# Adaptive SVDD-based Learning for False Alarm Reduction in Intrusion Detection

Tayeb Kenaza, Abdenour Labed, Yacine Boulahia and Mohcen Sebehi
*Ecole Militaire Polytechnique, BP-17, Bordj El-Bahri, 16111, Alger, Algerie*

Keywords: Intrusion Detection, False Alerts Reduction, Adaptive Learning, SVDD.

Abstract: During the last decade the support vector data description (SVDD) has been used by researchers to develop anomaly-based intrusion detection systems (IDS), with the ultimate objective to design new efficient IDS that achieve higher detection rates together with lower rates of false alerts. However, most of these systems are generally evaluated during a short period without considering the dynamic aspect of the monitored environment. They are never experimented to test their behavior in long-term, namely after some long period of deployment. In this paper, we propose an adaptive SVDD-based learning approach that aims at continuously enhancing the performances of the SVDD classifier by refining the training dataset. This approach consists of periodically evaluating the classifier by an expert, and feedback in terms of false positives and confirmed attacks is used to update the training dataset. Experimental results using both refined training dataset and compromised dataset (dataset with mislabeling) have shown promising results.

## 1 INTRODUCTION

IDS are very important tools that aim at monitoring information systems and reporting any security violation. However, current IDS have several shortcomings that must be avoided, especially the high rate of false alerts they generate. This latter is mainly due to the lack of data analysis techniques able to efficiently distinguish between legitimate and malicious traffic. In this perspective, we propose an adaptive learning approach to build an anomaly based IDS. This approach focuses only on the normal behavior and is able to progressively self-reduce the false positive. For that we use the Support Vector Data Description (SVDD) classifier which is a promising single-class classification technique. Indeed, as highlighted by several works of interest (Onoda and Kiuchi, 2012; Li and Wang, 2013), the use of SVDD shows encouraging results in the field of intrusion detection, since this area is characterized by the fact that large datasets on normal behavior may be collected whereas data on attacks are difficult to obtain, and hence SVDD seems to be an appropriate tool.

The SVDD classifier has been used by several researchers to design new approaches with the aim of achieving higher detection rates together with lower rates of false alerts. However, most of the developed systems are generally evaluated during a short period without considering the dynamic aspect of the intrusion detection environment. They are not experimented to test their behavior in long-term, namely after some long period of deployment.

Indeed, the learned classifier cannot be valid indefinitely, particularly in a changing environment such as intrusion detection in which the normal behavior of users may change and new attacks may arise. We propose in this paper a new learning approach that focuses on the continuous improvement of an SVDD classifier trough a periodic updating of the training dataset. It consists of periodically evaluating the classifier by an expert and feedback in terms of false positives and confirmed attacks will be used to update the training dataset. This approach, besides being valid for any classifier and in all fields, is more convenient for intrusion detection where both attacks and normal activities may continuously change. Experiments conducted in Section 5 confirm that anomaly-based intrusion detection is an appropriate application domain for the SVDD.

The remainder of this paper is organized as follows. In section 2, we briefly present some related works. In section 3, we give a short description of the one-class classification and the SVDD technique. In section 4, we give an overview of the improvement of the SVDD in the context of anomaly intrusion detection, and then we discuss in detail the qualified adap-

Kenaza T., Labed A., Boulahia Y. and Sebehi M..
Adaptive SVDD-based Learning for False Alarm Reduction in Intrusion Detection.
DOI: 10.5220/0005573204050412
In *Proceedings of the 12th International Conference on Security and Cryptography* (SECRYPT-2015), pages 405-412
ISBN: 978-989-758-117-5

tive SVDD by explaining the motivation behind this new approach. In section 5, some experiments are conducted to evaluate this approach. Section 6 concludes this paper.

