The Role of One-Class Classification in Detecting Cyberattacks in Critical Infrastructures

Patric Nader(✉), Paul Honeine, and Pierre Beauseroy

Institut Charles Delaunay (CNRS),
Université de Technologie de Troyes, Troyes, France
{patric.nader,paul.honeine,pierre.beauseroy}@utt.fr

Abstract. The security of critical infrastructures has gained a lot of attention in the past few years with the growth of cyberthreats and the diversity of cyberattacks. Although traditional IDS update frequently their databases of known attacks, new complex attacks are generated everyday to circumvent security systems and to make their detection nearly impossible. This paper outlines the importance of one-class classification algorithms in detecting malicious cyberattacks in critical infrastructures. The role of machine learning algorithms is complementary to IDS and firewalls, and the objective of this work is to detect intentional intrusions once they have already bypassed these security systems. Two approaches are investigated, Support Vector Data Description and Kernel Principal Component Analysis. The impact of the metric in kernels is investigated, and a heuristic for choosing the bandwidth parameter is proposed. Tests are conducted on real data with several types of cyberattacks.

Keywords: Critical infrastructures · Intrusion detection · One-class classification · SCADA systems

1 Introduction

Nowadays, the control of the majority of critical infrastructures is accomplished via Supervisory Control And Data Acquisition (SCADA) systems, which allow remote monitoring and control to physical systems such as electrical power grids, oil and natural gas pipelines, chemical processing plants, water distribution, wastewater collection systems and nuclear power plants [1]. The principal components of SCADA systems are: (a) The Human Machine Interface (HMI) allows operators to monitor the state of the process under control and modify its control settings, (b) the Master Terminal Unit (MTU) stores and processes the information from the field and transmits control signals, and (c) the Remote Terminal Units (RTU) receive commands from the MTU to control the local process, acquire data from the field and transmit it to the MTU. The common protocols (ModBus, Profibus, DNP3) used in the communication between these

C.G. Panayiotou et al. (Eds.): CRITIS 2014, LNCS 8985, pp. 244–255, 2016.
DOI: 10.1007/978-3-319-31664-2_25

components present many vulnerabilities [2]. These protocols do not perform any authentication mechanism between Master and Slave, do not check for the integrity of the command packets and do not apply any anti-repudiation or anti-replay mechanisms [3].

First generations of SCADA networks operate in isolated environments, with no connectivity to any system outside the network. Nowadays, SCADA systems use public network for system-to-system interconnection, which has introduced numerous vulnerabilities and has exposed the critical infrastructures to new sources of potential threats [4]. Many intentional cyberattacks against critical infrastructures relying on SCADA networks occured in the past few years. In 2000, an ex-employee of Maroochy Water Services in Australia released one million liters of untreated sewage into local parks and rivers [5]. In 2003, the Slammer worm penetrated Ohios Davis-Besse nuclear power plant and disabled a safety monitoring system for nearly five hours [6]. In 2006, a hacker penetrated a water filtering plant in Pennsylvania (USA) and installed malicious software capable of affecting the plants water treatment operations [7]. In 2009, cyberspies penetrated the U.S. electrical grid and left behind software programs that could be used to disrupt the system [8]. The most complex malware Stuxnet was discovered 2010. It installs malicious programs replacing the PLCs original file in a manner undetectable by the PLC operator [9]. The ultimate goal of Stuxnet was to sabotage nuclear centrifuges used for enriching uranium [10].

