Hindawi
Security and Communication Networks
Volume 2018, Article ID 8210614, 16 pages
https://doi.org/10.1155/2018/8210614

Review Article

Data Fusion for Network Intrusion Detection: A Review

Guoquan Li,[1] Zheng Yan ,[1,2] Yulong Fu ,[1] and Hanlu Chen[1]

[1]*State Key Laboratory of ISN, School of Cyber Engineering, Xidian University, Xi'an, China*
[2]*Department of Communications and Networking, Aalto University, Espoo, Finland*

Correspondence should be addressed to Yulong Fu; ylfu@xidian.edu.cn

Received 20 December 2017; Revised 26 March 2018; Accepted 11 April 2018; Published 15 May 2018

Academic Editor: Jesús Díaz-Verdejo

Rapid progress of networking technologies leads to an exponential growth in the number of unauthorized or malicious network actions. As a component of defense-in-depth, Network Intrusion Detection System (NIDS) has been expected to detect malicious behaviors. Currently, NIDSs are implemented by various classification techniques, but these techniques are not advanced enough to accurately detect complex or synthetic attacks, especially in the situation of facing massive high-dimensional data. Besides, the inherent defects of NIDSs, namely, high false alarm rate and low detection rate, have not been effectively solved. In order to solve these problems, data fusion (DF) has been applied into network intrusion detection and has achieved good results. However, the literature still lacks thorough analysis and evaluation on data fusion techniques in the field of intrusion detection. Therefore, it is necessary to conduct a comprehensive review on them. In this article, we focus on DF techniques for network intrusion detection and propose a specific definition to describe it. We review the recent advances of DF techniques and propose a series of criteria to compare their performance. Finally, based on the results of the literature review, a number of open issues and future research directions are proposed at the end of this work.

1. Introduction

Network Intrusion Detection System (NIDS) is a new generation of network security equipment following the traditional security measures such as firewall and data encryption [1], which has been rapidly developed in recent years. It successfully resists many attacks and malicious actions and is called the second line of defense in the Internet. However, in the current big data era, the large amount of traffic data makes NIDS face critical challenges. First, large amounts of high-dimensional data increase processing complexity and need huge computing and storage resources. Second, many redundant and unrelated data could adversely affect network security detection. Third, some new attacks are difficult to detect due to big data process and analytics. Besides, the inherent weakness of NIDSs, such as high false positives (FP) and high false negatives (FN), raises urgent requests on effective solutions. Data Fusion (DF), as a promising technology of big data, has been applied into the domain of network intrusion detection to overcome the above-mentioned challenges in recent years.

The concept of DF originated from the US Air Force project; the US Department of Defense first proposed a Joint Directors of Laboratories (JDL) DF model based on national defense monitoring needs in 1987 [2]. Subsequently, DF was gradually studied and applied in other fields, such as automatic control, image recognition, target detection, and cyber security, and many scholars have proposed definition of DF based on their own studies and researches [3]. In order to clearly show the role of DF technology in network intrusion detection, an expression of DF in the field of NIDS is presented in this article.

In general, DF can be applied into three layers according to where fusions are needed, namely, data layer, feature layer, and decision layer. The data layer is the lowest system layer, playing the role of processing and integrating raw network data; the feature layer is the middle layer, fusing and reducing features of the preprocessed data; the decision layer is the highest layer, fusing and combining the inferences or decisions of various processing units. In the field of NIDS, most researches of data fusion only focus on the feature layer and the decision layer. It is because the network data they

need to fuse comes from public datasets that have already been fused at the data layer. The use of DF technology at the feature level can greatly reduce the size of data processing, thereby enhancing the efficiency of NIDSs. Besides, useful and refined data generated by feature fusion can support decision-making and further improve the robustness and accuracy of the system. As for using of DF technology at the decision level, the decision fusion center fuses the decisions of multiple local detectors to obtain more accurate and reliable identifications of network behaviors.

Currently, a lot of research work has been carried out on DF for intrusion detection in order to improve the performance of NIDS. However, we found that the open datasets, the number of experimental data samples, and the fusion techniques used in many literatures are diverse. It is difficult to understand and analyze the strengths and weaknesses of different fusion techniques. Thus, it becomes essential to specify uniform criteria to evaluate them in view of a large number of references and give performance statistics of the current literature. This work is meaningful because it can make it easier for researchers and practitioners to understand the characteristics of the current DF techniques and methods.

In this article, we provide a thorough review on DF techniques in NIDS. We first describe DF for NIDS by representing the process and role of fusion for motivating this research work. We review existing DF techniques used in intrusion detection and propose evaluation criteria to analyze and compare the characteristics and performance of different fusion techniques. Besides, we simply analyze different open network datasets that can be used for testing intrusion detection techniques. Based on our review, we put forward current main challenges and point out promising research directions in this field.

The main contributions of this survey are listed as follows.

(1) We give a description of DF for NIDS in order to motivate related research in this field.

(2) We propose a number of evaluation criteria for evaluating fusion techniques for network intrusion detection.

(3) We further employ the proposed criteria to review the performance of different fusion techniques, which offers a good reference for scholars in the fields of network security and information fusion.

(4) We propose the challenges and promising research directions of DF for network intrusion detection based on our review.

The remainder of this article is organized as follows. Section 2 gives a brief introduction about the background knowledge of NIDS and DF. Several commonly used fusion techniques are elaborated in Section 3. Section 4 puts forward the evaluation criteria of data fusion techniques based on a large amount of literatures. The power of different fusion techniques is analyzed and compared in Section 5. In Section 6, the existing issues of DF are discussed, and some promising research directions are proposed. Section 7 summarizes the whole article.

2. Background Knowledge

In order to better understand this article, this section introduces some basic theory, including network intrusion detection and DF. Network intrusion detection is an old topic that has been repeatedly studied. We mainly present two kinds of intrusion detection techniques, anomaly-based and misuse-based, and explain their advantages and disadvantages, separately. As regards DF, we introduce it from its source, definitions, levels, and applications and put forward a general DF framework for intrusion detection to facilitate intuitive understanding.

2.1. Network Intrusion Detection. NIDS is a kind of network security scheme that can monitor the network transmission in real time and alert or take corresponding measures when detecting some behaviors that threaten network security. Actually, NIDS can be regarded as a pattern of recognition system that can distinguish malicious attacks from normal network behaviors. Intrusion detection technology plays an important role in the process of identifying malicious behaviors. The intrusion detection techniques based upon data mining generally fall into two categories: misuse detection and anomaly detection [4, 5]. The misuse-based detection, also called signature-based detection, is based on known attack signatures. It usually uses the well-known attack signatures to match and identify attacks. The advantages and disadvantages of the misuse-based detection are as follows [6].

(1) Advantages

(i) Fast and efficient detection of known attacks or specific attack tools.

(ii) Detecting attacks without generating an overwhelming number of false alarms.

(iii) Allowing system administrators, regardless of their security skills, to track their system security issues and run exception handlers.

(2) Disadvantages

(i) Hard to detect novel or unknown attacks.

(ii) Hard to detect the variants of known attacks.

Due to the efficient detection and low false positive rate (FPR), the misuse-based IDSs are widely used in commercial networks. Furthermore, much excellent open-source software has also been implemented, typically represented by Snort. The Snort IDS is one of the commonly used misuse-based NIDSs, which performs real-time traffic analysis, content searching, and content matching to discover attacks using preidentified attack signatures [7]. It is popular with many researchers because of its open source and adaptability to various platforms. In [1], Tian et al. fused the alerts through Snort to test the performance of their proposed detection fusion system.

Although the misuse-based detection is efficient, it can only detect known attacks and cannot detect novel or zero-day attacks [38]. To detect novel attacks, the anomaly-based NIDS have been proposed. In many related literatures,

most of the network behaviors acquired by researchers are normal, so NIDSs usually uses the anomaly-based detection techniques. Anomaly detection is a recognition model based on normal behaviors of the network connections. Any deviation from the established pattern of normal behaviors is considered to be a suspicious action. The anomaly detection seems to be able to detect all types of attacks, including unknown attacks. However, it indicates that some activities are suspicious but not malicious, resulting in high FP [39]. The advantages and disadvantages of the anomaly-based detection are as follows [6].

(1) Advantages

(i) It can detect novel or unknown attacks.
(ii) It Produces information that can in turn be used to define signatures for misuse detectors.

(2) Disadvantages:

(i) It requires extensive training data of network connections and behaviors.
(ii) FPR is not ideal.

The misuse-based detection is efficient in detecting known attacks but cannot detect novel attacks, while the anomaly-based detection can detect unknown attacks but usually has a high FPR. Therefore, NIDS used only one of these two which could be limited in performance and scope of application. To avoid the above defects, many hybrid approaches have been proposed, which combine the advantages of both misuse and anomaly detection [40]. Hybrid intrusion detection technology can be divided into three categories as follows.