## 2 RELATED WORKS

In this section, we summarize some previous works based on the SVDD for intrusion detection. The principle of the one-class classification, in general and the SVDD in particular, fits perfectly with the environment of intrusion detection, namely the large amount of data to be processed and the lack of information about attacks. Authors of (Onoda and Kiuchi, 2012) have underlined the power of the standard SVM technique to classify objects by optimizing the construction of a boundary between the classes. The SVDD, which is inspired by the conventional SVM but by considering only one class (the class of abundant objects), has been used. To confirm their hypothesis, tests on synthetic and real data have been conducted and comparison between the SVDD and the SVM given by the authors. In (Li and Wang, 2013), authors have applied the technique of SVDD to identify a specific type of attack, namely the denial of service (DDOS) attack. Indeed, the detection of a DDOS attack cannot be done by a conventional approach such as a detailed analysis of packets (as in misuse detection) because the system would be rapidly saturated. The authors have stated that it is more appropriate to apply the SVDD to detect this type of attack by targeting the DDOS attack class. Another similar work is presented in (Yu et al., 2008) where the objective is to apply the SVDD to detect traffic flooding. Considering that conventional SVDD is rigid even with the use of a kernel function, the authors of (Liu et al., 2010) introduced the concept of uncertainty in labeling objects for learning. The authors explained that it is possible to make mistakes when labeling objects in the training dataset and it would be therefore, interesting to associate each object with an uncertainty value. After some tests on real data, the authors have concluded the adaptability of their new approach to intrusion detection. Another technique for improving the SVDD has been proposed by (Ghasemi Gol et al., 2010), which is to surround objects of the target class by a hyperellipse instead of a hypersphere. Indeed, the authors assume that a hypersphere is a special case of hyperellipse, so using this latter could give better results. Tests achieved by the authors on different training sets confirm their assumptions and hence, introduce a new field of research that tries to improve the SVDD. Nevertheless, the mathematical formulation of this method is more complex than that of the conventional SVDD, so practical use is limited to sets of small size.

Generally, these works are interested in developing a new efficient SVDD. However, they do not consider the behavior of the developed systems in long-term. In other words, they never discuss the question if the classifier will keep the same performance after some period. Indeed, a trained classifier cannot be valid indefinitely, especially in very changing environments such as intrusion detection. During the monitoring of an information system, normal activities and attacks are often changing. In view of these findings, we propose in this paper a new learning approach that allows a continuous improvement of the SVDD classifier by updating the training dataset. This approach will be detailed in section 4.

## 3 SINGLE-CLASS CLASSIFICATION

Classification is a basic task in data analysis and machine learning. It consists of assigning a class to a set of attributes that characterize an object. Indeed, building a classifier from a set of labeled data is a central problem in machine learning. Several methods have been developed, such as decision trees, neural networks, association rules, etc (Liao et al., 2012). While it is usual to classify objects in two or more classes, the single-class classification is only focused on one class. It should be noted that the single-class classification is a recent concept in classification (Tax and Duin, 2004). In the following, we first give an overview of the single-class problem, then we present the SVDD technique used for this type of classification.

### 3.1 Motivation

In general, classification is used to classify objects in two or more classes. This classification is called "multi-class classification". But it is important to note that the use of multi-class classification requires a good knowledge of all classes of the problem being considered, that is to say the need to provide a representative number of samples of each class.

However, in the context of intrusion detection, it is difficult to have representative samples of all classes of possible behaviors of intruders (Mazhelis and Puuronen, 2007). This is because an intruder has a large number of variants to achieve the same attack. This difficulty is a constraint that prevent to completely

satisfy the assumptions of using multi-class classifications. Indeed, the inability to correctly define one or more attack classes may lead to error in the learning of a multi-class classifier, and the accuracy of the trained classifier may be strongly affected.

Taking into account this constraint requires the use of a particular type of classification called “single-class” or “one-class” classification (Tax and Duin, 2004). In the one-class classification, the training data contains only a labeled set of one class called “target”. The other classes, called “outliers”, are almost absent. So, the question is how to develop a decision boundary between the target and the outliers by considering only the target.