The vulnerabilities in the communication protocols between SCADA components and the intensive use of internet and communication technologies have increased the cyberthreats and opened new ways for carrying out cyberattacks against critical infrastructures relying on SCADA networks [11]. For these reasons, securing the critical infrastructures has become the ultimate priority of the researchers with the growth of cyberthreats and the diversity of aforementioned cyberattacks. Yang *et al.* proposed in [12] a signature-based approach that matches signatures of known attacks with the network traffic, and a model-based approach for detecting intrusions in SCADA systems. The first one cannot detect new attacks not existing in their databases, and the second one needs the existence of the exact system's model which is not the case for the majority of the critical infrastructures. A Bayesian network was implemented in [13] to reduce the false positive rate, but this statistical model relies on the conditional dependencies between the system's variables. A collaborative intrusion detection mechanism using a centralized server that dispatches activities coming from suspicious IP addresses was proposed in [14]. This approach do not provide any specific technique for identifying high level and complex cyberattacks. Carcano *et al.* presented in [15] a critical state-based IDS for a given industrial installation, which can only detect a particular type of cyberattacks against PLC systems. Morris *et al.* elaborated a SCADA testbed in [16,17], where false commands and responses were injected into the SCADA network to investigate the vulnerabilities of functional control systems. Cyberattacks studied in their testbed include command injection, response injection and denial of service (DOS) attacks. The complexity of the critical infrastructures and the diversity

of cyberthreats restrict the use of model-based approaches, and emphasize the potential role of non-parametric methods in detecting intrusions.

Real world data analysis problems require, most of the time, nonlinear methods for detecting patterns and interdependencies within the data [18]. Machine learning techniques have become very popular in the past few years since they provide a powerful way for detecting nonlinear relations using linear algorithms in the feature space [19,20]. This paper outlines the complementary role of machine learning algorithms to traditional IDS in detecting intrusions in critical infrastructures. Two distinct approaches are investigated, the Support Vector Data Description (SVDD) [21] and the Kernel Principal Component Analysis (KPCA) [22]. This paper also studies the impact of varying the metric norm in the kernel functions, and proposes a heuristic for choosing the bandwidth parameter without any computational costs. The tests are conducted on real data from the gas pipeline testbed [16,17]. The remainder of this paper is organized as follows. Section 2 provides an overview on kernel methods for one-class classification, namely the SVDD and the KPCA. Section 3 discusses the metric variation and the heuristic for choosing the bandwidth parameter. Section 4 describes the gas pipeline testbed and the results on the real datasets. Section 5 provides conclusion and future works.

2 Kernel Methods for One-Class Classification

kernel methods have been widely used in the past few years to discover hidden regularities in large volumes of data [18]. They use positive definite kernel functions to map the data into a reproducing kernel Hilbert space (RKHS) via the mapping function $\phi(\cdot)$, and provide an elegant way to learn a nonlinear system without the need of an exact physical model [23]. In industrial systems, the majority of the available data designates the normal functional mode, and it is very difficult to acquire data related to malfunctioning or critical states [24]. For this reason, the role of one-class classification has been growing in detecting machine faults and intrusions, especially in critical infrastructures [25–27]. Each training sample $\boldsymbol{x}_i$ can represent measurements such as the gas pressure in a specific time, the temperature, the water level, the pressure for three consecutive instants, etc. One-class classification algorithms learn the normal behavior of the system through the relations between these components, and a decision function tests new samples to classify them as normal or outliers (suspicious behavior).

2.1 Support Vector Data Description

Support Vector Data Description (SVDD) estimates a spherically shaped decision boundary with minimum radius that encloses most of the training data $\phi(\boldsymbol{x}_i)$ in the feature space $\mathcal{H}$ [21]. The hypersphere is characterized by its center $\boldsymbol{a}$ and its radius $R > 0$, and we minimize its volume by minimizing R^2. The presence of some outliers in the training set is allowed by introducing the slack

variables $\xi_i \geq 0$. Samples that lay outside this description are considered outliers, and they should be rejected. This boils down to the following constrained optimization problem:

$$\min_{\boldsymbol{a},R,\xi_i} R^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i \tag{1}$$

subject to $\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|_{\mathcal{H}}^2 \leq R^2 + \xi_i \quad \text{and} \quad \xi_i \geq 0\ \forall i = 1, ..., N$. The predefined parameter ν represents an upper bound on the fraction of outliers, and regulates the trade-off between the volume of the hypersphere and the number of outliers. Considering the Lagrangian of the above constrained optimization problem, and incorporating the relations from its partial derivatives with respect to R, $\boldsymbol{a}$ and ξ_i gives us the following objective functional to be maximized with respect to the Lagrangian multipliers α_i : $L = \sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}_i, \boldsymbol{x}_i) - \sum_{i,j=1}^{N}\alpha_i\alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, subject to $\sum_{i=1}^{N}\alpha_i = 1$ and $0 \leq \alpha_i \leq 1/\nu N$. The solution of this quadratic programming problem is found using any off-the-shelf optimization technique, i.e., matlab's function quadprog.

