Improved Deep Belief Networks (IDBN) Dynamic Model-Based Detection and Mitigation for Targeted Attacks on Heavy-Duty Robots

Lianpeng Li, Lun Xie *, Weize Li, Zhenzong Liu and Zhiliang Wang

College of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China; llpstudy@163.com (L.L.); liweizemail@163.com (W.L.); s20160647@xs.ustb.edu.cn (Z.L.); wzl@ustb.edu.cn (Z.W.)

* Correspondence: xielun@ustb.edu.cn; Tel.: +86-010-6233-2873

Received: 31 March 2018; Accepted: 24 April 2018; Published: 26 April 2018

Featured Application: The research is mainly used for heavy-duty robots to detect and mitigate target attacks, which come from both the cyber-domain and the physical-domain.

Abstract: In recent years, the robots, especially heavy-duty robots, have become the hardest-hit areas for targeted attacks. These attacks come from both the cyber-domain and the physical-domain. In order to improve the security of heavy-duty robots, this paper proposes a detection and mitigation mechanism which based on improved deep belief networks (IDBN) and dynamic model. The detection mechanism consists of two parts: (1) IDBN security checks, which can detect targeted attacks from the cyber-domain; (2) Dynamic model and security detection, used to detect the targeted attacks which can possibly lead to a physical-domain damage. The mitigation mechanism was established on the base of the detection mechanism and could mitigate transient and discontinuous attacks. Moreover, a test platform was established to carry out the performance evaluation test for the proposed mechanism. The results show that, the detection accuracy for the attack of the cyber-domain of IDBN reaches 96.2%, and the detection accuracy for the attack of physical-domain control commands reaches 94%. The performance evaluation test has verified the reliability and high efficiency of the proposed detection and mitigation mechanism for heavy-duty robots.

Keywords: security; heavy-duty robots; IDBN; dynamic model; detection; mitigation

1. Introduction

In the global industry, heavy-duty robots play an irreplaceable in many fields, such as heavy equipment manufacturing, hoisting, and fine assembly. They have become an essential equipment for solving heavy-duty operation problems, which can improve work efficiency and reduce labor costs [1,2]. By 2017, there have been more than 500 thousand heavy-duty robots in the world, and the size of the heavy-duty robot market will keep growing in the few years [3].

However, the rapid development of heavy-duty robots and the corresponding development of intrusion detection are seriously unbalanced. There are two main reasons for the above situation:

- Compared with lightweight robots such as medical robots and service robots, the heavy-duty robots have lower level of intelligence, more complex application environments, and relatively concentrated functions, which lead to the lack of attention to the security issues for researchers;
- It is difficult to implement the tests of cyber-physical systems (CPSs) on heavy-duty robots due to the large inertia, high load, and other physical characteristics. Once an attack occurs, the damage will be great. This is the reason why we conducted tests in the laboratory environment.
These two aspects have mainly led to the slow development of security research on heavy-duty robots. The targeted attacks on heavy-duty robots have expanded from communication protocol attacks to attacks on control systems. Meanwhile, the damage caused by targeted attacks covers the physical-domain and the cyber-domain [4].

With the increase in the intelligence and functional requirements, heavy duty robots are inevitably communicating with the network, which increases the threats from the cyber-domain. Attacks on control commands do not cause changes in the cyber-domain, making it difficult for current algorithms to detect these attacks. However, the corresponding damage to physical-domain is enormous. Few methods are specializing in heavy-duty robots’ intrusion detection and mitigation mechanism. Therefore, considering the above-mentioned situation, the security problem of heavy-duty robots is particularly important and urgent.

This paper attempts to solve the security problem of heavy-duty robots through the IDBN and dynamic model detection system. The main contributions of this paper are summarized as follows:

- The system component and targeted attacks of heavy-duty robots are analyzed;
- The detection and mitigation mechanism based on IDBN and dynamic model are established; and,
- The performance evaluation test is carried out to test the performance of established mechanism.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 analyses the component and targeted attacks of heavy-duty robots. Section 4 gives the intrusion detection and mitigation mechanism based on the IDBN and dynamic model, including the improvement of the composition of DBN and the establishment of dynamic model. In Section 5, the test platform is built and the performance evaluation test is carried out on the test platform. Finally, Section 6 concludes this paper.

2. Related Work

For intrusion detection of robots, related issues has been broadly studied and analyzed. Lin et al. [5] proposed general principles for detecting cyber-physical attacks. It has combined the technology of the cyber-domain and the physical-domain to estimate the adverse consequences of malicious activities in a real-time manner. Degeler et al. [6] demonstrated a method of using Danger Theory to protect industrial processes in robotic manufacturing facilities from cyber-domain attacks based on artificial immune system paradigms. By utilizing leveraging physical dynamics of robots, Guo et al. [7] have developed a new robotic intrusion detection system (IDS), which can detect actuator attacks as well as sensor attacks for lightweight robots subject to random noises. Through the method of decision tree to generate simple detection rules, Vuong et al. [8] evaluated the small-scale remote control of robots against denial of service and command injection attacks. In addition, a novel framework for integrity analysis of robotic systems has been presented, whose output can be used to avoid the inherent design flaws in the system and reduce the damage that may be caused by undiscovered attackers [9]. In [10], Khaitan et al. discussed security issues at various levels of the robot architecture, risk assessment, and technical security issues. An intelligent intrusion detection system has been proposed based on Integrated Circuit Metrics, which has significant defense capabilities against sudden attacks [11]. However, the above researches are mainly focused on lightweight robots, and there are few studies on heavy-duty robots. They are barely applicable for heavy-duty robots. This is because, firstly, the pertinence is insufficient. Secondly, heavy-duty robots have a large difference between lightweight robots in terms of driver, control, and physical hardware, such as the rigid structure, communication mode, and processor performance, which brings different attack effects and different mitigation strategies [12]. These differences have made it impossible for us to copy the intrusion detection methods of the lightweight robot.