The one-class problem is common, especially in areas where some classes are not obtainable for cost or practical reasons. For example, providing data on industrial machine failures can be done only by deteriorating this machine in all possible ways, which is very expensive. The objective of the one-class approach is to compensate for the absence of these “outliers” by using only available data of the target.

In Figure 1 we illustrate an example of the one-class problem. We represent the target class objects by the symbol “+”. The objective is to place a boundary with a minimum size (Figure 1) that encloses the maximum number of target objects. Recall that these outliers (objects by the symbol “-”) are absent during the learning phase. We particularly show the difficulty of this task when some outliers appear in the middle of the targets space during the test.

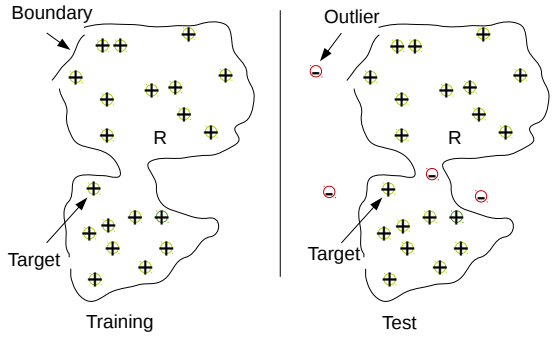


Figure 1: An example of single-class problem taken from (Desir, 2013).

## 3.2 Support Vector Data Description (SVDD)

The SVDD is a technique of one-class boundary proposed by the authors of (Tax and Duin, 2004). This technique is inspired by the SVM (applied in the case of two or more classes), proposed by Vapnik (Cortes and Vapnik, 1995). The fundamental difference with the SVM is that the SVDD do not aim at finding an optimal hyperplane separator such as in the SVM,but a hypersphere with a minimum volume (minimum radius) that encloses a maximum number of target objects in the training dataset (Figure 2). Then, a new object is accepted or rejected based on its membership or not to the hypersphere.

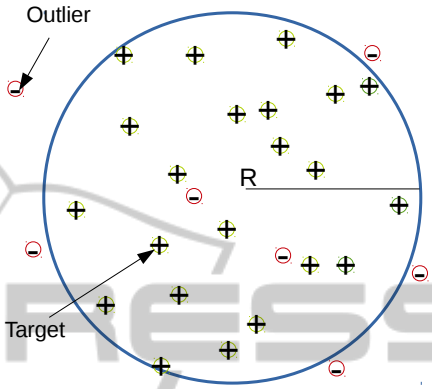


Figure 2: An example of SVDD hypersphere.

However, the SVDD such as implemented by (Tax and Duin, 2004), suffers from several drawbacks limiting its practical use. In fact, solving the associated optimization problem is feasible for a limited number of objects, but this is no longer possible for a large number of objects. Moreover, in the field of intrusion detection, we have a huge amount of data to be processed. To overcome this problem, Tax et al. (Tax and Laskov, 2003) have developed an incremental version of the SVDD. It solves the optimization problem by dividing it to a series of sub problems, so that only one sub problem is treated at once. This will enhance the performance of the SVDD in terms of execution time and memory consumption.

# 4 AN ADAPTIVE SVDD FOR ANOMALY-BASED INTRUSION DETECTION

Based on the incremental SVDD we propose an adaptive approach to build an SVDD classifier for intrusion detection. In other words, the classifier will not be trained in one step but it will be trained and refined through several iterations (Figure 3).

In the following sub sections we discuss our approach through two case studies. In the first one we use a refined training dataset, and in the second we use a training dataset with some mislabeled objects. An experimental study is given in section 5 of this paper.