In order to evaluate a new sample $\boldsymbol{z}$, we calculate the distance between the center of the sphere $\boldsymbol{a}$ and $\phi(\boldsymbol{z})$ in the feature space. If this distance is smaller than the radius, namely $\|\phi(\boldsymbol{z}) - \boldsymbol{a}\|_{\mathcal{H}}^2 \leq R^2$, $\boldsymbol{z}$ is accepted as a normal sample. Otherwise, $\boldsymbol{z}$ is considered as an outlier and an intrusion is detected. The radius of the optimal hypersphere is obtained with the distance in the feature space $\mathcal{H}$ from the center $\boldsymbol{a}$ to any sample $\phi(\boldsymbol{x}_k)$ on the boundary:

$$R^2 = k(\boldsymbol{x}_k, \boldsymbol{x}_k) - 2\sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}_k, \boldsymbol{x}_i) + \sum_{i,j=1}^{N}\alpha_i\alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

2.2 Kernel Principal Component Analysis

Kernel Principal Component Analysis (KPCA) is a nonlinear application of PCA in a kernel-defined feature space, where using $k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} \cdot \boldsymbol{y})$ is equivalent to performing the original PCA [22]. Hoffmann proposed in [26] the use the *reconstruction error* as a measure of novelty, since it takes into account the heterogeneous variance of the distribution of the data in the feature space. The first step in Hoffman's algorithm is to find eigenvalues λ and eigenvectors $\boldsymbol{v}$ of the covariance matrix $\widetilde{C}$ in the feature space $\mathcal{H}$, satisfying $\lambda\boldsymbol{v} = \widetilde{C}\boldsymbol{v}$. The second step is to project the data into the subspace spanned by the most relevant eigenvectors. Each $\boldsymbol{v}$ is a linear combination of the mapped data and takes the following form: $\boldsymbol{v} = \sum_{i=1}^{N}\alpha_i\phi(\boldsymbol{x}_i)$, and the coefficients α_i are given by solving the following eigen decomposition problem $N\lambda\alpha = \widetilde{K}\alpha$. The centered kernel matrix $\widetilde{K}$ is used in the optimization problem without the need to compute directly $\widetilde{C}$.

After projecting the data into the subspace spanned by the most relevant eigenvectors, the distance between each sample and its projection is computed. This distance is the reconstruction error, and it is used for novelty detection. Let $\mathcal{P}$ be the projection operator, the reconstruction error is computed as follows:

$$\|\widetilde{\phi}(\boldsymbol{z}) - \mathcal{P}\widetilde{\phi}(\boldsymbol{z})\|_{\mathcal{H}}^2 = \langle\widetilde{\phi}(\boldsymbol{z}), \widetilde{\phi}(\boldsymbol{z})\rangle - 2\langle\widetilde{\phi}(\boldsymbol{z}), \mathcal{P}\widetilde{\phi}(\boldsymbol{z})\rangle + \langle\mathcal{P}\widetilde{\phi}(\boldsymbol{z}), \mathcal{P}\widetilde{\phi}(\boldsymbol{z})\rangle. \tag{2}$$

Knowing in advance the number of outliers among the training dataset, an error threshold is fixed. If the reconstruction error of a new sample is smaller than this threshold, the corresponding sample is treated as a normal sample. Otherwise, it is considered as an outlier and an intrusion is detected.

3 Metric Variation and Parameter Optimization

The Gaussian kernel is adopted in our simulations, since it is the most common and suitable kernel for one-class classification problems [28]. The Gaussian kernel is given as follows: $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2})$, where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are input samples, $\|\cdot\|_2$ represents the l_2-norm in the input space, and σ is the bandwidth parameter of the kernel. The choice of the metric and σ has a great impact on the decision function of the classifier. The variation of the norm and the heuristic for choosing the bandwidth parameter are detailed in the next subsections.