As for algorithms, the commonly used intrusion detection algorithms are focused on intelligent algorithms, such as machine learning (ML). Loukas et al. [13] proposed a real-time detection system that used lightweight statistical learning techniques, which used approaches of much greater complexity
and detection strength based on deep learning to improve detection accuracy and save energy. Bezemskij et al. [14] proposed a method based on Bayesian networks, which could not only determine whether a robot was attacked, but also determine whether the attack came from the cyber-domain or the physical-domain. Maleh et al. [15] maked use of a support vector machine (SVM) algorithm to detect the presence of malicious behavior through a set of signature rule anomaly detection and provided global lightweight IDS. Subsequently, Jones et al. [16] proposed a two-level intrusion detection system, consisting of a signature detection component and an anomaly detection component, where the anomaly detection component trained through a deep neural network (DNN) to detect commands that deviated from expected behaviors. In view of the defects in the detection rate and convergence speed of the traditional back propagation (BP) neural network intrusion detection model, the improved PSO-BP neural network was used in the IDS model [17]. And, more advanced detection methods have been obtained by combining or integrating evolutionary algorithms and neural networks, which have shown better detection performance than general machine learning methods [18]. Guerrero et al. [19] studied that systems built using supervised learning could detect real-time positioning system attacks. Furthermore, Goldman et al. [20] introduced an automated process of active detection cyber-domain attacks based on theoretical ideas from decision theory and recent research results in neuroscience. Beton et al. [21] introduced an idea borrowed from computer vision and neuroscience to reduce CPS security effects. This idea was called active awareness, which proxy allocated computational and sensing resources to approximately optimize information value. The intrusion detection algorithms proposed in the above literatures have abundant theoretical researches, but the limitation is obvious. That is, they are mainly used to detect intrusions. No mitigation mechanism has been given, and the pertinence is weak. The security problems faced by heavy-duty robots cannot be better solved. Literature [22] has proposed mitigation mechanisms, but the targets of the study are lightweight robots.

In summary, although the intrusion detection methods for robots are continuously enriched, these researches are focused on lightweight robots such as medical robots and service robots. Furthermore, the existing research methods rarely provide mitigation mechanisms, and most of them adopt the emergency shutdown strategy to mitigate attacks impacts after detecting an intrusion. Due to the physical characteristics of heavy load and rigid structure of the heavy-duty robots, emergency shutdown usually causes more damage than cyber-physical attacks. The above situation motivates this study, which is aimed at improving the intrusion detection, mitigation, and the safe operation when an attack occurs in heavy-duty robots.

3. Heavy-Duty Robot Systems

In this section, we mainly analyze the component of the heavy-duty robot system and targeted attacks.

3.1. System Component

The heavy-duty robot system is an intelligent control system for processing and assembling. As shown in Figure 1a, the system consists of a heavy-duty robot body, a control system, and a human-robot interaction (HRI) control terminal. The heavy-duty robot body is composed of manipulator, motors and sensors. Manipulator is the execution unit of the system and motors are the power unit. These sensors consist of displacement sensor, vertical gyroscope, force sensor and laser radar, with the function of giving the feedback information of the joint position, force, torque and so on. HRI control terminal is the interaction unit between the heavy-duty robot and operators. It is responsible for the following: (1) The terminal receives operation commands from the operator and transmits the commands to the cloud server for data processing; (2) The feedback information of the system is given by the terminal to the operator.
Figure 1. (a) The heavy-duty robot system component; (b) Control system.

Figure 1b shows the control system of heavy-duty robot, mainly including control software and control hardware. The control system runs in a Linux environment and uses a ROS operation platform. The heavy-duty robot uses a hierarchical control system of the real-time kernel, which puts data processing on the cloud with the advantage of recording the operation logs conveniently and improving the effectiveness of the control system. The controller area network (CAN) bus adopts a ring-shaped redundant topology to realize the system’s multi-loop control. Based on the kinematics and inverse dynamics of the joint position controller and the joint torque controller, the control system can obtain the information about the desired joint angle, position and torque of each joint according to the commands, which make up of desired terminal joint position and trajectory from HRI control terminal. The motor controller converts the calculated desired angle, position, and torque of each joint into a corresponding motor drive command, which is transmitted from the control system via the CAN-bus to the motor driver. Subsequently, the motor driver adjusts the motor’s torque and rotation speed to drive the manipulator. In this process, the cloud server is responsible for the realization of software algorithms and data calculation. Finally, the information of force and position is fed back to the operator through the HRI terminal, and then the HRI terminal is waiting for the next operation commands.

3.2. Targeted Attacks

Targeted attacks on heavy-duty robots mainly make use of the vulnerabilities of communication protocol and operating system [23]. Once an attack occurs, it will bring serious security challenges to life and property of the heavy-duty robot system. As shown in Figure 1b, data processing is completed in the cloud server. The influence of this structure has two aspects. On the one hand, it brings about the improvement of data processing performance. On the other hand, it may cause the heavy-duty robot to face more attacks from the cyber-domain. What’s more, attacks on control commands can make heavy-duty robots vulnerable to physical-domain attacks. According to the heavy-duty robot system, some potential security issues for targeted attacks are proposed.