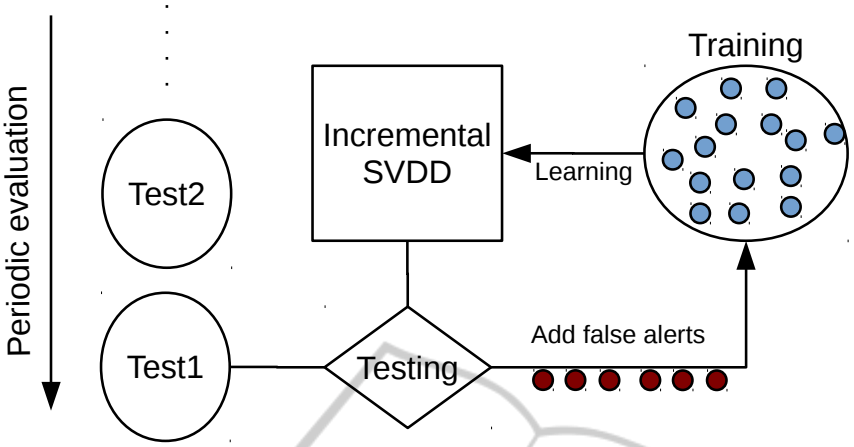


Figure 3: An iteration of the adaptive SVDD learning.

## Using a Refined Training Dataset

In this first case study we use a refined training dataset, which means that the dataset is carefully labeled and contains only elements of normal behavior. In other words, no attack is mislabeled in this dataset. We call such a dataset a refined dataset and we refer to as "Training1".

As shown in Figure 3, after the first learning iteration, the obtained classifier is tested by two sets called "Reference-Test" and "Test1". The "Reference-Test" will not be changed during all iterations of the learning phase, and this in order to measure the improvement of the classifier using the same test set. The second set "Test1" will be replaced within each new iteration, by another test set. Then, feedback of the evaluation will be used to refine the training dataset, and the new learning will be processed. Note that the learning in the next iteration will be faster because we use the incremental SVDD.

To evaluate a classifier we usually compute 4 values, detected normal behaviors, detected attacks, missed attacks and finally false alerts. All of these measures can be used to improve the training dataset. However, an IDS report only alerts which can be real attacks or false alerts. Therefore, using normal behavior and missed attack is difficult or impossible to achieve in practice. It would be very expensive for a security operator to recover missed attacks. This will imply to manually check all log files and the network traffic looking for malicious activities.

Now, using detected attacks and false alerts is possible. In practice, the recovery of these data can be easily done by analyzing alerts generated by the classifier which is periodically done by security operators. Revising the training dataset by detected attacks theoretically brings nothing to improve the classifier because this latter are already outside the normal behaviors boundary. In other words, no additional information will be reported[1].

Finally we conclude that only false alerts can be used to improve the classifier. Indeed, these data are easily provided when alerts are analyzed by the security operator. On the other hand, these data add to the classifier a new important information since it makes a wrong decision during the test. For that, after evaluating the initial classifier using "Test1", false alerts will be labeled as normal behavior and added to "Training1", and thus we obtain a new training datasets called "Training2". We then build a new classifier using "Training2" and we test again the obtained classifier using a new test set "Test2", and also using the "Reference-Test".

This procedure is repeated each time a security operator evaluates the reported alerts. Now, the improvement of the classifier can be checked by comparing the results of testing "Reference-Test" on all iterations.

## Using a Compromised Training Dataset

This case study is similar to the first, however the training dataset is not refined. In fact, it is rare to have a training dataset with no labeling errors. That is why we have chosen a small set of attacks, we labeled them as normal and we add them to the training dataset. The objective is to measure the robustness of our approach regarding mislabeling, which is common in practice.

This case study will add more realism to our adaptive approach, since it includes the possibility of mislabeling of normal behavior. After some experimen-

[1]In the experiments of the next section we will see that adding also confirmed attacks will improve the classifier.