3.1 Norm Variation

In order to understand the impact of l_p-norm on the classifier, the variation in the behavior of different norms in a 2-dimensional space is illustrated in Fig. 1. Each sample has two characteristics, feature 1 and feature 2, and p takes one of the values $\frac{3}{4}, 1, \frac{3}{2}, 2, 3, 4, 7$ and ∞. Each color represents equidistant contours with reference to the origin O. The following example clarifies the different behavior of several norms towards the same sample. The samples B and C are equidistant from the origin O with the l_2-norm, and D is much closer. However, for the l_1-norm, C and D are equidistant and much closer than B. Therefore, as p decreases, the norms are more sensitive on simultaneous variations of multiple features which become as important as large variation in a single one.

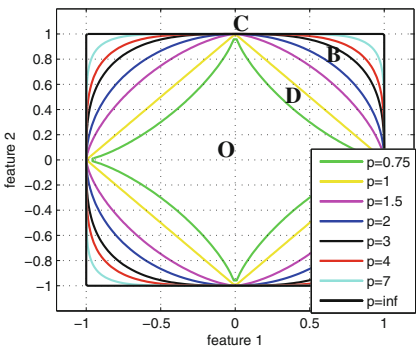

Fig. 1. The variation in the behavior of different norms ranging from $p = 0.75$ to $p = \infty$, where each color represents equidistant contours with reference to the origin O. The norms become more sensitive on simultaneous variation of multiple features as p decreases (Color figure online).

In critical infrastructures and industrial processes, the value of each variable is important to evaluate the state of the system, and to predict whether the process is leading to a critical state. The diversity of the studied physical processes requires more adapted kernels that depend on the behavior of the measured variables, i.e., the pressure inside a gas pipeline, the water level of a water distribution system, the temperature of a boiling water reactor, etc. For this reason, the choice of the norm in kernels affects the distribution of the data in the feature space, and has a great impact on the decision function of the classifier.

3.2 Choice of the Bandwidth Parameter

The performance of classification algorithms is highly related to the choice of the bandwidth parameter σ, as well as on the kernel's norm. σ plays a crucial role in defining the description boundary around the training data. With a large value of σ, the classifier underfits the data and we obtain a loose description boundary, where a small value of σ leads to overfitting. Several approaches were proposed in the literature for computing this parameter, but they are time consuming and do not always lead to an optimal choice [29–31].

The bandwidth parameter σ depends on multiple features, namely the spread of the training dataset, the number of input samples and the fraction of samples considered as outliers [32]. The estimation of σ should take into consideration all these factors. Therefore, we propose to use in the one-class classification algorithms the following expression for computing σ:

$$\sigma = \frac{d_{max}}{\sqrt{2M}},$$

where d_{max} refers to the maximal distance between any two samples in the input space, and M represents the upper bound on the number of outliers among the training dataset. The metric of the distance used in the kernel function is the same as the one in the expression of σ. The experiments showed that this proposed heuristic gives remarkable results without the need for the time consuming cross-validation step.

Fig. 2. Gas pipeline testbed

4 Results on the Gas Pipeline Testbed

In this paper, one-class classification algorithms are applied on real data from the gas pipeline testbed of the Mississippi State University SCADA Laboratory [16,17]. The gas pipeline illustrated in Fig. 2 is used to move natural gas or any other petroleum products to the market. Its control system contains an air pump that pumps air into the pipeline, a pressure sensor which allows pressure visibility at the pipeline and remotely on the HMI, a release valve and a solenoid release valve to loose air pressure from the pipeline. This testbed represents a typical SCADA system embracing a MTU, RTU and a HMI. Cyberattacks on the gas pipeline monitoring system can cause a loss of control of the physical process, and this may lead to huge financial and physical losses.