- Communication protocol: The attacker injects error logic or purposeful attack commands based on existing protocol vulnerabilities of heavy-duty robots.
- Traffic flow: The attacker continues to access to the cloud server with the purpose of occupying computing resources, making the server incapable of processing the data from the control system.
• System working state parameters: The attacker attacks the signal feedback unit in order to deceive the operator about the operation status of the heavy-duty robot system.

• Model parameters of control system: The attacker invades the control system and modifies the parameters of established model.

The above several potential security threats for targeted attacks come from both the physical-domain and the cyber-domain. Without dedicated effective intrusion detection and mitigation mechanisms, these potential issues may pose great threats to human-machine security, especially for attacks on control commands. Once accidental error logic or purposeful attack commands, such as motor torque commands, are injected into the control system, may cause the heavy-duty robot to move unexpectedly or jumping and other damage to all equipment [24]. One of the main reasons is that the injected intrusion command does not change the protocol length and transmission on the CAN-bus, nor does it cause changes in network traffic, making the existing intrusion detection algorithms difficult to detect. Therefore, it is necessary to propose a specialized intrusion detection algorithm for heavy-duty robots and establish a corresponding mitigation mechanism to reduce the adverse effects of targeted attacks.

4. Detection and Mitigation Mechanism

Combining with powerful feature extraction capabilities and data analysis capabilities of DBN, IDBN security checks are mainly used to detect targeted attacks from the cyber-domain; The dynamic model and security detection are used to simulate physical actions, according to control commands and ensure the security of the heavy-duty robot under the condition that IDBN security checks has not detected any intrusions. Meanwhile, the mitigation mechanism is used to ensure the cyber-physical security of the heavy-duty robot.

4.1. IDBN

DBN is a widely used intelligent classification algorithm, which has exhibited a good performance in intrusion detection of cyber-domain [25]. The data processing of the heavy-duty robot takes place in the cloud server, while the data is exchanged between control system and cloud server in real time. It is necessary to carry out intrusion detection against targeted attacks from the cyber-domain.

Traditional DBN has a complicated data classification step and a slow learning process, which lead to the low efficiency of the classifier. Furthermore, it also reduces the performance of the intrusion detection mechanism. Therefore, in order to meet the requirements for intrusion detection of heavy-duty robots, increasing the detection speed and performance, it’s essential for the traditional DBN to be optimized and improved. The improvement for DBN is mainly reflected in the data processing and DBN model design.

4.1.1. Data Processing

Appropriate data processing can better describe the laws among the data. Through the study of intrusion detection by related researchers [26], the performance of the data classifier mainly depends on the selected method and the manner of data processing.

Figure 2 shows the DBN improvement process based on the NSL-KDD dataset. The NSL-KDD dataset that provided by the Canadian institute for cybersecurity is used as the intrusion detection dataset. In this paper, the 41 dimensional data characteristics are divided into the nominal type and the numeric type (Table A1). The probability mass function (PMF) encoding of the nominal type and the normalization processing of the numeric type are carried out, respectively.
The NSL-KDD dataset comes from the information flow of the network. Each piece of the data contains multiple information characteristics, which include the basic numeric type as well as the nominal type. The dataset is formed by the data and data characteristics. It is used for the cloud server to train the classifier and detect abnormal data. In the actual process of intrusion detection, the input of training data by classifiers usually belongs to the numerical type [27]. However, the data of nominal type is crucial to the performance of the classifier. For example, the field protocol type and service type is necessary to train the classifier and detect abnormal data. In the actual process of intrusion detection, the input data, which belongs to the nominal type, contains multiple information characteristics, which include the basic numeric type as well as the nominal type.

For a dataset with \( n \) characteristics, it is mapped into \( M \) feature vectors, each of which is denoted as \( x = (x_1, x_2, \cdots, x_n) \). Assuming that \( x_j = (x_{1j}, x_{2j}, \cdots, x_{Mj}) (j \in \{1,2,\cdots,n\}) \) represents the \( j \)th feature of each piece of data, which belongs to the nominal type. \( x_j \) contains \( K \) kinds of nominal values \( nom_{1j}, nom_{2j}, \cdots, nom_{kj} \). Letting \( r_{kj} \in N \) be the number of occurrences of the value \( nom_{kj} \) in \( x_j \), and it can be obtained:

\[
r_{kj} = | \{ i \in N \mid x_{ij} = nom_{kj}, i = 1,2,\cdots,M \} |, \quad k = 1,2,\cdots,N
\]

(1)

According to Equation (1), the frequency value \( f_{kj} \) of \( nom_{kj} \) appearing in \( x_j \) can be expressed as:

\[
f_{kj} = \frac{r_{kj}}{M}
\]

(2)

and

\[
\sum_{k=1}^{k} r_{kj} = M, \quad 0 \leq r_{kj}/M \leq 1.
\]

(3)

For the nominal type data \( x_{kj} \) in the dataset, literature [28,29] encode it with the digital coding method. The corresponding field is coded as 0, 1, \( \cdots \), \( N \) according to the number of possible values for the field. This process can achieve the conversion of nominal type to numeric type, but it is also
necessary to normalize the converted data again. Moreover, in [30, 31], the method of binary encoding is used. Although the converted data are all in \([0, 1]\), this method has an obvious disadvantage. That is, it will seriously increase the dimension of the original data, resulting in the demand for multiple layers, which leads to a complex DBN model structure needed to extract effective characteristics of the classification.