(1) Anomaly-based detection on top of misuse-based detection
(2) Misuse-based detection followed on top of anomaly-based detection
(3) Misuse-based and anomaly-based detection in parallel

Zhang et al. [15] implemented a hybrid system through the following first approach. This hybrid system can be used to detect known intrusions in real time and to detect unknown intrusions offline. Generally, in the past two decades, NIDSs have been fully studied. Intrusion detection technologies continue to improve and update. The performance of NIDSs has been greatly optimized accordingly, but NIDSs still face many challenges. The use of DF technology in the field of NIDS is a very promising research direction, which holds great potential to deal with these challenges.

2.2. Data Fusion

2.2.1. Data Fusion Definition. The concept of DF first appeared and applied in the military field in the 1980s, with strong military characteristics, which was called "intelligence synthesis." Joint Directors of Laboratories (JDL) defines DF from the perspective of military applications as follows: DF is a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, complete and timely assessments of situations, threats, and their significance. Waltz and Llinas [41] supplemented and modified the above definition in their work, replaced the "position estimate" with the "state estimate," and added the detection function, which gave the definition: data fusion is a multilevel and multifaceted process and mainly completes the detection, integration, correlation, estimation, and combination of data from single and multiple data sources. Its purpose is to achieve an accurate estimate of the status and identity of the target and to make a complete and timely assessment of the situation and threats. Many other DF definitions are presented by some scholars based on their own researches and analysis. Although these definitions give us inspiration and guidance to some extent, they are not exhaustive in a particular area. A more specific expression of DF in the field of intrusion detection is beneficial to researchers within the field and motivates their own work. Therefore, based on these facts, we presented a specific description of DF in NIDS: "single source or multisource data collected from the network is preprocessed to obtain a uniform data format. More refining data of greater quality is obtained through feature fusion and association, which greatly improves the identification of malicious network behaviors. The initial decisions generated from multisource data are integrated in a decision fusion center to achieve more accurate and comprehensive inferences or decisions." This expression is based on network intrusion detection; the goal of DF is to improve efficiency, accuracy rate (ACC) and robustness while reducing FNR and FPR, saving computing resources of system. We believe that the proposed definition is helpful to practitioners and researchers in the field of intrusion detection.

2.2.2. Data Fusion Levels. The data fusion is mainly applied at three levels with respect to the processing stage of the fusion [42]. Normally, three main levels are discerned: data, feature, and decision. At different levels, the representation of information is different: the outputs of the data level fusion and the feature level fusion are the "states," "characteristics," and "attributes," and the outputs of the decision level fusion are "inferences" or "decisions." Different fusion techniques and methods are usually used in different levels to improve overall performance of data processing.

The brief description of fusion levels is shown as below.

(1) Data level fusion: it is also called low level fusion, which combines several different raw data sources to produce refined data that is expected to be more informative and synthetic.
(2) Feature level fusion: it combines many data features and is also known as intermediate level fusion. The objective of feature fusion is to extract or select a limited number of important features for subsequent data analysis through feature reduction methods, which can reduce computation and memory resources.

(3) Decision level fusion: it is also called high level fusion, which fuses decisions coming from multiple detectors. Each detector completes basic detection locally including preprocess, feature reduction, and identification to establish preliminary inferences on observed objectives. And then these inferences are fused into a comprehensive and accurate decision through the decision fusion techniques.

2.2.3. Data Fusion Applications. As a technology, DF is a multidisciplinary research field with a wide range of potential applications in such areas as automatic control, image recognition, target detection, and intrusion detection. The following is a brief introduction to DF applications based on the review of some related literatures.

In [43], Cao et al. presented a fire automation control system based on DF by applying it into intelligent building. The control system consists of six layers (sensor layer, sensor subsystem layer, primary fusion subsystem layer, decision management subsystem layer, actuator subsystem layer, and actuator layer). It can be applied into intelligent building to automatically realize accurate fire alarm and fire protection.

Zhang et al. proposed a DF based smart home control system [44]. The proposed smart home control framework includes the Internet access module, information acquisition module, and internal network service module with Bluetooth connection, data fusion controller that uses fuzzy logic and fuzzy neural network, and embedded computer in household appliances. It integrates information from multiple sources to control household appliances to create an intelligent home environment.

In [45], DF system based on D-S (Dempster-Shafer) evidence reasoning was proposed, in which two Charge Coupled Device (CCD) cameras and an Infrared Radiation (IR) sensor are used to extract the characteristics for identifying a missile target. Based on the D-S evidence reasoning, the authors recognized missile target and jamming light on region square feature and clutter and fire pile on position feature, respectively. The probability of identification obtained by integrating the three sensors with D-S evidence is greatly improved comparing with the method of using a single sensor.

Hu and Wang applied DF fuzzy theory to develop a fire alarm system based on a wireless sensor network [46]. This system not only offers detection correctness, but also improves the intelligence of monitoring. The proposed method has excellent performance and it is superior to traditional diagnostic methods with a single sensor.

In [47], a deep model for remote sensing DF and classification was proposed. The Convolutional Neural Network (CNN) is used to efficiently extract abstract information characteristics from Hyperspectral Image/Multispectral Image (MSI/HSI) and Light Detection and Ranging (LIDAR) data, respectively. Then, Deep Neural Networks (DNN) was used to fuse the heterogeneous characteristics obtained by CNN. The proposed depth fusion model provides competitive results in terms of classification accuracy. In addition, the proposed deep learning idea opens a new window for future remote sensing data fusion.

In [48], Yan et al. applied DF to reputation generation and proposed a reputation generation method based on opinion fusion and mining. The opinions were fused and classified into a number of major opinion sets containing opinions with similar or identical attitudes. Based on these opinion sets, the rating is aggregated to normalize the reputation of the entity. The experimental results from actual data analysis of several popular Chinese and English commercial websites demonstrated the versatility and accuracy of the method.

Liu et al. collected four articles to study the application of DF in the Internet of Things (IoT) [49]. With a large number of wireless sensor devices, IoT generates a large amount of data, which are massive, multisourced, heterogeneous, dynamic, and sparse. In the special issue, they believed that DF was an important tool for processing and managing these data to improve processing efficiency and provide advanced intelligence. By exploiting the synergy among the datasets, DF can reduce the amount of data, filter noise measurements, and make inferences at any stage of data processing in IoT.

A DF model for intrusion detection was presented in [42], based on clustering. The model uses a centralized approach to fuse data from different analyzers and then make a final analysis decision. The main strength of the proposed approach lies in its accuracy to fuse information from different detection modules and its adaptability to scalability. In addition, the DF module takes into account the efficiency of each analyzer in the process of fusion and can predict upcoming network threats.

2.2.4. A General Fusion Framework for Network Intrusion Detection. Herein, we specify a general fusion framework for network intrusion detection, as shown in Figure 1. The framework is comprised of the following parts.

(a) Input/Data Source. In order to monitor network status and detect and prevent attacks, we need to collect data from multiple sources in the network. These data include different types of packets and the statistical logs of network devices, for example, hosts, routers, and switches. They have different types and formats and cannot be processed directly.

(b) Data Preprocessing. The function of data preprocessing is to eliminate obviously wrong, invalid, or duplicate data and to get the valid data that can be used. The raw data is normalized and digitized through data preprocess, which is then converted into a unified format for analysis and processing.

(c) Feature Fusion. The network data has the characteristics of big data. Massive network data not only overly consumes computing and storage resources, but also cause dimensional disasters. Feature fusion occurs at the feature level and can reduce a large number of features to few features. The more streamlined data after feature fusion play a more important role in decision-making than the original features while accelerating data processing and increasing the detection accuracy of NIDS.

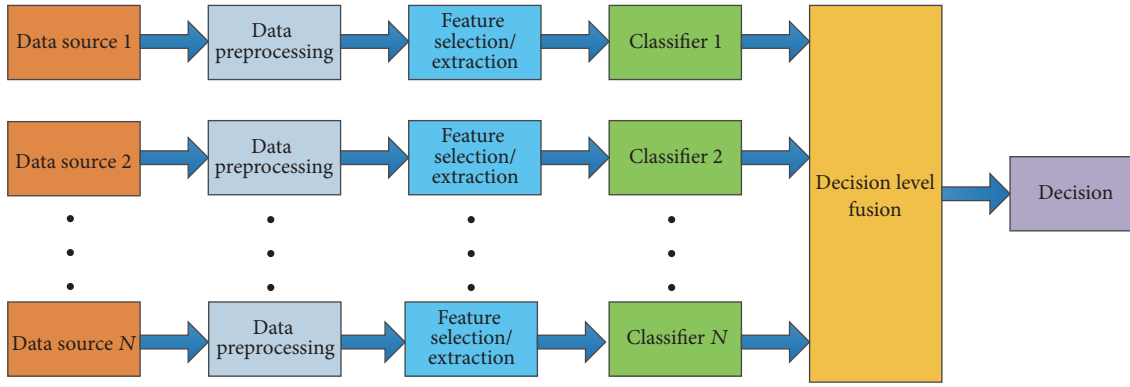

Figure 1: A general fusion framework for network intrusion detection.

(d) Classification. Intrusion detection can be seen as a pattern recognition system. Its performance is determined by the classifiers. Classifier models are obtained through training to identify abnormal network behaviors and make timely responses to the network attacks.