The pipeline operates in three principal modes; the first mode is characterized by a very low pressure maintained around 0.1 PSI, the second mode keeps it around 10 PSI (9 to 11 PSI), while the third mode should maintain the pressure around 20 PSI (18 to 22 PSI). The pressure greater than 22 PSI and the transitional states between different modes are considered as outliers. Several types of false commands and responses are injected into the normal behavior zone of the system to make its behavior abnormal. The *fast change response attack* returns measurements that change very fast opposed to the normal behavior of the pipeline. The *burst response injection attack* injects at high frequency a single value equals to 20 PSI while the system is running in several modes. The *wave response injection attack* injects pressure responses that vary in a wave form around 9 PSI which imitates exactly the second normal mode, while the real system is dealing with high pressures in the third mode. The primary objective of this paper is to detect these common and dangerous attacks that imitate the normal behavior of the system, and hide the real functioning status.

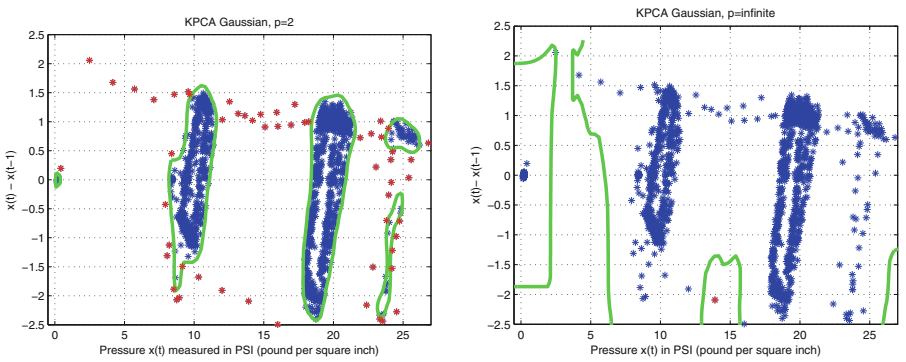

Fig. 3. Results on the gas pipeline real data with the KPCA approach. The decision boundaries are given by the green lines, the outliers correspond to the red samples and the normal samples are in blue. The l_2-norm (left) gives a good description while the infinite norm (right) underfits the data with a loose descriptions (Color figure online).

Table 1. Time computational cost of several approaches for computing the bandwidth parameter.

approach	5-fold CV	11-value range for σ	limited range (5 values)	proposed heuristic
SVDD	8 h 5 min	2 h 58 min	1 h 26 min	14.78 s
KPCA	3 h 47 min	1 h 32 min	34.6 min	14.78 s

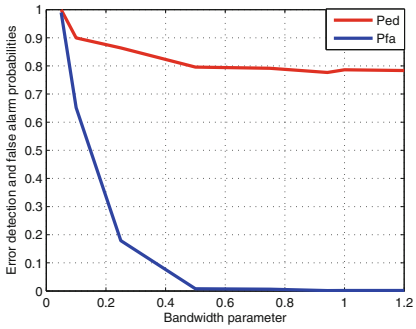

Fig. 4. The error detection and the false alarm probabilities as a function of the bandwidth parameter σ. The proposed heuristic leads to $\sigma = 0.9427$ with the highest error detection and the lowest false alarm rates.

Let $x(t)$ be the pressure in the pipeline at instant t. In normal functioning modes, the pressure measurements of two consecutive instants must be close to each other, and a gap between two consecutive instants may be a strong sign of a cyberattack. Therefore, the time series is folded into 2-dimensional input vectors composed of the pressure at instant t and the difference in the pressure between instants t and $t-1$, namely $\boldsymbol{x}_t = [x(t) \quad x(t) - x(t-1)]$. The training phase is made on a train set of 2000 samples, and the tests are conducted on five different test sets containing several types of cyberattacks. The outliers in the test sets represent the simulated attacks that have to be detected by one-class classification algorithms. The different types of attacks are shown in Fig. 5.