To solve the above-mentioned problems, this paper converts the nominal type \(x_{kj}\) into the numeric type \(x_{kj} = f_{kj}\) through the calculation method of PMF. The dimension of the converted values has not changed. In addition, it can guarantee that all converted values are all in \([0, 1]\), which is equivalent to completing the two operation steps of type conversion and data normalization at the same time.

In a dataset, the magnitude of data is often not in the same order [32]. This leads to a decrease in the rate at which the gradient descent to find the optimal solution when learning the data laws. It is also possible that heavy-duty robots cannot complete data processing and detection due to the slow convergence or even inability to converge, which may affect the classification accuracy. Therefore, the data needs to be normalized to \([0, 1]\).

Assuming that the selected dataset contains a total of \(N\) samples, each feature attribute column of all samples can be mapped to \(x = (x_1, x_2, \cdots, x_n)^T\). If \(x_i\) is the attribute value of the \(i\)th sample corresponding to the numeric data, then, we can use the method of Statistical normalization to make the processed data converge to \([0, 1]\). The specific form is as follows:

\[
f(x_i) = \frac{|x_i - \mu|}{\sigma}, \quad (4)
\]

where, the value of \(\mu\) is the average value of all \(x\) values, and \(\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}\) is the standard deviation of \(x\).

4.1.2. Intrusion Detection Based on IDBN

The intrusion detection based on IDBN makes use of the DBN model to detect unknown targeted attack types of heavy-duty robots. The DBN model includes multi-layer RBM for feature extraction and the contrastive divergence (CD) learning method to repeatedly optimize the network weights. It can achieve good learning ability and adaptability to detect unknown samples by training known samples.

In the DBN model, the processed data are used as the input data of the first visible layer of the RBM. The CD algorithm is used to carry out the layer-by-layer training on the RBM. Meanwhile, the RBM output values of the previous layer are used as the input values of next layer until the completion of multilayer RBM training [33].

The process of IDBN training and detection is as follows:

- A fixed amount of training data \(\text{min\_batch}\) was sampled randomly and inputted to the DBN at each time;
- The network weight would be updated once for each \(\text{min\_batch}\) number of training. This process would continue until all samples have been trained;
- After the multi-layer RBM training was completed, the BP neural network fine-tuned the parameters obtained by training the RBM through the back-propagation process;
- Through the repeated fine-tuning, the optimal value of related parameters could be obtained. These parameters include the connection weight matrix \(w\), the bias of visible unit \(a\) and the bias of hidden unit \(b\);
- Test data and corresponding attribute labels were entered in the trained DBN;
- The actual classification result of each test data was obtained by forward propagation calculation;
- The result was compared with the input attribute label to obtain the correct detection rate of the test sample.

During this process, the data dimension of the test sample was the same as the training data.
4.1.3. Model Structure and Parameters Selection

The determination of DBN structure for a particular system has not yet been fully supported by theory. It is necessary to confirm the relatively superior structure and parameters through the verification of theories and simulation experiments [34]. In order to determine the model structure and parameters selection of the proposed IDBN, a simulation experiment was conducted based on the NSL-KDD dataset. The NSL-KDD dataset contains 25,192 instances of training data. Each data in the dataset consists of an attribute label and 41 attribute characteristics. The attribute label divides the data into 4 types of attacks and 1 type of normal data. Moreover, the 4 types of attack data are divided into 39 subtypes. Furthermore, the 41 attribute characteristics are divided into 3 nominal types and 38 numeric types [35].

The normalized training set and test set were obtained by adopting the methods that the data of nominal type is encoded by PMF and the data of numeric type is normalized. Since the dataset had 41 data characteristics and had divided into 1 type of normal data and 4 types of abnormal data, the IDBN input data was 41-dimensional and the output data was 5-dimensional.

Considering the efficiency problem of intrusion detection, the number of DBN hidden layers was limited to 6 layers, which may sacrifice parts of the detection accuracy. The number of nodes in each hidden layer was the same and can be selected in [10, 20, 40, 60, 80, 100]. According to [36] and the iterative cross experiments, the other parameters of the IDBN model were set as follows: the learning rate in the pre-training and fine-tuning stages was set to 0.04; the number of iterations in the pre-training phase was 5, and after 20 iterations in fine-tuning stage, the results were stable in the area. The number of nodes and hidden layers were determined through simulation experiments.

As shown in Figure 3a, the model obtains a relatively high detection rate in the detection process when the number of hidden layers is 2. In addition, increasing the depth of DBN is not a positive correlation with the ability of feature extraction. Instead, it may lead to the decrease of generalization ability and the problem of overfitting.

In the condition where the number of hidden layers is fixed at 2 and the number of nodes in each hidden layer is changing, the change trend of the correct detection rate is shown in Figure 3b. When the number of nodes is 10, it is not enough to extract the feature set, which is suitable for classification, because of the less connected nodes to each other. Meanwhile, when the number of nodes is too large, there will be a problem of overfitting. Therefore, the correct detection rate is relatively highest when the number of hidden layer nodes is set to 20. Then, the design of IDBN is completed.