(e) Decision Fusion. Decision fusion is the integration of multiple results of basic detectors. The so-called decisions in the intrusion detection can be understood as the detection results of network behaviors. Decision fusion can achieve improved accuracy and more specific inference than the way of using a single detector alone. Besides, decision fusion can effectively detect complex attacks by integrating multiple decisions.

(f) Output/Decision. Output is the final decision, which usually is a judgment in the NIDS, either an abnormal behavior (e.g., an attack) or a normal behavior.

3. Data Fusion Techniques for NIDS

This section introduces the data fusion techniques, mainly focusing on feature fusion and decision fusion. We classify the fusion techniques shown in Figure 2 and describe the commonly used fusion techniques.

As mentioned above, DF techniques in NIDS can be classified into the data layer fusion, the feature layer fusion, and the decision layer fusion. To the best of our knowledge, the majority of researches on NIDS are based on open datasets, which leads to the result that the data level fusion is omitted in the related literatures. Therefore, we mainly review the DF techniques at the feature layer and the decision layer.

There are two main categories for feature fusion in NIDS: filters and wrappers [50]. The filters are applied through statistical methods, information theory based methods, or searching techniques [51], such as Principal Component Analysis (PCA), Latent Dirichlet Allocation (LDA), Independent Component Correlation Algorithm (ICA), and Correlation-Based Feature Selection (CFS). The wrapper uses a machine learning algorithm to evaluate and fuse features to identify the best subset representing the original dataset. The wrapper is based on two parts: feature search and evaluation algorithms. The wrapper approach is generally considered to generate better feature subsets but costs more computing and storage resources than the filter [27]. The filter and the wrapper are two complementary modes, which can be combined. A hybrid method is usually composed of two stages. First, the filter method is used to eliminate most of the useless or unimportant features, leaving only few important ones, which can effectively reduce the size of data processing. In the second stage, the remaining few features representing the original data are used as input parameters to send into the wrapper to further optimize the selection of important features.

The decision fusion methods are divided into two classes: winner-take-all and weighted sum, by considering how to combine decisions from basic classifiers [32]. Majority vote, weighted majority vote, Naïve-Bayes, RF (Random Forest), Adaboost, and D-S evidence theories are classified as the type of winner-take-all because they all have measured values for each basic classifier and the final decision depends on the classifier with the highest measured value. In case of the weighted sum, the weight of each basic classifier depends on its own capabilities. The weights of basic classifiers are calculated, and then their outputs with the weights are added to give a final decision. The method of weighted sum mainly includes average and neural network. Figure 2 gives the categories of fusion techniques. In what follows, we briefly described several commonly used feature fusion and decision fusion techniques, respectively.

3.1. Feature Fusion Techniques. There are many types of feature fusion methods in the literature. We introduce some of them due to space limitations. Some classic fusion techniques are described below.

3.1.1. PCA. Principal Component Analysis (PCA) is a multivariate statistical technique used for feature reduction [12, 52]. The goal of PCA in intrusion detection is to extract n (small integer) most important features representing the dataset. It can achieve dimensionality reduction while removing noise from the data and improving the performance of

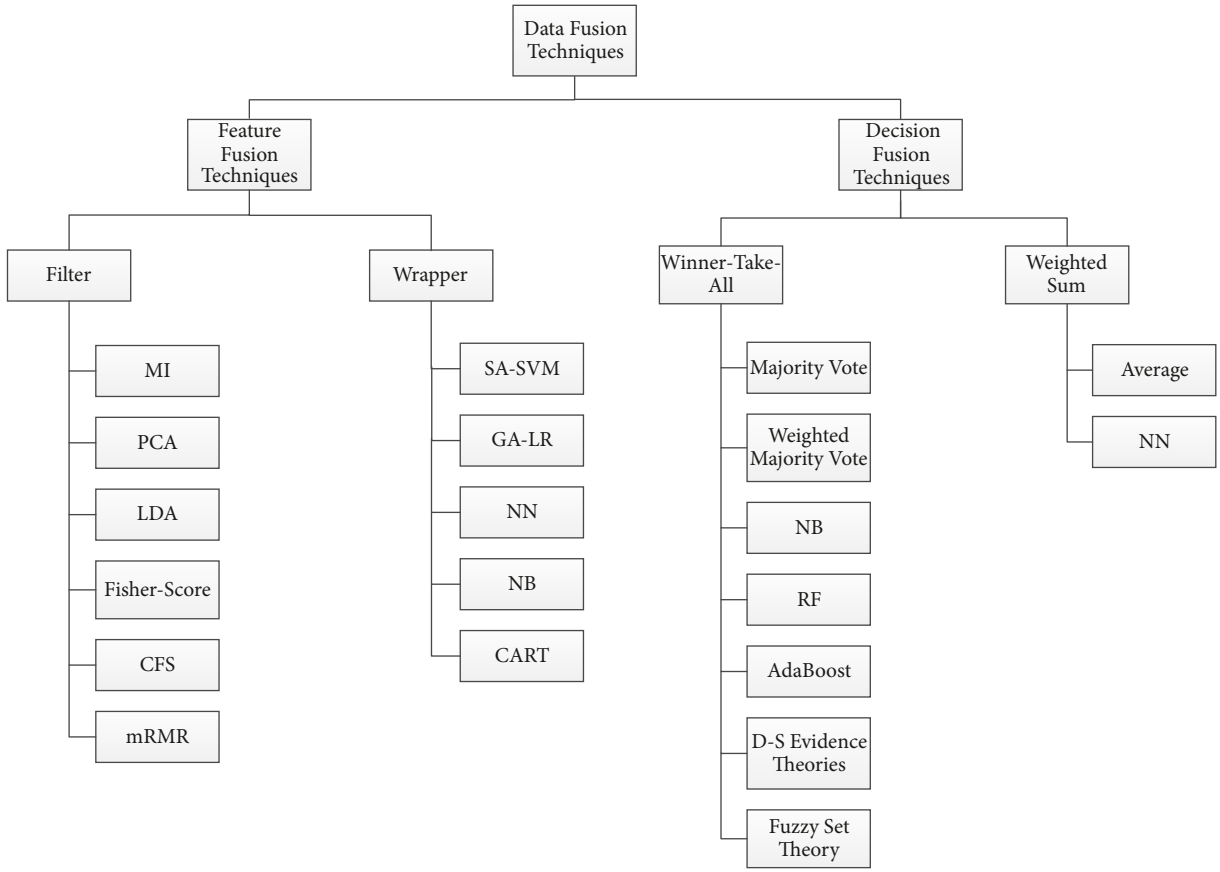

Figure 2: Categories of fusion techniques.

the system. In order to achieve these goals, PCA needs to extract new variables, that is, the main components. The first principal component has the largest variance that is the most representative of the entire dataset. The second principal component is computed under the constraint of being orthogonal to the first component and to have the largest possible variance. The other principal elements are calculated in the same way. These principal components form the new features of the original data. Before applying PCA, the data must be averaged and normalized to avoid the imbalance between the data values. PCA is popular in feature fusion because its simplicity and high precision. Nonetheless, in fact, each principal component can be represented by a linear combination of primitive features, which leads to a lack of interpretability for these principal components, especially when a large number of features are involved.

3.1.2. CFS. Correlation-Based Feature Selection (CFS) evaluates and ranks feature subsets rather than individual features [27]. It tends to have a set of attributes highly correlated with the class but with low intercorrelation. CFS often uses a variety of heuristic search strategies (such as hill climbing and best-first) to search a feature subset space within a reasonable time period. It first calculates the matrix of feature-class and the feature-feature correlation from the training data and then uses best-first to search the feature subset space [50]. The equation for CFS is

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}, \quad (1)$$

where M_s is the heuristic of the feature subset S containing k features, $\bar{r}_{cf}$ is the average value of all feature-classification correlations, and $\bar{r}_{ff}$ is the average value of all feature-feature correlations. The molecular $k\bar{r}_{cf}$ means the predictive ability of features, and the denominator $\sqrt{k + k(k-1)\bar{r}_{ff}}$ indicates the redundancy between features.

3.1.3. GA. Genetic Algorithm (GA) is a search heuristic model for simulating natural selection processes [53]. This heuristic approach is often used to generate useful solutions for optimization and search problems. GA is a kind of Evolutionary Algorithm (EA), which uses natural evolution-inspired techniques (such as genetic, mutation, selection, and crossover) to generate solutions for optimizing results. We can use the evaluation function to calculate the goodness of each chromosome. This operation begins with the initial population of randomly generated generations of chromosomes,

and the quality of each individual is gradually increased. Each individual chooses three basic GA operators, namely, selection, crossover, and mutation. In intrusion detection, in the face of a large number of features of original data, the GA can search for a subset of the raw features through Support Vector Machine (SVM), Neural Networks (NN), or other classifiers as evaluation functions. The advantage of this approach is that it has a flexible and powerful global search capability that converges from multiple directions without regard to previous knowledge of system behaviors. The main drawback is the high consumption of computing resources.

3.2. Decision Fusion Techniques. Comparing with feature fusion, the level of decision fusion is higher, and the data to be merged is more abstract. The decision fusion further improves the performance of the detection system, especially when a single detector is difficult to identify complex network behaviors. In what follows, we introduce several common decision fusion techniques.