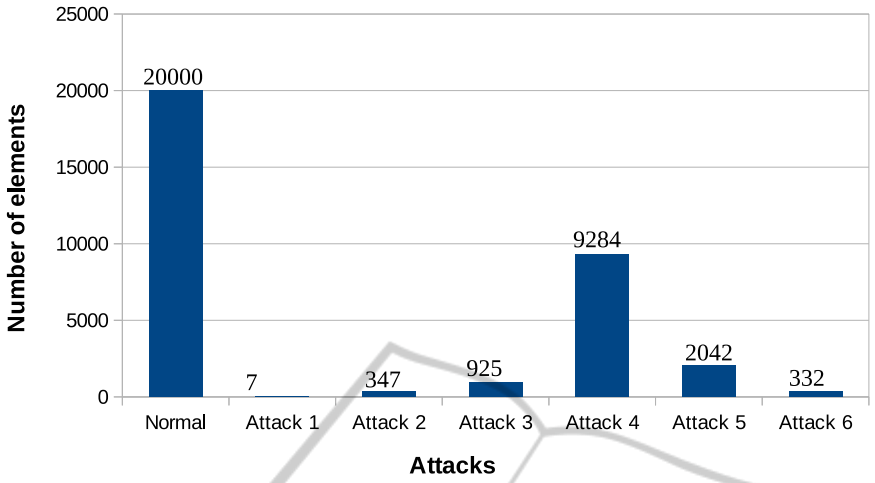


Figure 4: Attacks distribution.

tal tests, carried out in the next section, we will confirm the validity of this approach in realistic conditions, namely when training dataset contains inaccurate measurements.

## 5 EXPERIMENTS

In this experiment we use a benchmark that we generated using our attack simulation framework. We simulated traffic for normal behavior and for 6 attacks, which are presented in Table 1. Figure 4 gives attacks distribution of the benchmark.

Table 1: Attacks description.

| Attacks | Description |
|---|---|
| Attack 1 | IP and port scanning |
| Attack 2 | SMTP user enumerating |
| Attack 3 | IRC Daemon backdoor exploit |
| Attack 4 | BailiWicked Domain Attack (CVE-2008-1447) |
| Attack 5 | Windows smb client exploit (ms10-006) |
| Attack 6 | Sending emails with compromised attached files |

Now, as a refined "Training1" dataset, we select a subset from our benchmark which contains 20000 normal objects and 0 attacks. And as "Reference-Test" we select an other subset which contains 143392 normal objects and 2784 attacks. Note that both "Training1" and "Reference-Test" datasets are randomly selected from the benchmark.

Then, as explained in section 4 the "Reference-Test" dataset is divided into 16 subsets to be used for the successive iterations. Each subset contains 8962 normal objects and 174 attacks. After each iteration we extract false alerts and detected attacks and we add them to the training dataset. Note that each test subset will be tested by the classic SVDD and by the proposed approaches, namely the A-SVDD.

A comparison of these 2 approaches in terms of detection rate, false positive rate and the Percentage of Correct Classification (PCC) are given in Figure 5.

In the Figure 5 (a) we can see that the A-SVDD considerably enhance the accuracy of the SVDD. Indeed, as we can see in Figure 5 (b), the A-SVDD reduce much better the false positive. Moreover, in the Figure 5 (c) we see that the detection rate of the classic SVDD and the A-SVDD is exactly the same, which means that A-SVDD does not induce any deterioration in the ability to detect attacks. Thus, we conclude that the A-SVDD is better because, it reduces the false alerts rate over the time, while it keeps unchanged the detection rate.

This can be explained as follows. The SVDD require only two parameters the rejection rate of the target class and the rejection rate of the outliers[2].

In the A-SVDD, after each iteration we add some normal objects to the target class and we try to find an other hypersphere which may causes that some outliers (i.e attacks) will be inside the new hypersphere. This may deteriorate the detection rate. For that, we add also confirmed attacks and we set to zero the rejection rate of the outliers which means that no attacks already learned are tolerated to be inside the new hypersphere. So, the A-SVDD will try to find a new hypersphere containing all normal objects (older and

[2]The rejection rate of the target (resp. outliers) is the percentage of objects tolerated to be outside (resp. inside) the hypersphere.

