal that the proposed RQGPR outperformed other cutting-edge kernels for prediction efficiency and reduction in false alarm rates, as seen by the 71.73% reduction in MAE. SCADA network communication traffic is generated regularly due to the polling method used for data collection. As a result, traffic patterns are not as dependent on human activity as in traditional IoT networks. To validate the proposed model's applicability in a scenario other than the SCADA network,

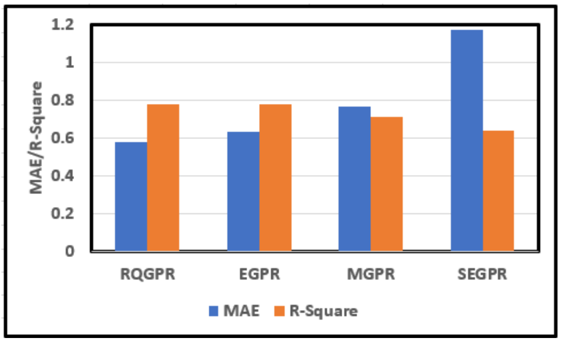

Fig. 7. When compared to other kernels, the proposed RQGPR performance shows that it has the best MAE and R-Squared results. It is worth noting that it has the lowest error rate.

the proposed model evaluated the CIRA-CIC-DoHBrw-2020 dataset. The choice of this dataset is a result of being a network traffic cyber-security dataset with present-day attacks from DNS tunneling. The experimental results demonstrate the efficiency of the proposed model with an MAE reduction of 67.61%, with a superior detection accuracy in a comparative analysis with other-state-of-art models. In the future, it will be interesting to broaden the comparison by examining more current cyber-security datasets with additional attributes to reveal the flexibility and tenacity of the kernels.

Acknowledgment

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] M. A. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin, and M. Samaka, "Scada system testbed for cybersecurity research using machine learning approach," *Future Internet*, vol. 10, no. 8, p. 76, 2018.

[2] D.-S. Kim and H. Tran-Dang, "Industrial Sensors and Controls in Communication Networks," *Computer Communications and Networks. Springer International Publishing, Cham*, 2019.

[3] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm," *IEEE Access*, vol. 9, pp. 154 892–154 901, 2021.

[4] C. M. Morsey, *Supervisory Control and Data Acquisition (SCADA) Systems and Cyber-Security: Best Practices to Secure Critical Infrastructure*. Robert Morris University, 2017 [Online], Available: https://eric.ed.gov/?id=ED578127.

[5] H. Tran-Dang, N. Krommenacker, P. Charpentier, and D.-S. Kim, "Towards the Internet of Things for Physical Internet: Perspectives and Challenges," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4711–4736, 2020.

[6] J. Choi, "Data-Aided Sensing for Gaussian Process Regression in IoT Systems," *arXiv:2011.11725v1*, pp. 1–10, 2020.

[7] C. Beecks, K. W. Schmidt, F. Berns, and A. Grass, "Gaussian Processes for Anomaly Description in Production Environments," in *EDBT/ICDT 2019 Joint Conference, Lisbon, Portugal*, 2019.

[8] J. N. Njoku, M. E. Morocho-Cayamcela, A. Caliwag, P. Xiao, and W. Lim, "Predicting Target Data Rates for Dynamic Spectrum Allocation using Gaussian Process Regression," *ICT Express*, 2021.

[9] S. Wu, "Moments of Complex Gaussian Ratios," *IEEE Communications Letters*, vol. 23, no. 1, pp. 88–91, 2018.

[10] B. Farzaneh, M. A. Montazeri, and S. Jamali, "An Anomaly-Based IDS for Detecting Attacks in RPL-Based Internet of Things," in *2019 5th International Conference on Web Research (ICWR)*, 2019, pp. 61–66.

[11] F. Farivar, M. S. Haghighi, A. Jolfaei, and M. Alazab, "Artificial Intelligence for Detection, Estimation, and Compensation of Malicious Attacks in Nonlinear Cyber-Physical Systems and Industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2716–2725, 2020.

[12] Z. Wei, P. Li, and H. Yue, "A Foreground-Background Segmentation Algorithm for Video Sequences," in *2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 2015, pp. 340–343.

[13] C. I. Nwakanma, L. A. C. Ahakonye, J.-M. Lee, and D.-S. Kim, "Selecting Gaussian Process Regression Kernels for IoT Intrusion Detection and Classification," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2021, pp. 462–465.

[14] M. Injadat, A. Moubayed, and A. Shami, "Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach," in *2020 32nd International Conference on Microelectronics (ICM)*, 2020, pp. 1–4.

[15] W. Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, and N. Bouguila, "Network Anomaly Intrusion Detection Using a Nonparametric Bayesian Approach and Feature Selection," *IEEE Access*, vol. 7, pp. 52 181–52 190, 2019.

[16] A. Gupta, C. G. Christodoulou, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Gaussian Processes for Direction-of-Arrival Estimation With Random Arrays," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2297–2300, 2019.

[17] T. Teng, J. Chen, Y. Zhang, and B. K. H. Low, "Scalable Variational Bayesian Kernel Selection for Sparse Gaussian Process Regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 5997–6004.

[18] W. Cui and H. Wang, "Anomaly Detection and Visualization of School Electricity Consumption Data," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017, pp. 606–611.

[19] W. Wan, Y. Wang, C. Long, J. Wei, J. Zhao, and G. Du, "An Attack Behaviors Prediction Model Based on Bag Representation in Time Series," in *2019 IEEE 44th Conference on Local Computer Networks (LCN)*, 2019, pp. 113–116.

[20] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[21] A. Singh, J. Nagar, S. Sharma, and V. Kotiyal, "A Gaussian Process Regression Approach to Predict the K-Barrier Coverage Probability for Intrusion Detection in Wireless Sensor Networks," *Expert Systems with Applications*, vol. 172, no. 114603, pp. 1–1, 2021.

[22] U. Adhikari, S. Pan, T. Morris, R. Borges, and J. Beave, "Industrial Control System (ICS) Cyber Attack Datasets," *datasets used in the experimentation.[Online]. Available: https://sites. google. com/a/uah. edu/tommy-morris-uah/ics-data-sets*, 2019.

[23] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM computing surveys (CSUR)*, vol. 50, no. 6, pp. 1–45, 2017.

[24] "CIRA-CIC-DoHBrw-2020 Dataset," 2020. [Online]. Available: https://www.unb.ca/cic/datasets/dohbrw-2020.html

[25] M. MontazeriShatoori, L. Davidson, G. Kaur, and A. H. Lashkari, "Detection of DoH Tunnels using Time-series Classification of Encrypted Traffic," in *The 5th IEEE Cyber Science and Technology Congress, Calgary*, 06 2020.

[26] D. Cao, J. Zhao, W. Hu, Y. Zhang, Q. Liao, Z. Chen, and F. Blaabjerg, "Robust Deep Gaussian Process-based Probabilistic Electrical Load Forecasting against Anomalous Events," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[27] D. K. Duvenaud, "Automatic Model Construction with Gaussian Processes," Ph.D. dissertation, University of Cambridge, Pembroke College, 2014.