Figure 3. (a) The correct detection rate changes with the number of hidden layers (20 nodes); (b) The correct detection rate varies with the number of hidden layer nodes (2 hidden layers nodes).
4.2. Dynamic Model

The execution unit of the heavy-duty robot is a manipulator. The dynamic model for motors has been detailedly described in [37–39], and this paper will focus on the derivation of the manipulator dynamic model. The conventional dynamic modelling methods of manipulator usually adopt Newton-Euler method, Kane method, and Houston method [40]. These modelling methods are cumbersome and require large computational workloads, which are not suitable for the application requirements of the dynamic model in this paper. Aiming at the manipulator, this paper adopts the equivalent finite element method. The core idea of this method is to replace the real components in the system with equivalent units, and the equivalent system consisting of equivalent units replaces the actual heavy-duty robot system.

As shown in Figure 4, the assumed modal method is adopted. The point O at any position on the boom is the pedestal point of the manipulator, and the seven local coordinate systems formed by the ideally constrained mass points $O_1$, $O_2$, $O_3$, $O_4$, $O_5$, $O_6$, $O_7$ constitute the generalized coordinate. $\eta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]^T$ is the generalized coordinate vector. $\eta$ is the generalized velocity vector, and $l_1$, $l_2$, $l_3$, $l_4$, $l_5$, $l_6$, $l_7$ are the length of each joint.

![The 7-DOF manipulator coordinate diagram.](image)

The manipulator can be viewed as Euler-Bemoulli beam [41]. Then the expression of Lagrange equation for the dynamic equation is:

$$\frac{d}{dt}\left( \frac{\partial(T - U)}{\partial\dot{q}} \right) - \frac{\partial(T - U)}{\partial q} = Q,$$

where, $T$ is the kinetic energy of the manipulator joints, $U$ is the potential energy of the manipulator joints, $Q$ is the generalized force, $q$ is the generalized coordinate vector, and $\dot{q}$ is generalized coordinate velocity vector.

The kinetic energy of the manipulator joints $T$ can be expressed as follows:

$$T = \frac{1}{2} \int \rho v^2dl = \frac{1}{2} \dot{q}^T M(q)\dot{q}$$

where, $\rho$ is the density of joints, $v$ is the speed of any point on the joint, $l$ is the length of the corresponding joint, and $M(q)$ is the mass matrix of manipulator.

The potential energy of the manipulator joint includes the gravitational potential energy and the elastic potential energy generated by deformation:

$$U = U_l + U_{l1} = U_l + \frac{1}{2} \sum_{i=1}^{7} E_i I_i \int \frac{\partial^2 q_i}{\partial x^2} dx = U_l + \frac{1}{2} \dot{q}^T K\dot{q},$$

Figure 4. The 7-DOF manipulator coordinate diagram.
where, $U_i$ is the joints’ gravitational potential energy, $U_{ii}$ is the joints’ elastic potential energy, $E_i$ is the elastic modulus of the corresponding joint, $I_i$ is the moment of inertia of the joint to the z-axis, and $\varphi_i$ is the elastic deformation of the corresponding joint.

Substituting Equations (6) and (7) into the Lagrange equation:

$$M\ddot{\eta} + \dot{M}\ddot{\eta} - \frac{1}{2}\left( \frac{\partial M}{\partial \eta} \right)^T \ddot{\eta} + K\ddot{\eta} + \frac{\partial M}{\partial \eta} = Q_k$$

where, $Q_k$ is the generalized coordinate force.

Then, the manipulator dynamic equation can be described as:

$$
\begin{align*}
& \quad \left\{ \begin{array}{l}
m_{\theta\theta}\ddot{\theta} + m_{\theta\phi}q + v_\theta\dot{\theta}^2 + D_\theta\dot{\theta} = Q_{\theta} \\
m_{\phi\phi}\ddot{q} + m_{\theta\phi}q + v_\phi\dot{q}^2 + K_{\phi\phi}q = Q_q
\end{array} \right.
\end{align*}
$$

where, $m_{\theta\theta}$, $m_{\phi\phi}$ and $m_{\theta\phi}$ make up the mass matrix $M$, $K_{\phi\phi}$ is the stiffness matrix. $D$ and $v$ are the first-order and second-order coefficient matrices of speed.

According to the dynamic equation and the manipulator dynamic parameters given in Table 1, the manipulator dynamic model is established in Adams 2013, as shown in Figure 5.

**Table 1. Parameters of heavy-duty manipulator.**

<table>
<thead>
<tr>
<th>Joint</th>
<th>Mass $m_i$</th>
<th>Centroid Position $c_i/m$</th>
<th>Terminal Position $c_i/m$</th>
<th>Density $\text{kg/m}^3$</th>
<th>Elastic Modulus $\text{kN-mm}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>745</td>
<td>[0 0 0.21]</td>
<td>[0 0 0.4]</td>
<td>7800</td>
<td>200–235</td>
</tr>
<tr>
<td>2</td>
<td>910</td>
<td>[0.48 0 0.4]</td>
<td>[1.13 0 0.4]</td>
<td>7850</td>
<td>200–235</td>
</tr>
<tr>
<td>3</td>
<td>882</td>
<td>[1.26 0 0.4]</td>
<td>[1.54 0 0.4]</td>
<td>7650</td>
<td>200–235</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>[1.7 0 0.25]</td>
<td>[1.7 0 0.12]</td>
<td>7820</td>
<td>200–235</td>
</tr>
<tr>
<td>5</td>
<td>301</td>
<td>[1.83 0 0.12]</td>
<td>[1.92 0.12]</td>
<td>7810</td>
<td>200–235</td>
</tr>
<tr>
<td>6</td>
<td>279</td>
<td>[2.01 0 0.1]</td>
<td>[2.08 0 0.1]</td>
<td>7800</td>
<td>200–235</td>
</tr>
<tr>
<td>7</td>
<td>198</td>
<td>[2.27 0 0.1]</td>
<td>[2.5 0 0.1]</td>
<td>7800</td>
<td>200–235</td>
</tr>
</tbody>
</table>