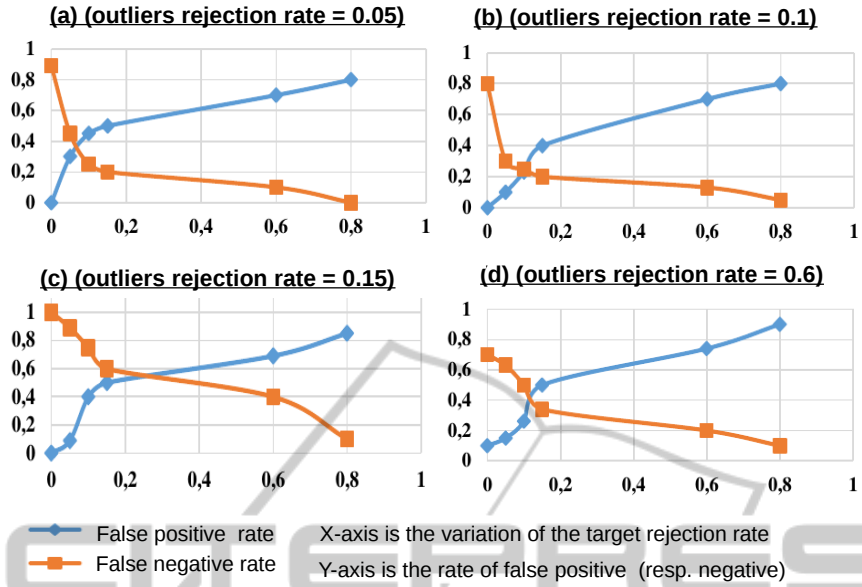


Figure 6: ROC curves of the A-SVDD using a refined training dataset and tested on "Reference-Test".

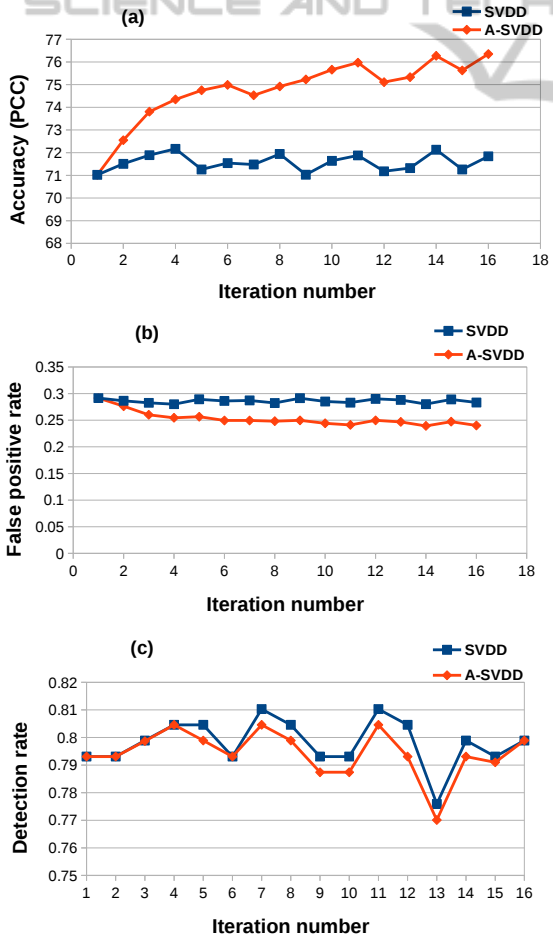


Figure 5: Detection rate, false positive rate and accuracy using a refined training dataset.

new ones) while keeping outside all already learned attacks.