3.2.1. Weighted Majority Vote. Weighted majority vote can assign weights to each basic classifier, which indicates the importance of the outputs of different classifiers for a final decision [32]. The weight varies according to the ability of the basic classifier to separate the samples. The formula is as below.

$$\sum_{i=1}^{L} b_i d_{i,k}(x) = \max_{1 \le j \le c} \sum_{i=1}^{L} b_i d_{i,j}(x), \tag{2}$$

where $d = [d_{i,1}, \ldots, d_{i,c}]^T \in \{0,1\}^c$, $i = 1 \cdots L$ is the outputs of the classifiers from the decision vector d, where L is the number of classifiers and $d_{i,j} = 1$ is 1 or 0 depending on whether classifier i chooses j, or not, respectively. The final decision to fuse multiple classifiers is determined by the base classifier's output $d_{i,j}(x)$ and corresponding weights b_i. This method assigns a higher weight to the basic classifier with higher accuracy, but it ignores other inaccurate base classifiers. The weights for the base classifiers are difficult to obtain and adjust. Therefore, it is difficult to detect new network attacks.

3.2.2. Bayesian Estimation. Bayesian estimation is applied to DF for a long time. It is an excellent method if prior probability is known. In order to obtain the most accurate and comprehensive information, this method first analyzes the compatibility of various sensors, removes false information with low confidence, and makes the Bayesian estimate of useful information under the assumption that the corresponding prior probabilities are known. The advantages of Bayesian approach include explicit uncertainty characterization and fast and efficient computation. Moreover, Bayesian networks offer good generalization with limited training data and easy maintenance when adding new features or new training data [23]. The disadvantage of Bayesian estimation is that it cannot distinguish unaware and uncertain information, and it can only handle the related events. In particular, it is difficult to know the prior probabilities in practical applications. When the hypothetical prior probabilities are contradictory to reality, the results of the inference will be undesirable and will become quite complicated when dealing with multiple hypotheses and multiple conditions. In fact, the Bayesian inference methods are now rarely applied in DF because of these defects.

3.2.3. D-S Evidence Theory. The Dempster-Shafer evidence theory, abbreviated as D-S theory, is a complete theory of dealing with uncertainty. Its most notable feature is the usage of "interval estimates" rather than "point estimates" for the description of uncertainty information. It shows great flexibility in distinguishing between unknown and uncertain. These advantages make it widely applicable to information fusion, expert systems, intelligence analysis, and multiattribute decision analysis.

In the NIDS using the DS evidence theory, the results of each basic classifier are considered to be different "evidences." Different pieces of evidence of the same hypothesis (e.g., network connection categories, such as normal or attack) are integrated to obtain the supporting degree of the hypothesis. On the basis of the supporting degree, whether the network connection is normal or intrusion can be finally judged [31]. Zhao et al. used D-S theory to fuse several basic classifiers [33]. The correct rates of fused results in terms of every kind of intrusions are all close to, or even higher than, the highest correct rates of all basic detectors, which achieves a high correct rate to all intrusions. D-S Evidence Theory is considered as the generalization of the Bayesian theory. It can well represent "uncertainty" and does not need to know prior probabilities, compared with the Bayesian theory. Besides, it also has some drawbacks, such as the fact that the evidence is required to be independent and there is a potential exponential explosion in computation.

3.2.4. Neural Network. Neural Network (NN) is a supervised learning method that consists of input neurons, output neurons, and hidden neurons. In order to represent the relationship between the input neuron and the output neuron, the neural network needs a large amount of labelled data to train and obtain an accurate model. NN has the characteristics of self-learning, self-adaptation, self-organization, and fault-tolerant, which enable it to solve complex nonlinear problems. Furthermore, the advantage of NN is that it can automatically adjust the connection weights without any domain-specific knowledge, while other methods use preselected weights to combine outputs [32]. Therefore, its strong capabilities can be well adapted to the requirements of multisource DF in NIDS. In network intrusion detection, the classification results of multiple detectors are used as input neurons, and the output neurons are integrated classification results. The output of the neural network is used as feedback to adjust the training parameters. With the improved parameters, the detectors can be fused to produce an improved resultant output. The main drawback of NN is the lack of valid criteria for creating, selecting, and combining the results of the base classifiers. For example, one may use a Multilayer Perceptron (MLP) or a radial basis function to find fusion weights with different structure.

Please note that the DF techniques are not limited to the above-mentioned ones. Other techniques are no longer described in detail. These techniques can be applied to fuse network data. The performance comparison of different fusion techniques is given in Section 5 based on the criteria proposed in Section 4.

4. Evaluation Criteria of DF Techniques

The application of DF techniques in intrusion detection has received particular attention in the field of network security. Many studies on DF have been conducted to improve the performance of NIDS. However, DF in NIDS still faces many serious challenges, such as how to reduce the complexity of massive data, how to ensure data security, and how to overcome the complexity and improve the efficiency of the fusion. Therefore, in order to facilitate the analysis and comparison of different fusion techniques, we propose a number of criteria for evaluating the performance of fusion techniques in NIDS based on the traditional criteria of IDS performance. Herein, we introduce specific evaluation criteria. Since most of the experiments for NIDS performance testing are based on a few public datasets, we firstly introduce the commonly used datasets for intrusion detection.

4.1. Datasets. Since real-time network data brings personal or organizational privacy issues and cannot be used for comparison of different algorithms, most of researches conduct experiments based on open datasets. Fusion techniques may show different performance based on different datasets. Herein, we introduce some classic datasets and new but more realistic datasets that are used in the field of intrusion detection research.

4.1.1. DARPA Dataset. In order to evaluate difficult intrusion detection techniques, the United States MIT Lincoln Laboratory successfully constructed a complete dataset in 1998, namely, DARPA 1998. The dataset is a 9-week network connection data collected from a simulated US Air Force LAN, dividing into training data and testing data. The testing data contains some types of attacks that do not appear in the training data, which makes the dataset more realistic. The KDD99 dataset was generated for the KDD cup competition, which extracts 41 features from the DARPA 1998 dataset. It is one of the most popular and comprehensive intrusion detection datasets and is widely applied to evaluate the performance of NIDSs [54]. It includes a complete training set, 10% training set, and a testing set. Each connection record in the KDD99 training dataset contains 41 feature attributes and an attack type label. The type of attack in KDD99 training dataset mainly includes Denial-of-Service (DOS) attacks, Probe attacks, User-to-Root (U2R) attacks, and Remote-to-Local (R2L) attacks. The KDD99_10% packet is a 10% sample of KDD99 packets, with approximately 490,000 data records, which is used in most of the literatures. However, there are many problems in KDD99; for example, the number of different types of attacks is not balanced and some data records are duplicate or invalid. To address these problems in the KDD99 dataset, as a new revision of the KDD99, NSL-KDD was proposed by Tavallaee et al. [55]. The training and testing datasets of the NSL-KDD consist of approximately 125,973 and 22,544 connection records, respectively. Similar to the KDD99 dataset, each record in this dataset has 41 quantitative and qualitative features.

4.1.2. Kyoto 2006+ Dataset. There is a fatal problem in the existing dataset benchmark (KDD99) for network security, which does not reflect the current network security situation and the latest attack characteristics. This is because it generated from a simulated network nearly 20 years ago. To overcome its limitations, the Kyoto 2006+ dataset was presented by Song et al. [56]. It is a dataset based on actual traffic data from 2006 to 2009, which comes from different types of honeypots installed in the Kyoto University. The dataset consists of 14 conventional features captured by honeypots based on the KDD99 dataset and 10 additional features. Conventional features include the duration of the session, service, source byte, and destination byte, which is meaningful and important for subsequent data processing or decision-making. In addition to 14 statistical features, additional features were extracted, which may enable us to investigate effectively what kinds of attacks happened in networks. It can be used for further analysis and evaluation of NIDSs. The Kyoto 2006+ dataset includes about 50,033,015 normal sessions and 434,343,255 attacks, in which 425,719 attacks are unknown. Each connection in the dataset has 23 features. Compared to the KDD99 dataset, the Kyoto 2006+ dataset is generated in the real network. By using the Kyoto 2006+ dataset, researchers can access more realistic and practical network security attacks.

4.1.3. UNSW-NB15 Dataset. The above-mentioned datasets cannot meet the needs of research on the current network security situation, especially KDD99 and NSL-KDD. The UNSW-NB15 [57] was created by the IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) for generating a dataset that consists of real modern normal activities and synthetic contemporary attacks. The data collection period was 16 hours on January 22, 2015, and 15 hours on February 17, 2015. Tcpdump tool is used to capture 100 GB of the raw traffic. This dataset contains nine types of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. Moreover, the Argus and Bro-IDS tools are used and twelve algorithms are developed to generate in total 49 features with class labels. There are 175,341 records in the training set and 82,332 records in the testing set. The key characteristics of the UNSW-NB15 dataset are a hybrid of the real modern normal behaviors and the synthetic attack activities. Thus, this dataset is considered as a new benchmark dataset that can be used for evaluating NIDSs by the NIDS research community [57]. It is worth noting that the IXIA tool contains all the information about the new attacks that are continuously updated from CVE site 4. This site is a public information security vulnerability and exposure dictionary. However, it is undeniable that the UNSW-NB15 dataset is more complex than the KDD99 dataset [58].