**Figure 5.** (a) Heavy-duty manipulator; (b) Adams dynamic model.

In order to test the correctness and accuracy of the manipulator dynamic model, the dynamic simulation experiment was performed by defining the rotation drive and torque of the seven joints. The simulation adopted the basic mode and the simulation type selected default. Moreover, the simulation time was 16 s. In this simulation, we added the same drive function to joint 5 and joint 6. The force of each joint in heavy-duty manipulator could be obtained through the Adams post-processing module. As shown in Figure 6a, the forces of the three joints at the end of the manipulator is given. And the force curves of joint 5 and joint 6 are similar, which is consistent with our drive equations, demonstrating the correctness of the dynamic model. In addition, Figure 6b shows the follow-up of
dynamic model’s trajectory of the first three joints and the actual joint trajectory for the same control input based on a heavy-duty robot, where the blue line represents the manipulator and the red line represents the dynamic model. It can be seen from the figure that the established model and the actual trajectory are mostly consistent, ignoring the leading or lagging in some places, which proves the accuracy of the established model.

Figure 6. (a) The forces of three joints in Adams; (b) Trajectory following.

4.3. Detection and Mitigation Mechanism

Based on the establishment of the IDBN and the heavy-duty robot dynamic model, we established a detection and mitigation mechanism. Figure 7 shows the mechanism for intrusion detection and mitigation.

Figure 7. Detection and mitigation mechanism.

The manipulator is used to execute the control commands and complete tasks. Once the manipulator runs abnormally because of targeted attacks, the consequences are serious.
The HRI control terminal consists of Force interaction and Simulation action. It transmits the manipulator desired terminal position and trajectory to the control software unit. At the same time, it receives the feedback from the control system based on the dynamic model, while the desired action is pre-presented on the Simulation action.

The IDBN security checks is based on the IDBN intrusion detection algorithm to detect known attacks, such as dos, u2r, r2l, probe and other common attacks, and unknown attacks from cyber-domain. In order to eliminate the susceptibility of the sample statistic to the outliers and possible noises in the detection results, we set up the detection threshold to provide a trustworthy range of volatility [16]. Furthermore, we choose that the detection result of continuous twice attacks in a communication cycle is the sign to give the alarm, avoiding the above conditions and improving the continuity of operation.

The control software converts the desired terminal position and trajectory into drive commands for the motors through the joint controller and motor controller. The joint position controller is a proportional amplifier, and the output position is controlled by inverse kinematic feedback. The torque controller is a PI controller that calculates the torque of each joint based on inverse dynamics. And then, the motor controller converts the control command signals into motor torque and speed drive signals.

The dynamic model is modelled by the equivalent finite element method. It can simulate the physical system behavior of the manipulator according to the driving commands and display it on the Simulation action. The security detection is used to detect the analog output of dynamic model. Furthermore, this part is mainly for control commands, including joint position, joint angle and motor torque. The next joint information is estimated by the dynamic model based on the control commands. As for joint position, joint angle and motor torque, if two or three of the consecutive data reach the threshold, an intrusion alarm will be triggered. One of the most serious situations is that, if the manipulator joint jumps or stops unexpectedly in the dynamic model, the alarm mechanism will be triggered immediately. The dynamic model and security detection mainly focus on the control commands of error logic injection or purposeful attacks. They detect and mitigate targeted attacks before the physical system generates the attack effects.

The detection and mitigation mechanism is established based on IDBN and dynamic model. For cyber-domain attacks, IDBN security checks mainly detects the aforementioned four major types of attacks. Once the attack has been determined, the mitigation mechanism will take a series of measures in response to the type of intrusion, including giving an alarm, breaking the intruder’s connection, collecting intrusion information, reporting the results to the operator, and waiting for the operator’s response. If the existing mitigation mechanism cannot block the intrusion attack, it will stop sending drive commands to the motors and start the safe shutdown mechanism. At the same time, the new detected data will be brought into the IDBN training set to improve the ability of detecting unknown attacks. Security detection of attacks on the physical-domain is based on the dynamic model. It estimates the joint action changes that generated by the commands. The threshold is determined by the upper and lower limits of each parameter in 100 trouble-free operations of the heavy-duty robot. In order to reduce false alarms raised by model inaccuracies and natural noises, it is trustable in the range of 98–102% of the threshold. Meanwhile, the information of the joint position, joint angle and motor torque are estimated according to the dynamic model after the discovery of attacks. If the instruction mechanism has been detected five times in a row, the safety shutdown mechanism of the heavy-duty robot will be triggered in consideration of the safety of the robot. Subsequently, the motor drive system hasn’t receive commands from the control software anymore. According to the current position of the manipulator, the safety shutdown will be allowed within the allowable range of the impact, and the manipulator information will be fed back to the HRI control terminal in real time. The above measures are used to mitigate the adverse effects brought by targeted attacks before they occur, which can also ensure the safety of the human, robots and the continuous operation.
5. Performance Evaluation

5.1. IDBN Performance Evaluation

For the heavy-duty robot training dataset, there is a large number of redundant data in the NSL-KDD dataset, and there are also existing problems, such as data imbalance, which can’t fully reflect the ability of identifying unknown samples [42]. Therefore, it is necessary to preprocess the dataset. Table A2 shows the details of dataset preprocessing, and the training set and test set are obtained.