Now, the question is how we can choose the appropriate value for these rejection rates. For instance, this can be done only experimentally by setting one rate and varying the other. For each pair (target rejection rate, outliers rejection rate), detection rate and false positive rate are computed to draw a ROC curve as given in Figure 6. Intuitively, one might think that setting these rates to (0,0) will give the better results. This is false because target objects and outliers are not uniformly distributed, many objects of the class "Target" and "outliers" must be rejected to ensure a good trade-off between detection rate and false alerts.

4 ROC curves are given in Figure 6 which allow us to consider four optimal points. Recall that the intersection between the false positive and the false negative represent the optimum point in a ROC curve. Figure 6 (b) gives as the best optimum which corresponds to a rejection rate of 10% for the target and 10% for the attacks. These values are used in the experiments presented above.

## Using a Compromised Training Dataset

In this experiment we intentionally add some attacks to the training dataset, namely 1000 attacks are labeled as normal. This makes our experiment more realistic because mislabeling is common in intrusion detection. We repeat this experiment 16 times as in the case of refined training and results are given in Figure 7. Note that the test dataset is the same.

Figure 7 shows that even using a compromised

Table 2: Comparison with other classifiers.

| | Detection rate | False positive rate | Accuracy |
|---|---|---|---|
| SVM | 0,216 | 0,082 | 69,40 |
| SVDD | 0,78 | 0,33 | 70,77 |
| A-SVDD | 0,78 | 0,20 | 76,83 |
| Decision tree | 0,012 | 0,0001 | 68,50 |
| Naive bayes | 0,047 | 0,003 | 69,41 |

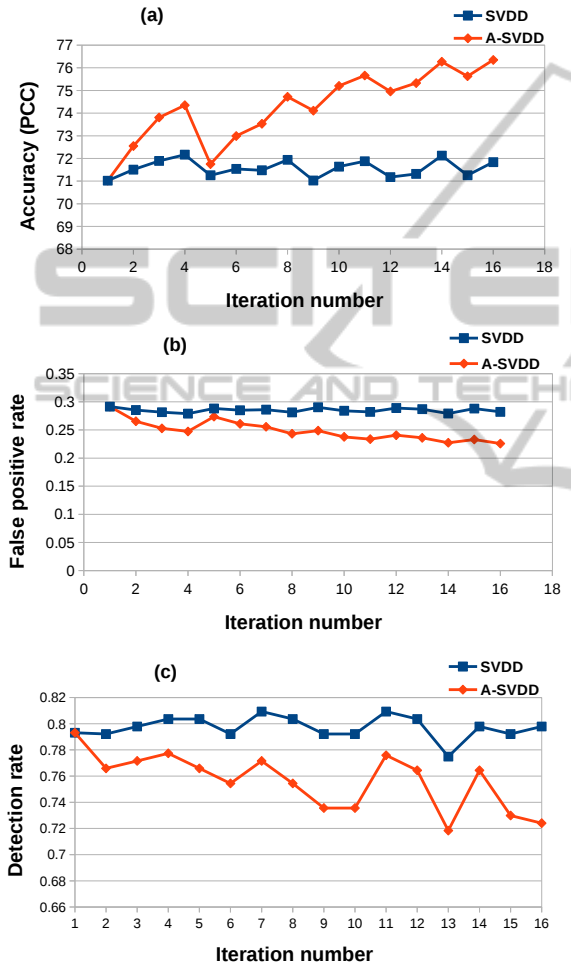


Figure 7: Detection rate, false positive rate and accuracy using a compromised training dataset.

training, the A-SVDD give a good results as in the case of a refined training dataset. This means that A-SVDD is able to reduce the rate of false alerts while keeping slightly unchanged the detection rate. The advantage of this new experiment is to say that the A-SVDD can be used directly to a collected training dataset without any additional work to separate normal traffic from attacks. This will make the learning of an anomaly intrusion classifier more easily and more efficient.

## Comparison with other Classifiers

In this subsection we compare the A-SVDD with some other classifiers, namely the SVM, decision trees and naive Bayesian networks. We will use in the learning of these classifiers the refined dataset described above, to which we add a single outliers object (one attack). We do that because the other classifiers are multi-class and need objects in all classes to be able to build a classifier. Then, we test the different classifiers using "Reference_Test". The results are shown in table 2.