Table 1: The formulas of the metrics.

Measures	Equations
ACC	(TP + TN)/(TP + FP + TN + FN)
PR	TP/(TP + FP)
RR	TP/(TP + FN)
F-Measure	(2 ∗ PR ∗ RR)/(PR + RR)
FPR	FP/(TN + FP)
FNR	FN/(FN + TP)
FAR	(FPR + FNR)/2

4.2. Validity. The validity is the key to measuring the quality of the NIDS. The purpose of the application of fusion technology is to improve the performance of intrusion detection. Therefore, the validity can still be used to measure the fusion technology.

The elements of the validity evaluation metrics include TP (the number of positive samples predicted to be positive), FP (the number of negative samples predicted to be positive), FN (the number of positive samples predicted to be negative), and TN (the number of negative samples predicted to be negative). Based on these measurement elements, the accuracy (ACC), precision rate (PR), recall rate (RR), F-Measure, FPR, and FNR are applied to evaluate the performance of the fusion techniques. These metrics' formulas are listed in Table 1.

4.3. Efficiency. In the big data era, communications and activities between people generate high volume and high-dimensional network data that require real-time classification. In NIDS, not only the network behavior classification technology needs to be efficient, but also the efficiency of data fusion is crucial [59], which determines the efficiency of NIDS. Training time and testing time can be used to measure the efficiency of fusion technology. Besides, the number of features produced by feature fusion also measures the efficiency of the fusion technique.

4.4. Data Security. In actual network monitoring, DF and classification techniques concern data security issues in order to provide trustworthy data fusion results, such as data confidentiality, integrity, and creditability. We must consider that the privacy of individuals or organizations cannot be compromised when we analyze and fuse network data. Therefore, data security and data privacy also need to be considered in data fusion.

4.5. Scalability. Digital communications will enter the era of 5G with the rapid technology development. Large-scale heterogeneous networks have become the trend of network development, and mass data and heterogeneous DF technologies are increasingly important. Fusion techniques and frameworks should take scalability into consideration, such as compatibility with different data formats and scalability of memory and CPU, which, therefore, becomes a measure of fusion technologies.

5. Comparisons and Discussions

Based on the above evaluation criteria, we conduct a rigorous review and analysis on 31 related studies, of which 23 are feature fusion techniques and the remaining 8 are decision fusion techniques. The results of research and analysis are listed in Tables 2 and 3, respectively. The experiments reported in the above work were conducted based on published datasets, including KDD99, NSL-KDD, Kyoto 2006+, and UNSW-NB15. We analyzed and compared the performance of different fusion techniques in terms of the feature fusion and the decision fusion based on the proposed criteria and specified metrics. It must be mentioned that the following comparisons are made based on different datasets. In addition, the experimental details in the literature are different, which may affect the performance evaluation of data fusion techniques.

5.1. Comparison of Feature Fusion Techniques. In this part, we review feature fusion techniques based on our proposed criteria and show our evaluation results in Table 2.

The original intention of feature fusion is to reduce the size of data and improve the operation efficiency of NIDS. Therefore, the efficiency is the key to measure the quality of feature fusion. We concern with training time compared with testing time in evaluating the efficiency of feature fusion. This is because the training time is usually far longer than the corresponding testing time. We first analyze and compare the training time of classifiers using different feature fusion techniques based on different datasets. For the KDD dataset series (DARPA99, KDD99, and KDD99_10%), we can find that the training time of network intrusion classifier using the following feature fusion techniques is shorter than others, such as GFR, FRM-SFM [18], and CART [23]; CFS-GA [25] is very efficient for the NSL-KDD dataset; based on the Kyoto 2006+ dataset, PLS [12] helps to reduce time consumption of classifier training. In summary, these mentioned fusion techniques are outstandingly efficient in the training time of network behavior classifier. What these fusion techniques have in common is that fewer features are generated regardless of the dataset, with a minimum of 4 features in [25]. The filter is more efficient than the wrapper among these feature fusion techniques, and the hybrid methods usually have excellent efficiency.

In addition to efficiency, the validity is also an important measure of feature fusion techniques. For the KDD dataset series, SA-SVM [20], GA-LR [16], (Filter-MISF, FMIFS) [17], PCA [11], MIFS [24], (FRM-SFM, GFR) [18], SVM [9, 10, 19], (GeFS-mRMR, GeFS-CFS) [19], and NN [9] achieved very high accuracy, exceeding 99.20%, and the highest was 99.96% of SA-SVM. In addition, the FPR of Filter-MISF, GA-LR, Filter, MIFS, MLCFS [24], and SVM are less than 0.50%. We found out that GA-LR, SVM, Filter-MISF, and MISF perform very well in terms of validity in the KDD dataset series. As for the NSL-KDD dataset, (FMISF, MIFS, FLCFS) [24], Chi-Square [21], FVBRM [27], and CFS [30] performed excellently in accuracy, both exceeding 96.75% and up to 99.91% of FMIFS. The FPR of FMISF, MIFS, and FLCFS are all lower than 0.53%, and Chi-Square's FAR is 0.13%. These

Table 2: The performance of different feature reduction algorithms.

(a)

Feature fusion techniques	Article	Dataset	Number of training/testing data	Number of features	Classifier	Identified attack types	Metrics										
							Validity							Efficiency		Data security	Scalability
							ACC	PR	RR	F-Score	FPR	FNR	FAR	Training time (s)	Testing time (s)		
NN	[8]	DARPA99	6819/3679	84/37	MLF	All	*	*	*	*	*	*	*	*	*	×	×
	[9]	KDD99	7000/7000	41/13	NN	Attack/Normal	99.41%	*	*	*	*	*	*	*	*	×	×
	[10]	KDD99_10%	*	41/34	NN	Attack/Normal	81.57%	*	*	*	18.19%	0.25%	9.22%	*	*	×	×
PCA	[11]	KDD99_10%	5000/5000	41/10	SVM	Probe	99.78%	99.85%	99.70%	99.77%	*	*	*	276	*	×	×
		KDD99_10%	5000/5000	41/10	SVM	R2L	99.70%	99.50%	99.39%	99.53%	*	*	*	237	*	×	×
	[12]	Kyoto 20606+	31360/47040	18/5	MLP	Attack/Normal	97.12%	*	*	*	4.29%	1.44%	2.87%	22.14	*	×	×
	[13]	NSL_KDD	125971/22000	41/23	NN	All	*	86.49%	83.95%	83.78%	*	*	*	*	*	×	×
Fisher-Score	[14]	KDD99_10%	1500/1500	41/13	RBF-NN	Attack/Normal	*	85.33%	*	*	5.40%	*	*	6.71	0.13	×	×
				41/19			*	91.27%	*	*	5.20%	*	*	8.38	0.13	×	×
				41/28			*	95.70%	*	*	6.31%	*	*	9.07	0.18	×	×
				41/41			*	96.51%	*	*	11.74%	*	*	13.05	0.41	×	×
RF	[15]	KDD99_10%	16919/49838	41/34	RF	Attack/Normal	*	*	94.20%	*	1.10%	*	*	*	*	×	×
GA-LR	[16]	KDD99_10%	1000/1000	41/18	RF	Attack/Normal	99.90%	*	99.81%	*	0.11%	*	*	*	*	×	×
		UNSW-NB15	2000/2000	49/20	C4.5		81.42%	*	*	*	6.39%	*	*	*	*	×	×
Filter-MISF	[17]	KDD99_10%	15246/478775	41/6	LS-SVM	Attack/Normal	99.90%	*	99.93%	99.53%	0.07%	*	*	63.19	27.32	×	×
Filter				41/19			99.75%	*	99.43%	99.34%	0.17%	*	*	87.83	30.64	×	×
MISF				41/25			99.70%	*	99.38%	99.34%	0.23%	*	*	*	*	×	×
GFR	[18]	KDD99_10%	550/115705	41/19	Multi-class SVM	All	98.62%	*	*	*	*	*	*	0.12	4.63	×	×
FRM-SFM				41/10			98.68%	*	*	*	*	*	*	0.16	7.8	×	×
SVM	[10]	KDD99_10%	*	41/30	Multi-class SVM	All	99.61%	*	*	*	*	*	*	81.18	6.36	×	×
	[19]	KDD99_10%	424/106	41/17	SVM	Dos and Probe	99.30%	*	*	*	*	*	*	*	*	×	×
	[9]	KDD99	7000/7000	41/13	SVM	Attack/Normal	99.52%	*	*	*	0.50%	*	*	163	1.06	×	×
SA-SVM	[20]	KDD99	93969/10441	41/23	SA-DT	All	99.96%	*	*	*	*	*	*	*	*	×	×
Chi-Square	[21]	NSL_KDD	8325/24975	41/31	Multi-class SVM	All	98.00%	*	*	*	*	*	0.13%	*	*	*	*

LR: logistic regression; MISF: Mutual Information-Based Feature Selection; GFR: gradually feature removal method; FRM: feature removal method; SFM: sole feature method; SA: simulated annealing; MLF: multilayer feed-forward; LS: least square; and DT: decision tree. *Not given. ×Not mentioned. Number of features (m/n): m and n represent the number of features before and after fusion, respectively.