In this experiment, 20%, 40%, 60%, and 100% of the optimized NSL-KDD dataset were extracted and trained according to the SVM, BP, IDBN, and DBN algorithms. Finally, we obtained the accuracies of each algorithm with different data volumes and the corresponding training time.

As shown in Figure 8 and Table 2, it is found that the proposed IDBN has significantly improved the identification ability and detection efficiency of cyber-domain intrusion compared with the traditional DBN. As for the BP algorithm and SVM algorithm, IDBN has a prominent advantage in accuracy on the basis of similar time. In addition, we find that a high detection accuracy can be obtained by using a 40% of the NSL-KDD dataset, and there is also a great advantage in terms of efficiency. Therefore, IDBN security checks for the heavy-duty robot use the 40% of NSL-KDD dataset as the training set, not only improving the detection capability of the cyber-domain intrusion, but also having a good efficiency.

![Figure 8](image_url)

Figure 8. Accuracy rate with training data volume.

<table>
<thead>
<tr>
<th>Training Data</th>
<th>SVM</th>
<th>BP</th>
<th>IDBN</th>
<th>DBN</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>10.4s</td>
<td>0.11s</td>
<td>0.13s</td>
<td>0.24s</td>
</tr>
<tr>
<td>40%</td>
<td>11.7s</td>
<td>0.18s</td>
<td>0.20s</td>
<td>0.41s</td>
</tr>
<tr>
<td>60%</td>
<td>20.87s</td>
<td>0.25s</td>
<td>0.28s</td>
<td>0.58s</td>
</tr>
<tr>
<td>100%</td>
<td>32.30s</td>
<td>0.47s</td>
<td>0.52s</td>
<td>0.93s</td>
</tr>
</tbody>
</table>
5.2. Performance of Detection and Mitigation Mechanism

This subsection mainly tests the detection and mitigation mechanism based on the dynamic model and security detection. Considering the safety of the heavy-duty robot and the simplicity of the test, we established a test platform in the laboratory environment.

As shown in Figure 9, we use a self-developed 7-DOF manipulator to simulate the heavy-duty robot. The application operating system and communication bus are the same as the heavy-duty robot. The data processing is completed in the cloud server, which communicates with the robot controller via the Ethernet. An IPC is used to simulate attackers, while receiving processed data from the cloud and randomly injecting some malicious control commands to replace data from the cloud server. The simulation attacker here adopts a PC software written in C# to send control commands containing malicious instructions to detect the effectiveness of the intrusion detection and mitigation mechanism. In this process, there is no change in the transmission of the communication protocol and the volume of data. Furthermore, in order to improve the detection efficiency of the system, only the dynamic model of the first three joints of heavy-duty robots was modelled. This is reasonable for that the first three joints contribute most to the robot’s positions, while the other four joints mainly affecting the orientation of the manipulator [23].

The 10 frames of data are sent per second between the simulated attacker and robot controller. Of course, the frequency is adjustable. In order to test the performance of the detection and recovery mechanism, we did something based on the established test platform as follows:

Step 1: From the 5ths of communication establishment, the injected commands randomly changed 1–2 continuous data with the information of joint torque, rotation angle, and trajectory.
Step 2: From the 15ths, the aforementioned information in 3–4 continuous data was randomly changed.
Step 3: From the 25ths, the aforementioned information in 3–5 continuous data was randomly changed.
Step 4: The above steps were repeated in 15 sets of data. Meanwhile, the experimental phenomena and the number of alarms record were recorded by log records.

Finally, we got the results of the test. Figure 10a gives part of the attack data in the red line, and one group of data detection results is shown in Figure 10b. In the Figure 10b, 1 indicates that an attack is detected and the mitigation mechanism was triggered. 0 means that no attack is detected, that is, the data are trusted. There is no security alarm on the left side of the red line, while a security alarm occurs per second on the right side of the red line except the 23ths.
The statistics of the corresponding log records are shown in Table 3. We find that the log records are in accordance with the corresponding detection attack results. There is an undetected attack at 23.8 s, which means that the modified data (at least one of the data) are within the range of trustworthy thresholds. The analysis is consistent with the Table 3. In the 33.4–33.8 s, targeted attacks have been detected five times in a row, triggering a safe shutdown mechanism. During the test, there was no manipulator shaking or sudden braking occurred from the log records, which proved that the pre-estimation instructions of the dynamic model could eliminate transient and discontinuous attacks. When the attacks are continuous and can’t be relieved by the mitigation mechanism, the mitigation mechanism will ensure a safe shutdown.

![Image](https://via.placeholder.com/150)