The results on real data from the gas pipeline testbed for the KPCA approach are shown in Fig. 3. The decision boundary encloses the samples accepted as normal data, while the samples considered as outliers are rejected outside the boundary. The best results are obtained with the l_2-norm and the l_1-norm, having a tight decision boundary enclosing the normal behavioral modes. For small values of p, the norms become very sensitive to simultaneous variation of multiple features, and this leads to overfitting the data. On the other hand, the results for the values of p greater than $p = 2$ become worse as p increases, with a loose decision boundary that underfits the data. We have similar results with the SVDD approach. The prediction time for testing a new sample is 0.096 s with SVDD and 0.049 s with KPCA, which is very interesting in monitoring critical infrastructures. The error probabilities of the different types of cyberattacks are detailed in Table 2. The l_1-norm outperforms the l_2-norm in the wave and

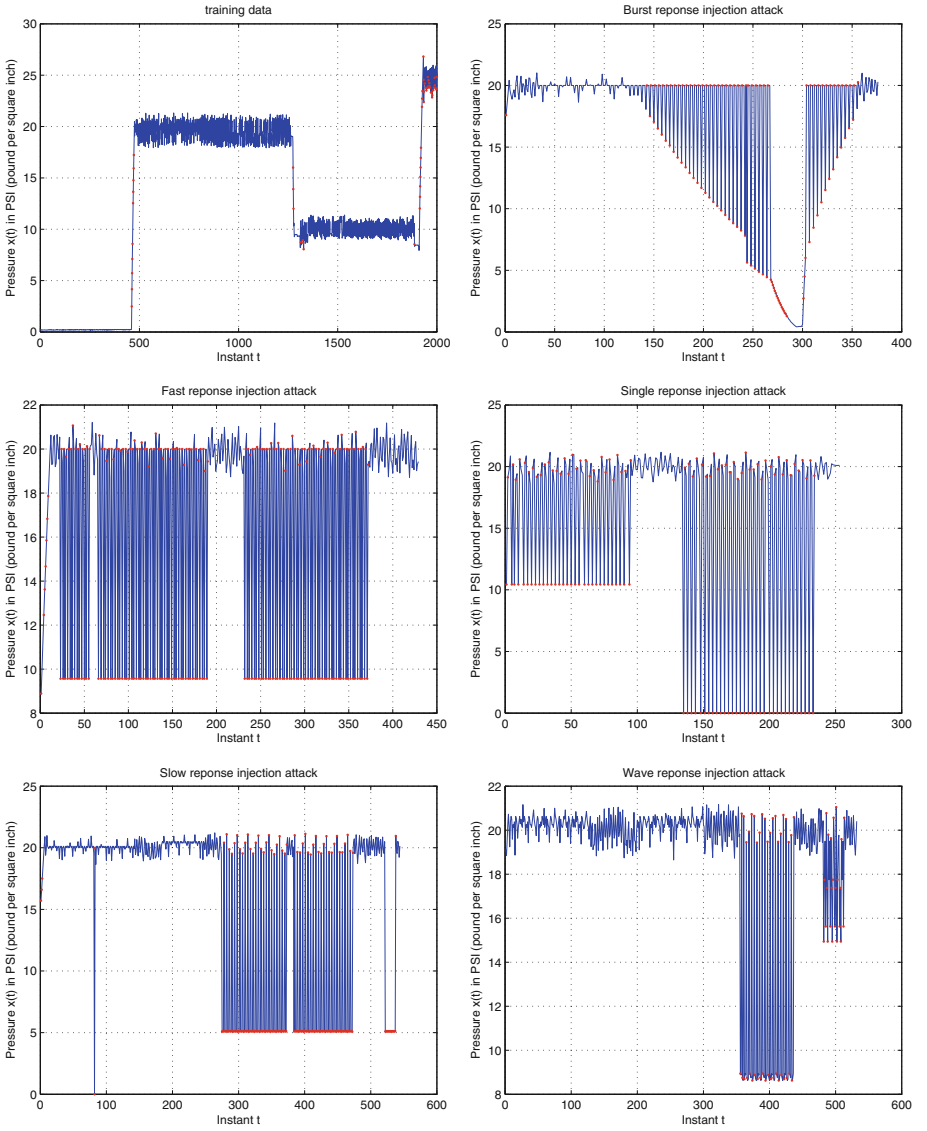

Fig. 5. Detection of outliers for several types of attacks with the SVDD approach using the l_1-norm. The blue samples refer to the data accepted as normal data while the red samples are considered as outliers (Color figure online).