The results in table 2 confirm the failure of the multi-class methods (binary in this case) for intrusion detection, this domain is more appropriate for single class. Indeed, while the accuracy (computed using the PCC) of these classifiers is of the order of 69%, the rate of detection is very low and this because we have provided only one element of the class "Intruder" in the learning.

This experiments highlight, with more recent and consistent data, in the field of intrusion detection, that the mono-class classification is more appropriate, especially with an adaptive learning as explained in section 4.

# 6 CONCLUSION

In this paper, an adaptive SVDD-based learning approach for the anomaly detection is presented. This new approach consists in continuously enhancing the performance of the learned SVDD classifier by refining the training dataset. This requires that after every evaluation of the classifier by a security operator, feedback concerning false alerts and confirmed attacks will be re-injected into the training dataset in order to reduce the false alerts rate in the next iteration.

This approach has been tested on a refined training dataset and on a compromised dataset. Experimental results confirm that the adaptive SVDD allows a significant reduction of the false alarms rate, while keeping unchanged the detection rate. In fact, reducing false alarms is due to the injection of false alarms

to the training dataset and the ability of the SVDD to do an incremental learning. However, to keep unchanged the detection rate, we injected also confirmed detected attacks and we force the outliers rejection rate to zero, so the new constructed hypersphere will contain new normal objects, while keeping outside all already known attacks.

Another important point is that these tests have confirmed that the use of multi-class classifiers (especially binary classifiers) is not recommended in the field of intrusion detection with the main hypothesis of an abundant data on the normal class and a small data of the class of attacks.

Finally, we state that this approach can be adopted to reduce false alerts without affecting the detection rate. Moreover, this approach can be applied directly to the training dataset even if it contains mislabeling, which will facilitate the work to security operators.

In future work, we are interested in using a most flexible form than hypersphere or hyperellipse, namely a kernel function. However, we have first to overcome some computational complexity due to the amount of data to be processed, in the domain of intrusion detection.

## REFERENCES


Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Desir, C. (2013). *Classification Automatique d'Images, Application à l'Imagerie du Poumon Profond*. PhD thesis, Université de Rouen.

Ghasemi Gol, M., Monsefi, R., and Yazdi, H. S. (2010). Intrusion detection by new data description method. In *Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on*, pages 1–5. IEEE.

Li, J. L. and Wang, B. Q. (2013). Detecting app-ddos attacks based on marking access and d-svdd. *Applied Mechanics and Materials*, 347:3734–3739.

Liao, S.-H., Chu, P.-H., and Hsiao, P.-Y. (2012). Data mining techniques and applications a decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12):11303 – 11311.

Liu, B., Yin, J., Xiao, Y., Cao, L., and Yu, P. S. (2010). Exploiting local data uncertainty to boost global outlier detection. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 304–313. IEEE.

Mazhelis, O. and Puuronen, S. (2007). A framework for behavior-based detection of user substitution in a mobile context. *computers & security*, 26(2):154–176.

Onoda, T. and Kiuchi, M. (2012). Analysis of intrusion detection in control system communication based on outlier detection with one-class classifiers. In *Proceedings of the 19th International Conference on Neural Information Processing - Volume Part V*, ICONIP'12, pages 275–282, Berlin, Heidelberg. Springer-Verlag.

Tax, D. M. and Duin, R. P. (2004). Support vector data description. *Machine learning*, 54(1):45–66.

Tax, D. M. and Laskov, P. (2003). Online svm learning: from classification to data description and back. In *Neural Networks for Signal Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on*, pages 499–508. IEEE.

Yu, J., Lee, H., Kim, M.-S., and Park, D. (2008). Traffic flooding attack detection with snmp mib using svm. *Computer Communications*, 31(17):4212–4219.