(b)

Feature fusion techniques	Article	Dataset	Number of training/ testing data	Number of features	Classifier	Metrics: Identify attack types	Validity: ACC	PR	RR	F-Score	FPR	FNR	FAR	Efficiency: Training time (s)	Testing time (s)	Data security	Scalability
GeFS-CFS	[19]	KDD99_10%	424/106	41/4.5	C4.5	Dos and Probe	99.20%	*	*	*	*	*	*	*	*	×	×
GeFS-mRMR				41/18			99.60%	*	*	*	*	*	*	*	*	×	×
Markov-Blanket				41/17	BN		98.70%	*	*	*	*	*	*	*	*	×	×
BN	[22]	KDD99	4700/4700	41/30	BN	All	83.13%	*	*	*	*	*	*	37.44	*	×	×
	[23]	KDD99	5092/6890	41/17	BN	All	91.06%	*	*	*	*	*	*	112.11	59.4	×	×
CART	[23]	KDD99	5092/6890	41/12	CART	All	88.52%	*	*	*	*	*	*	3.86	0.19	×	×
	[19]	KDD99_10%	424/106	41/12	CART	Dos and Probe	94.30%	*	*	*	*	*	*	*	*	×	×
FMIFS	[24]	KDD99	*	41/19	LS-SVM	Attack/Normal	99.79%	*	99.46%	*	0.13%	*	*	*	*	×	×
		NSL_KDD	*				99.91%	*	98.76%	*	0.28%	*	*	*	*	×	×
		Kyoto 2006+	*				99.77%	*	99.64%	*	0.13%	*	*	*	*	×	×
MIFS	[24]	KDD99	*	41/25	LS-SVM	Attack/Normal	99.70%	*	99.38%	*	0.23%	*	*	*	*	×	×
		NSL_KDD	*				97.96%	*	95.96%	*	0.53%	*	*	*	*	×	×
		Kyoto 2006+	*				99.32%	*	98.59%	*	0.16%	*	*	*	*	×	×
FLCFS	[24]	KDD99	*	41/17	LS-SVM	Attack/Normal	97.63%	*	89.26%	*	0.34%	*	*	*	*	×	×
		NSL_KDD	*				96.75%	*	93.26%	*	0.47%	*	*	*	*	×	×
		Kyoto 2006+	*				99.12%	*	98.10%	*	0.58%	*	*	*	*	×	×
CFS-GA	[25]	NSL_KDD	125973/22544	41/4	J48	Attack/Normal	91.86%	*	*	*	*	*	*	1.37	0.22	×	×
BBAL-NB	[26]	NSL_KDD	9566/4500	41/15	NB	Attack/Normal	*	*	91.62%	*	5.73%	*	*	*	0.76	×	×
BBAL-SVM				41/16	SVM		*	*	95.87%	*	2.89%	*	*	*	68.77	×	×
FVBRM	[27]	NSL_KDD	56687/6299	41/24	NB	All	97.78%	*	*	*	*	*	*	*	9.42	×	×
ML	[28]	KDD99	90000/10000	41/16	SVM	All	90.36%	*	*	*	*	*	*	*	*	×	×
		NSL_KDD	90000/10000	41/16		Attack/Normal	89.35%	*	*	*	*	*	*	*	*	×	×
		Kyoto 2006+	90000/10000	23/8		Attack/Normal	87.12%	*	*	*	*	*	*	*	*	×	×
ARM	[29]	KDD99	*	41/11	NB	All	62.02%	*	*	*	*	*	*	*	*	×	×
		UNSW-NB15	*	49/11			37.50%	*	*	*	*	*	*	*	*	×	×
HVS	[12]	Kyoto 2006+	31360/47040	18/5	MLP	Attack/Normal	98.28%	*	*	*	3.05%	0.35%	1.70%	64.47	0.02	×	×
PLS			*	18/5	PLS	Attack/Normal	94.72%	*	*	*	6.52%	4.02%	5.27%	0.02	0.03	×	×
CFS	[30]	NSL_KDD	25192/11850	41/8	RandomTree	All	97.49%						2.50%				

GeFS: generic feature selection; mRMR: minimal redundancy maximal relevance; BN: Bayesian networks; CART: classification and regression tree; FMIFS: flexible mutual information feature selection; MIFS: mutual information feature selection; FLCFS: flexible mutual information feature selection; BBAL: binary bat algorithm with lévy flights; FVBRM: feature vitality based reduction method; ML: feature vitality based reduction method; ARM: association rule mining; HVS: heuristic for variable selection; and PLS: partial least squares regression. *Not given. ×Not mentioned. Number of features (m/n): m and n represent the number of features before and after fusion, respectively.

Table 3: The performance of different decision reduction algorithms.

Decision fusion techniques	Article	Dataset	Number of training/testing data	Classifier	Identified attack types	Metrics							
						Validity						Data security	Scalability
						ACC	PR	RR	F-Score	FPR	FNR		
D-S Evidence Theory	[31]	KDD99	*	Multiclass SVM	Attack/normal	*	*	95.10%	*	0.19%	4.74%	×	×
	[32]	KDD99	*	RBF-NN	Dos	99.08%	*	*	*	0.71%	*	×	×
	[33]	KDD99	30000/30000	C4.5	Attack/normal	98.90%	*	*	*	*	*	×	×
				BN		96.70%	*	*	*	*	*	×	×
				NN		99.20%	*	*	*	*	*	×	×
				MDT		86.30%	*	*	*	*	*	×	×
				D-S fusion		99.10%	*	*	*	*	*	×	×
RF	[15]	KDD99_10%	16919/49838	RF	Attack/normal	*	*	94.20%	*	1.10%	*	×	×
Adaboost	[34]	KDD99	494021/311029	Decision stumps	Attack/normal	*	*	90.02%	*	1.68%	*	×	×
NN	[35]	DARPA99	*	PHAD	All	99%	35%	28.00%	31%	*	*	×	×
				ALAD		99%	38%	32.00%	35%	*	*	×	×
				Snort		99%	9%	51.00%	15%	*	*	×	×
				Data-dependent fusion		99%	39%	68.00%	50%	*	*	×	×
RBF-NN	[32]	KDD99	*	RBF-NN	Dos	99.59%	*	*	*	0.63%	*	×	×
Majority voting rule	[36]	NSL_KDD	8105/11695	BN	Attack/normal	*	93.10%	91.90%	92.20%	*	*	×	×
				IBK		*	99.60%	99.60%	99.60%	*	*	×	×
				J48		*	98.50%	98.50%	98.50%	*	*	×	×
				SVM		*	98.50%	92.90%	92.60%	*	*	×	×
				Classifier fusion		*	99.10%	99.40%	99.20%	*	*	×	×
MLP	[37]	KDD99	833/7436	MLP-4 intrinsic features	Attack/normal	*	*	*	*	3.19%	*	×	×
				MLP-7 content features		*	*	*	*	2.25%	*	×	×
				MLP-19 traffic features		*	*	*	*	23.94%	*	×	×
				MLP-30 features		*	*	*	*	3.57%	*	×	×

PHAD: packet header anomaly detection system; ALAD: application layer anomaly detector; MDT: Multirandom Decision Tree; and IBK: lazy classifier. *Not given. ×Not mentioned. Number of features (m/n): m and n represent the number of features before and after fusion, respectively.

feature fusion techniques have outstanding characteristics in NIDSs based on NSL-KDD datasets. In the Kyoto 2006+ dataset, the accuracy of (FMIFS, MIFS, FLCFS) [24] and (HVS, PCA) [12] was all higher than 97.12%, and the FPR of FMIFS, MIFS, and FLCFS are all below 0.58%.

A notable fact is that the accuracy of the classification in the new dataset (UNSW-NB15) is not as good as the old datasets mentioned earlier (such as KDD dataset series). The major reason is that the UNSW-NB15 dataset is considered complex due to the similar behaviors of the modern attack and normal network traffic compared to the KDD99 dataset [55]. So far, the effectiveness of network intrusion detection is not good based on the UNSW-NB15 dataset. The accuracy in [16] reached the highest accuracy 81.42% based on our statistics, and the corresponding feature fusion technique and classifier are GA-LR, C4.5, respectively. Decision Tree (DT) classifier has indeed performed better in the UNSW-NB15 dataset [55] than other methods. The misfortune is not alone. The FAR of NIDSs in the UNSW-NB15 dataset is also bad. Therefore, advanced classification techniques and feature fusion techniques need further study. In general, GA-LR, SVM, Filter-MISF, and MISF show excellent validity in the KDD dataset series; FMISF, MIFS, FLCFS, and Chi-Square are more valid in the NSL-KDD dataset; the feature fusion techniques with high-validity are FMIFS, MIFS, and FLCFS in the Kyoto 2006+ dataset. Because the performance of network intrusion detection based on UNSW-NB15 dataset is not very good, more advanced fusion and classification techniques should be further investigated in order to identify the anomalies from this complex dataset.