the slow response injection attacks, where the data contain small simultaneous variation of its features. The best results are achieved with the slow and the single attacks having error detection probabilities around 99.52 %. We note that since these injections have already bypassed IDS and firewalls, the detection of the malicious attacks by operators comes mostly far too late after some severe

Table 2. The confusion matrix of several types of attacks with the KPCA approach.

		l_2-norm		l_1-norm	
		Normal	Outlier	Normal	Outlier
Slow injection	Normal	99.41	0.59	99.7	0.3
	Outlier	0.95	99.05	0.48	99.52
Fast injection	Normal	98.3	1.7	99.35	0.65
	Outlier	11.6	88.4	11.6	88.4
Burst injection	Normal	99.3	0.7	99.3	0.7
	Outlier	27.9	72.1	31.33	68.67
Single injection	Normal	98.37	1.63	99.2	0.8
	Outlier	0.78	99.22	0.78	99.22
Wave injection	Normal	98.8	1.2	98.09	1.91
	Outlier	35.1	64.9	34.21	65.79

consequences on the industry. This is where machine learning techniques play a crucial role to learn the industrial systems in order to detect all kinds of intrusions and avoid physical, financial and human lives losses.

The bandwidth parameter is computed as detailed in the previous section. We compared the time computational cost of the proposed heuristic with three other common methods existing in the literature as shown in Table 1. Our approach is clearly hundreds of times faster than the other methods, and it takes exactly the same time with SVDD and KPCA. In addition, the error detection and the false alarm probabilities for several values of σ are computed, and the results are illustrated in Fig. 4. The proposed heuristic leads to $\sigma = 0.9427$ having the highest error detection rates and the lowest false alarm rates, which confirms its relevance.

5 Conclusion

In this paper, we showed the importance of the complementary role of one-class classification algorithms in detection malicious cyberattacks in critical infrastrucres relying on SCADA systems. The tests were conducted on real data containing several types of cyberattacks. We studied the impact of varying the norm in the kernels on the decision function of the classifier. We also proposed a simple heuristic for computing the bandwidth parameter of the Gaussian kernel, which led to the highest error detection and the lowest false alarm rates with minimum time computational cost. For future works, we are investigating a sparse one-class classification approach that should fasten the learning phase of the system. We are also working on increasing the performance of the algorithm by decreasing the time to test new samples. Finally, online one-class classification techniques should be integrated in the security systems critical infrastructures to improve the live detection of cyberattacks and reduce their consequences.

Acknowledgment. The authors would like to thank Thomas Morris and the Mississippi state university SCADA Laboratory for providing the real SCADA dataset.