**Figure 10.** (a) The partial test data; (b) The results of detection attacks.

<table>
<thead>
<tr>
<th>Attack Time (s)</th>
<th>Attack Number</th>
<th>Attack Time (s)</th>
<th>Attack Number</th>
<th>Attack Time (s)</th>
<th>Attack Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5–5.6</td>
<td>2</td>
<td>14.8–14.9</td>
<td>2</td>
<td>23.8–23.9</td>
<td>2</td>
</tr>
<tr>
<td>6.1–6.2</td>
<td>1</td>
<td>15.6–15.9</td>
<td>3</td>
<td>24.5–24.7</td>
<td>3</td>
</tr>
<tr>
<td>7.7–7.8</td>
<td>2</td>
<td>16.1–16.4</td>
<td>4</td>
<td>25.7–25.9</td>
<td>3</td>
</tr>
<tr>
<td>8.4–8.5</td>
<td>2</td>
<td>17.0–17.3</td>
<td>4</td>
<td>26.3–26.6</td>
<td>4</td>
</tr>
<tr>
<td>9.0</td>
<td>1</td>
<td>18.6–18.8</td>
<td>3</td>
<td>27.2–27.4</td>
<td>3</td>
</tr>
<tr>
<td>10.6–10.7</td>
<td>2</td>
<td>19.4–19.7</td>
<td>4</td>
<td>28.5–28.7</td>
<td>3</td>
</tr>
<tr>
<td>11.8–11.9</td>
<td>2</td>
<td>20.7–20.9</td>
<td>3</td>
<td>29.4–29.7</td>
<td>4</td>
</tr>
<tr>
<td>12.2–12.3</td>
<td>2</td>
<td>21.5–21.8</td>
<td>4</td>
<td>30.4–30.6</td>
<td>3</td>
</tr>
<tr>
<td>13.7–13.8</td>
<td>2</td>
<td>22.2–22.5</td>
<td>4</td>
<td>31.7–31.9</td>
<td>3</td>
</tr>
</tbody>
</table>

Then, we expanded the number of trials to 100 sets. We got that the detection accuracy is 94.0% and the true positive rate is 97.8%, which proved the effectiveness of the proposed intrusion detection and mitigation mechanism.

In summary, the detection and mitigation mechanism of heavy-duty robots based on IDBN and dynamic model can effectively mitigate targeted attacks, which come from cyber-domain and physical-domain. For the attacks of the cyber-domain, the IDBN detection accuracy reaches 96.2% under the premise of taking into account the efficiency. For the physical-domain control instruction attacks, the detection accuracy reaches 94%. The undetected attacks will not cause damage to the heavy-duty robots under limited attack instructions.

6. Conclusions

In this paper, the security status and targeted attacks for heavy-duty robots are analyzed. For targeted attacks, we establish the intrusion detection and mitigation mechanism based on IDBN and dynamic model. The IDBN is proposed by improving the traditional DBN for targeted attacks.
from the cyber-domain. Meanwhile, the specific improvement process and improved performance are described detailedly. In addition, the dynamic model of the heavy-duty robot is derived and the correctness and accuracy are verified by simulation. Finally, we set up the test platform in the laboratory environment, and carry out the performance evaluation test. The test results prove that the IDBN detection accuracy reaches 96.2% for the cyber-domain attacks, and the detection accuracy for the attack of physical-domain control commands reaches 94%. Moreover, the proposed detection mitigation mechanism can ensure the normal operation of heavy duty robots under transient and discontinuous attacks. The proposed detection mitigation mechanism is correct and effective.

It is important to note that there are still a few shortages in this paper: (1) This research has not been able to conduct operations and detect attacks in real time, which means that it is only applicable to the objects that carry out work in an operation planning manner; (2) The test is conducted in a laboratory environment. It’s necessary to conduct the performance evaluation on actual heavy-duty robots; and (3) The definitely robotic structures inside commercial buildings should be within our next research work. The above analyses should be further studied in the future.

Author Contributions: Lianpeng Li and Lun Xie conceived this study and wrote the paper; Weize Li analyzed the data; Zhenzong Liu and Zhiliang Wang established the test platform.

Acknowledgments: This work is supported in part by the National Key Research and Development Program under Grant 2017YFB1302104 and in part by the National Natural Science Foundation of China under Grants 61672093 and 61432004.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

- IDBN: Improved deep belief networks
- AI: Artificial intelligence
- IDS: Intrusion detection system
- CPSs: Cyber-physical systems
- ML: Machine learning
- DNN: Deep neural network
- BP: Back propagation
- PSO: Particle swarm optimization
- CAN: Controller area network
- HRI: Human-robot interaction
- PMF: Probability mass function
- CD: Contrastive divergence
- DOF: Degree of freedom
- PI: Proportion integral
- IPC: Industrial personal computer

Appendix A

<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal</td>
<td>Protocol_type, service, flag</td>
</tr>
<tr>
<td>numeric</td>
<td>duration, src_bytes, dst_bytes, land, wrong_fragment, urgent, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_creations, num_shells, num_access_files, num_outbound_cmds, is_host_login, is_guest_login, count, srv_count, serv_rate, srv_serror_rate, srv_rerror_rate, srv_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_rerror_rate, dst_host_srv_rate, st_host_srv_rate, dst_host_same_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_rerror_rate, dst_host_serror_rate, dst_host_srv_rerror_rate, dst_host_rerror_rate</td>
</tr>
</tbody>
</table>
Table A2. Dataset preprocessing.

<table>
<thead>
<tr>
<th>Attack Type</th>
<th>Attack Subtype</th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>normal</td>
<td>6000</td>
<td>4000</td>
</tr>
<tr>
<td>dos</td>
<td>neptune</td>
<td>4000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Smurf</td>
<td>3000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>back</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>apache2</td>
<td>0</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>teardrop</td>
<td>0</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td>processstable</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>probe</td>
<td>satan</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ipsweep</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>nmap</td>
<td>0</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>portswep</td>
<td>0</td>
<td>900</td>
</tr>
<tr>
<td>u2r</td>
<td>warezmaster</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>warezclient</td>
<td>700</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>snmpguess</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>guess_password</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>r2l</td>
<td>ps</td>
<td>150</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>loadmodule</td>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>perl</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>xterm</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>18450</td>
<td>8700</td>
</tr>
</tbody>
</table>

References


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).