Unfortunately, the fusion techniques in the literature we have reviewed have not considered the security of data fusion. The data privacy issues were not covered because existing experiments were based on the public datasets. In addition, the scalability of fusion technologies and frameworks were normally not mentioned in the past work. However, these properties of data fusion are particularly important in the big data era. More efforts are needed in order to solve these issues.

5.2. Comparison of Decision Fusion Techniques. In this subsection, we analyze the performance of different decision fusion techniques based on the proposed criteria and show our evaluation results in Table 3.

According to Table 3, we can find that the training and testing time of the classifiers are not recorded. The reason is that decision fusion techniques fuse the recognition results of basic classifiers. Although the training and testing time of classifiers can reflect the efficiency of classifiers, it cannot reflect the merits of decision fusion techniques. Besides, the KDD dataset series are used in the most statistical literature. So herein, we mainly analyze the validity of decision fusion techniques based on the KDD dataset series. The accuracy of D-S Evidence Theory [32, 33] and NN [33] is over 99%, which is usually higher than the accuracy of a single basic classifier. The FPR is also reduced through the integration of basic classifiers. The FPR in [31] (D-S Evidence Theory) is as low as 0.19%. As a group, D-S Evidence Theory, Data-Dependent Fusion, NN, RF, and Adaboost show good fusion performance in combining multiple basic decisions.

Like the feature fusion techniques, the existing decision fusion techniques did not consider the credibility of basic decisions and data security in the process of integration, which will affect the reliability of the final results or cause privacy leakage. Besides, most of the literatures also fail to analyze the scalability of decision fusion. We believe that these aspects are very important and should attract special attention.

6. Open Issues and Future Research Directions

In recent years, DF has achieved special attention and has developed rapidly in many fields. In the field of network intrusion detection, scholars have conducted extensive researches in DF and have made significant progress. However, the current data fusion techniques still face some serious challenges or open issues, which are summarized as below according to our literature review.

First, most of the existing researches were conducted based on open datasets and the practicability of these fusion algorithms or techniques needs further validation. Few researches used real network data because it is easy to expose privacy and cannot measure or compare with other existing works, which is not conductive to the development of data fusion technology. In fact, this is a difficult contradiction, which hinders the further development of network intrusion detection.

Second, in the era of big data, the network security monitoring and prevention may need real-time fusion and processing of massive network data. However, large data communication overhead and long computation delay are obviously a big challenge to overcome.

Third, existing DF technologies do not consider data security, including confidentiality and credibility. The feature fusion techniques could reveal the privacy of individuals or organizations, and the decision fusion techniques need to identify the credibility of local decisions. All above are not considered in the past work.

Fourth, since most of the researches conducted their work over some public datasets and these datasets are preprocessed, there are few data level fusion techniques used in intrusion detection. However, we are facing a large number of different types of raw data in actual networks. Thus, the data layer fusion becomes indispensable for intrusion detection. Special efforts are expected on data fusion with regard to network intrusion detection.

Fifth, there is a lack of studies on the visualization of data fusion. Through utilizing the visualization algorithm, we can not only deeply understand the features and effectiveness of the fusion technology, but also easily identify the distribution characteristics of the fused data. Few articles use a visual method to analyze classical datasets. In [60], Ruan et al. performed a visual analysis of the KDD99 dataset using MDC and PCA techniques to clearly identify normal and attack clusters. Based on this research, we believe that it is also necessary to provide a beautiful and comprehensive data fusion expression.

In addition, based on the above open issues, we further proposed a number of promising research directions in the field of data fusion for network intrusion detection.

First, the improvement of data fusion technology depends on new datasets to evaluate and verify. Most of the fusion techniques and intrusion detection technology show excellent performance on some old datasets, such as KDD99 and NSL-KDD. However, these datasets are out of date and do not represent the current network security situation, which deviates from the actual network security detection. More research needs to be done on new dataset collection, such as UNSW-NB15. The existing problem is that the performance of feature fusion based on the UNSW-NB15 dataset is not good. We should further study more advanced or appropriate fusion techniques to better identify abnormalities from complex network data.

Second, big network data fusion techniques should be investigated. The current fusion techniques are difficult to effectively and adaptively integrate network data of high-velocity, varieties of formats and types. In the era of big data, in addition to the large amount of data, the network data that needs to be collected come from different sources in different types of networks. Therefore, the collection of heterogeneous network data is required to research more advanced fusion methods.

Third, universal, flexible, and extensible fusion framework should be studied. There are many kinds of data fusion technologies, and the principles and mathematical theories of some fusion technologies are not easy to understand. Therefore, the simple, easy-to-use, universal, and easy-to-expand network data fusion architecture is worth studying. It can modularize mature fusion techniques and provide open interfaces for new fusion methods and architectures; thus, it greatly promotes the development of data fusion in the field of intrusion detection.

Fourth, data security in data fusion should be ensured. Most of the existing researches are based on public datasets, and security issues were not considered at all. In an actual network, network data includes personal or organizational information, which is easily revealed during the integration process, and the credibility and integrity of network data are difficult to guarantee. Data security and privacy should be protected and ensured in order to achieve trustworthy data fusion.

Finally, data layer fusion is an essential part of study towards efficient and practical data fusion in real-time network intrusion detection. The data layer fusion has not been seriously studied by relevant literature because of the widespread use of public datasets. The study of data layer fusion is also very significant, especially for practical applications. However, it is very difficult to collect and evaluate the original network data containing various modern attacks.

7. Conclusion

In this article, we categorically presented a detailed review on the feature fusion techniques and the decision fusion techniques used in NIDSs. A specific description of DF in the field of intrusion detection was presented in order to motivate this work. Based on the literature study, we proposed the evaluation criteria of data fusion techniques in terms of NIDS. The performance of different data fusion techniques is measured using the proposed criteria. We found that, in the feature fusion, in addition to some excellent fusion techniques, such as SVM and MIFS, the improved types of fusion techniques and hybrid fusion techniques are generally efficient and valid. For the decision fusion techniques, D-S Evidence Theory, NN, RF, and Adaboost can combine multiple decisions more precisely than other methods regarding the studies based on KDD dataset series. In addition, we found many effective classification algorithms in NIDS, namely, RF, C4.5, NN, and SVM, as well as their variants. Unfortunately, the current fusion techniques normally did not consider the security and the scalability of DF.

DF has been regarded as one of the most important technologies in improving the performance of the NIDSs. The use of DF can well alleviate the defects of network intrusion detection and improve the comprehensive performance of NIDSs. However, there are still many deficiencies in current DF techniques. Based on our review, we pointed out the main challenges and promising future research directions in this field of research. In summary, this article provides a good reference for researchers and practitioners in the field of network intrusion detection.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is sponsored by the National Key Research and Development Program of China (Grant 2016YFB0800700), the NSFC (Grants 61602359, 61672410, and U1536202), the Project Supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program no. 2016ZDJC-06), the Fundamental Research Funds for the Central Universities (JB181503), the 111 Project (Grants B08038 and B16037), and Academy of Finland (Grant no. 308087).

References

[1] J. Tian, W. Zhao, R. Du, and Z. Zhang, "A New Data Fusion Model of Intrusion Detection-IDSFP," in *Parallel and Distributed Processing and Applications*, vol. 3758 of *Lecture Notes in Computer Science*, pp. 371–382, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[2] F. E. White, "Data Fusion Lexicon," Defense Technical Information Center, 1991.

[3] H. Boström, S. F. Andler, M. Brohede et al., "On the definition of information fusion as a field of research," *Neoplasia*, vol. 13, pp. 98–107, 2007, IN1.

[4] B. R. Raghunath and S. N. Mahadeo, "Network Intrusion Detection System (NIDS)," in *Proceedings of the 2008 First International Conference on Emerging Trends in Engineering and Technology*, pp. 1272–1277, Nagpur, Maharashtra, India, July 2008.

[5] Y. Fu, Z. Yan, J. Cao, O. Koné, and X. Cao, "An Automata Based Intrusion Detection Method for Internet of Things," *Mobile Information Systems*, vol. 2017, Article ID 1750637, 13 pages, 2017.

[6] L. Wang and H. Xiao, "An integrated decision system for intrusion detection," in *Proceedings of the 1st International Conference on Multimedia Information Networking and Security, MINES 2009*, pp. 417–421, chn, November 2009.

[7] M. A. Aydin, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Computers and Electrical Engineering*, vol. 35, no. 3, pp. 517–526, 2009.

[8] H. Wang, X. Liu, J. Lai, and Y. Liang, "Network security situation awareness based on heterogeneous multi-sensor data fusion and neural network," in *Proceedings of the International Multi-Symposiums on Computer and Computational Sciences*, pp. 352–359, 2007.

[9] S. Mukkamala, G. Janoski, and A. Sung, "Audit data reduction for intrusion detection," Training, 2008.

[10] A. H. Sung and S. Mukkamala, "Identifying important features for intrusion detection using support vector machines and neural networks," in *Proceedings of the International Symposium on Applications and the Internet*, pp. 209–216, IEEE, Orlando, Fla, USA, January 2003.

[11] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of PCA and optimized SVM," in *Proceedings of the 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, pp. 879–884, ind, November 2014.