References

1. Stouffer, K., Falco, J., Kent, K.: Guide to supervisory control and data acquisition (scada) and industrial control systems security. Technical report, National Institute of Standards and Technology (NIST) (2006)
2. Fovino, I., Masera, M., Guidi, L., Carpi, G.: An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants. In: 3rd Conference on Human System Interactions (HSI), pp. 679–686 (2010)
3. Fovino, I., Coletta, A., Carcano, A., Masera, M.: Critical state-based filtering system for securing SCADA network protocols. IEEE Trans. Ind. Electron. **59**, 3943–3950 (2012)
4. Ten, C.W., Hong, J., Liu, C.C.: Anomaly detection for cybersecurity of the substations. IEEE Trans. Smart Grid **2**, 865–873 (2011)
5. Slay, J., Miller, M.: Lessons learned from the maroochy water breach. In: Goetz, E., Shenoi, S. (eds.) Critical Infrastructure Protection, pp. 73–82. Springer, US (2007)
6. Christiansson, H., Luiijf, E.: Creating a European SCADA security testbed. In: Goetz, E., Shenoi, S. (eds.) Critical Infrastructure Protection. IFIP International Federation for Information Processing, vol. 253, pp. 237–247. Springer, US (2007)
7. Cárdenas, A.A., Amin, S., Lin, Z.S., Huang, Y.L., Huang, C.Y., Sastry, S.: Attacks against process control systems: risk assessment, detection, and response. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, pp. 355–366. ACM, New York (2011)
8. Gorman, S.: Electricity grid in U.S. Penetrated by spies. Wall Street J. (2008)
9. Chen, T., Abu-Nimeh, S.: Lessons from stuxnet. Computer **44**, 91–93 (2011)
10. Langner, R.: Stuxnet: dissecting a cyberwarfare weapon. IEEE Secur. Priv. **9**, 49–51 (2011)
11. Urias, V., Van Leeuwen, B., Richardson, B.: Supervisory command and data acquisition (SCADA) system cyber security analysis using a live, virtual, and constructive (LVC) testbed. In: Military Communication Conference - MILCOM, pp. 1–8 (2012)
12. Yang, Y., McLaughlin, K., Littler, T., Sezer, S., Pranggono, B., Wang, H.: Intrusion detection system for IEC 60870-5-104 based SCADA networks. In: 2013 IEEE Power and Energy Society General Meeting (PES), pp. 1–5 (2013)
13. Bigham, J., Gamez, D., Lu, N.: Safeguarding SCADA systems with anomaly detection. In: Gorodetsky, V., Popyack, L.J., Skormin, V.A. (eds.) MMM-ACNS 2003. LNCS, vol. 2776, pp. 171–182. Springer, Heidelberg (2003)
14. Gross, P., Parekh, J., Kaiser, G.: Secure selecticast for collaborative intrusion detection systems. In: 3rd International Workshop on Distributed Event-Based Systems (DEBS 2004), Edinburgh, Scotland, UK (2004)
15. Carcano, A., Coletta, A., Guglielmi, M., Masera, M., Fovino, I., Trombetta, A.: A multidimensional critical state analysis for detecting intrusions in SCADA systems. IEEE Trans. Ind. Inf. **7**, 179–186 (2011)
16. Morris, T., Vaughn, R.B., Dandass, Y.S.: A testbed for SCADA control system cybersecurity research and pedagogy. In: CSIIRW, Oak Ridge, Tennessee (2011)

17. Morris, T., Srivastava, A., Reaves, B., Gao, W., Pavurapu, K., Reddi, R.: A control system testbed to validate critical infrastructure protection concepts. Int. J. Crit. Infrastruct. Prot. **4**, 88–103 (2011)
18. Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning. Ann. Stat. **36**, 1171–1220 (2008)
19. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, New York (2004)
20. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**, 15:1–15:58 (2009)
21. Tax, D.M.J., Duin, R.P.W.: Support vector data description. Mach. Learn. **54**, 45–66 (2004)
22. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**, 1299–1319 (1998)
23. Noumir, Z., Honeine, P., Richard, C.: Online one-class machines based on the coherence criterion. In: Proceedings of the 20th European Conference on Signal Processing, Bucharest, Romania (2012)
24. Khan, S.S., Madden, M.G.: A survey of recent trends in one class classification. In: Coyle, L., Freyne, J. (eds.) AICS 2009. LNCS, vol. 6206, pp. 188–197. Springer, Heidelberg (2010)
25. Mazhelis, O.: One-class classifiers : a review and analysis of suitability in the context of mobile-masquerader detection. S. Afr. Comput. J. **36**, 29–48 (2006)
26. Hoffmann, H.: Kernel PCA for novelty detection. Pattern Recogn. **40**, 863–874 (2007)
27. Nader, P., Honeine, P., Beauseroy, P.: Intrusion detection in SCADA systems using one-class classification. In: Proceedings of the 21th European Conference on Signal Processing, Marrakech, Morocco (2013)
28. Schölkopf, B., Platt, J.C., Shawe-Taylor, J.C., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Comput. **13**, 1443–1471 (2001)
29. Soares, C., Brazdil, P.B., Kuba, P.: A meta-learning method to select the kernel width in support vector regression. Mach. Learn. **54**, 195–209 (2004)
30. Cherkassky, V., Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw. **17**, 113–126 (2004)
31. Gurram, P., Kwon, H.: Support-vector-based hyperspectral anomaly detection using optimized kernel parameters. IEEE Geosci. Remote Sens. Lett. **8**, 1060–1064 (2011)
32. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River (1998)