[12] A. Ammar, "Comparison of Feature Reduction Techniques for the Binominal Classification of Network Traffic," *Journal of Data Analysis Information Processing*, vol. 03, pp. 11–19, 2015.

[13] N. A. Biswas, F. M. Shah, W. M. Tammi, and S. Chakraborty, "FP-ANK: An improvised intrusion detection system with hybridization of neural network and K-means clustering over feature selection by PCA," in *Proceedings of the 18th International Conference on Computer and Information Technology, ICCIT 2015*, pp. 317–322, bgd, December 2015.

[14] J. Zhou, J. Wang, and Z. Qun, *The Research on Fisher-RBF Data Fusion Model of Network Security Detection*, Springer, Berlin, Heidelberg, Germany, 2012.

[15] J. Zhang, M. Zulkernine, and A. Haque, "Random-forests-based network intrusion detection systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 5, pp. 649–659, 2008.

[16] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," *Computers & Security*, vol. 70, pp. 255–277, 2017.

[17] M. A. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and U. T. Nagar, "A novel feature selection approach for intrusion detection data classification," in *Proceedings of the IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 82–89, 2015.

[18] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Systems with Applications*, vol. 39, no. 1, pp. 424–430, 2012.

[19] H. T. Nguyen, S. Petrović, and K. Franke, "A comparison of feature-selection methods for intrusion detection," in *Lecture Notes in Computer Science*, I. Kotenko and V. Skormin, Eds., vol. 6258, pp. 242–255, 2010.

[20] S.-W. Lin, K.-C. Ying, C.-Y. Lee, and Z.-J. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection," *Applied Soft Computing*, vol. 12, no. 10, pp. 3285–3290, 2012.

[21] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 4, pp. 462–472, 2017.

[22] Y. Xu and W.-B. Zhang, "A novel IDS model based on a Bayesian fusion approach," in *Proceedings of the 1st International Conference on Multimedia Information Networking and Security, MINES 2009*, pp. 546–549, chn, November 2009.

[23] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," *Computers & Security*, vol. 24, no. 4, pp. 295–307, 2005.

[24] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[25] K. S. Desale and R. Ade, "Genetic algorithm based feature selection approach for effective intrusion detection system," in *Proceedings of the International Conference on Computer Communication and Informatics*, pp. 1–6, 2015.

[26] A.-C. Enache, V. Sgarciu, and A. Petrescu–Niţă, "Intelligent feature selection method rooted in Binary Bat Algorithm for intrusion detection," in *Proceedings of the Jubilee IEEE International Symposium on Applied Computational Intelligence and Informatics, SACI 2015*, pp. 517–521, 2015.

[27] S. Mukherjee and N. Sharma, "Intrusion Detection using Naive Bayes Classifier with Feature Reduction," *Procedia Technology*, vol. 4, pp. 119–128, 2012.

[28] M. A. Ambusaidi, X. He, and P. Nanda, "Unsupervised Feature Selection Method for Intrusion Detection System," in *Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA*, pp. 295–301, 2015.

[29] N. Moustafa and J. Slay, "The significant features of the UNSW-NB15 and the KDD99 data sets for Network Intrusion Detection Systems," in *Proceedings of the International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, 2017.

[30] S. Thaseen and C. A. Kumar, "An analysis of supervised tree based classifiers for intrusion detection system," in *Proceedings of the International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pp. 294–299, 2013.

[31] F. Xie, H. Yang, Y. Peng, and H. Gao, "Data fusion detection model based on SVM and evidence theory," in *Proceedings of the 2012 IEEE 14th International Conference on Communication Technology, ICCT 2012*, pp. 814–818, chn, November 2012.

[32] A. P. Chan, D. S. Yeung, E. C. Tsang, and W. W. Ng, "Empirical study on fusion methods using ensemble of RBFNN for network intrusion detection," in *Lecture Notes in Artificial Intelligence*, S. Yeung, Z. Q. Liu, X. Z. Wang, and H. Yan, Eds., vol. 3930, pp. 682–690, 2006.

[33] X. Zhao, H. Jiang, and L. Jiao, "A Data Fusion Based Intrusion Detection Model," in *Proceedings of the 2009 First International Workshop on Education Technology and Computer Science*, pp. 1017–1021, Wuhan, Hubei, China, March 2009.

[34] W. Hu and S. Maybank, "Adaboost-Based Algorithm for Network Intrusion Detection," *IEEE Transactions on Systems Man Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man Cybernetics Society*, vol. 38, pp. 577-83, 2008.

[35] C. Thomas and N. Balakrishnan, "Advanced sensor fusion technique for enhanced intrusion detection," in *Proceedings of*

the IEEE International Conference on Intelligence and Security Informatics (ISI '08), pp. 173–178, Taipei, Taiwan, June 2008.

[36] K. Saleem Malik Raja and K. Jeya Kumar, "Diversified intrusion detection using Various Detection methodologies with sensor fusion," in *Proceedings of the 2014 International Conference On Computation of Power , Energy, Information and Communication (ICCPEIC)*, pp. 442–448, Chennai, India, April 2014.

[37] G. Giacinto, F. Roli, and L. Didaci, "Fusion of multiple classifiers for intrusion detection in computer networks," *Pattern Recognition Letters*, vol. 24, no. 12, pp. 1795–1803, 2003.

[38] J. Song, H. Takakura, and Y. Kwon, "A generalized feature extraction scheme to detect 0-day attacks via IDS alerts," in *Proceedings of the 2008 International Symposium on Applications and the Internet (SAINT'08)*, pp. 55–61, fin, August 2008.

[39] M. Beheshti, J. Han, K. Kowalski, J. Ortiz, J. Tomelden, and D. Alvillar, "Packet information collection and transformation for network intrusion detection and prevention," in *Proceedings of the 2008 International Symposium on Telecommunications (IST'08)*, pp. 42–48, irn, August 2008.

[40] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks," *Expert Systems with Applications*, vol. 29, no. 4, pp. 713–722, 2005.

[41] E. Waltz and J. Llinas, *Handbook of Multisensor Data Fusion*, CRC Press, 2001.

[42] B. A. Fessi, S. Benabdallah, Y. Djemaiel, and N. Boudriga, "A clustering data fusion method for intrusion detection system," in *Proceedings of the 11th IEEE International Conference on Computer and Information Technology, CIT 2011 and 11th IEEE International Conference on Scalable Computing and Communications, SCALCOM 2011*, pp. 539–545, cyp, September 2011.

[43] L. Cao, J. Tian, and W. Jiang, "Information fusion technology and its application to fire automatic control system of intelligent building," in *Proceedings of the International Conference on Information Acquisition (ICIA'07)*, pp. 445–450, Seogwipo-si, South Korea, July 2007.

[44] L. Zhang, H. Leung, and K. C. C. Chan, "Information fusion based smart home control system and its application," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1157–1165, 2008.

[45] Y. Xiao and Z. Shi, "Application of multi-sensor data fusion technology in target recognition," in *Proceedings of the 3rd IEEE International Conference on Advanced Computer Control (ICACC'11)*, pp. 441–444, chn, January 2011.

[46] X. Hu and X. Wang, "Application of fuzzy data fusion in multi-sensor fire monitoring," in *Proceedings of the 2012 International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA '12)*, vol. 1, pp. 157–159, August 2012.

[47] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep fusion of remote sensing data for accurate classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1253–1257, 2017.

[48] Z. Yan, X. Jing, and W. Pedrycz, "Fusing and mining opinions for reputation generation," *Information Fusion*, vol. 36, pp. 172–184, 2017.

[49] J. Liu, Z. Yan, and L. T. Yang, "Fusion - An aide to data mining in Internet of Things," *Information Fusion*, vol. 23, pp. 1-2, 2015.

[50] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant Features and the Subset Selection Problem," in *Proceedings of the Eleventh International Conference on International Conference on Machine Learning*, pp. 121–129, 1994.

[51] E. De La Hoz, E. De La Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," *Knowledge-Based Systems*, vol. 71, pp. 322–338, 2014.

[52] H. Chen, Y. Fu, and Z. Yan, "Survey on big data analysis algorithms for network security measurement," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 10394, pp. 128–142, 2017.

[53] D. Paudel, *A hybrid network intrusion detection system using SVM and GA*, 2016.

[54] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: a review," *Expert Systems with Applications*, vol. 36, no. 10, pp. 11994–12000, 2009.

[55] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proceedings of the International Conference on Computational Intelligence for Security and Defense Applications*, pp. 33–58, 2009.

[56] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation," in *Proceedings of the Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, pp. 29–36, 2011.

[57] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proceedings of the Military Communications and Information Systems Conference*, pp. 1–6, 2015.

[58] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Information Security Journal*, vol. 25, no. 1-3, pp. 18–31, 2016.

[59] D. Wu, B. Yang, and R. Wang, "Scalable privacy-preserving big data aggregation mechanism," *Digital Communications and Networks*, vol. 2, no. 3, pp. 122–129, 2016.

[60] Z. Ruan, Y. Miao, L. Pan, N. Patterson, and J. Zhang, "Visualization of big data security: a case study on the KDD99 cup data set," *Digital Communications and Networks*, vol. 3, no. 4, pp. 250–259, 2017.