IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 19, 2020, accepted March 12, 2020, date of publication March 18, 2020, date of current version March 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981517

A Survey of Privacy Protection and Network Security in User On-Demand Anonymous Communication

YUN HE[1], MIN ZHANG[1], XIAOLONG YANG[1], (Member, IEEE), JINGTANG LUO[2], (Member, IEEE), AND YIMING CHEN[2]

[1]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]State Grid Sichuan Economic Research Institute, Chengdu 610041, China

Corresponding author: Min Zhang (zhangm@ustb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671057 and Grant 61941113.

ABSTRACT In an untrusted Internet environment, there is a large amount of user privacy information being vulnerable to various attacks, and it is an important security concern for users on how to protect their privacy, and to further provide the anonymity guarantee. Recently, most privacy-sensitive users prefer anonymous communication to protect their private information from eavesdropping by attackers. Each user has different privacy protection demands for anonymous communication. However, anonymous communication methods cannot meet the various anonymity needs of users. Hence, the user on-demand anonymous communication is proposed, which can dynamically adjust the anonymity level according to a user's anonymity needs. However, the defense capabilities against attacks are insufficient in the existing user on-demand anonymous communication. Some malicious users in the network employ anonymous communication to hide their identities and attack the Internet. Moreover, the existing user on-demand anonymous communication is weak against traffic classification attacks based on machine learning. Therefore, this paper investigates its progresses and defects in privacy protection and network security. User on-demand anonymous communication is mainly composed of the user side and the transmission side. We separately investigate privacy protection and network security of user on-demand anonymous communication from the user side and the transmission side. By surveying various identity anonymity of users on the user side and the hierarchical anonymous transmission on the transmission side, we find two kinds of serious attacks in user on-demand anonymous communication, i.e., abuse of anonymous communication and traffic classification attacks based on machine learning. To solve the above security issues, we suggest the balancing mechanism of anonymous communication and behavior tracking which is used to prevent anonymous abuse, and the secure defense mechanism which is used to resist traffic classification attacks. To inspire follow-up research, we identify some open problems and emphasize future trends concerning user on-demand anonymous communication.

INDEX TERMS Privacy protection, network security, machine learning, traffic classification attacks, user on-demand anonymous communication.

I. INTRODUCTION

With the rapid development of the Internet, it has been widely used in all aspects of our daily living. For example, each of us uses the Internet to handle important business transactions such as banking, shopping, and taxation. However, in an untrusted Internet environment, there is a large number of security threats and attacks to user's privacy and private communications. The communication sessions of users on the Internet may be subject to attacks such as eavesdropping or traffic analysis attacks, which result in the leakage of users' privacy. Moreover, some attackers sell a user's privacy information stolen from the Internet to illegal organizations, posing a huge threat to the user's privacy and personal security.

With the outbreak of the Snowden incident [1], the fact that users' private information is monitored for a long time is revealed. If the user's private information is used for fraud

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Babu Thanikanti.

and extortion by illegal persons, it will pose a huge threat to users' personal and property security. Therefore, more and more users have begun to pay more attention to the security of their private data in the process of communication in untrusted Internet environments. Users usually use encryption to guarantee the confidentiality and security of transmitting information. However, encryption technology [2] only processes the data itself, and the header of the data packet still contains the identity information of the communication parties. The attackers can still obtain the identity information and communication relationship of the communication parties through an interception or traffic correlation attack [3], which poses a large threat to the privacy of users. Aiming at encryption technology, there are some limitations to the user's anonymity. Therefore, some users prefer to use anonymous communication [4] to hide the real communication relationship in the service flow without changing the existing network protocol and to achieve privacy protection for user identity so that an eavesdropper could not directly know or indirectly infer the communication relationship or identities of the two parties. Anonymous communication can protect the identity information of the communication initiator and communication receiver, and the association relationship between the communicating parties. Current anonymous communication is very versatile and can be used in anonymous mail, instant messaging, electronic voting, online payments, and military communications [5].

However, different users have different needs for communication anonymity through the Internet. Current anonymous communication cannot meet the variety of anonymous needs of different users [6]. User on-demand anonymous communication is that users can dynamically adjust the anonymity level of their communication according to the sensitivity of their transmitted information, which meets the user's various anonymity needs. Therefore, some scholars begin to research user on-demand anonymous communication.

Nevertheless, the defense capabilities against attacks in user on-demand anonymous communication are insufficient in anonymous protection ability [7] and resistance to traffic analysis attacks [8], [9]. Some malicious users in the network use anonymous communication to hide their identities, attack and destroy the network [10], [11], and send spam virus information, which affects the normal operation of the network and causes damage to users' privacy. Moreover, with the rapid development of machine learning and deep learning techniques [12], some attackers use machine learning technology to extract the statistical characteristics of user's communication traffic, so as to identify the traffic and the communication relationship between users.

Therefore, there are many challenges and difficulties for privacy protection and network security in user on-demand anonymous communication. It is necessary to carry out a survey of privacy protection and network security in on-demand anonymous communication, which will provide a reference for future research for the user on-demand anonymous communication. Shirazi *et al.* [13] and Luo *et al.* [14] survey the key technologies of typical anonymous communication systems, as well as the security of anonymous systems. But, none of them discuss how to meet user's different privacy protection demands for anonymous communication. As far as we know, there has been no survey on privacy protection and network security of user on-demand anonymous communication so far. In order to better protect user's privacy security and promote researches on user on-demand anonymous communication, it is necessary and urgent to review the privacy protection and network security in user on-demand anonymous communication. So, we will give a detailed survey on the privacy protection and network security of user on-demand anonymous communication.

User on-demand anonymous communication is mainly composed of the user side and the transmission side. Therefore, we separately investigate privacy protection and network security of user on-demand anonymous communication from the user side and the transmission side. The main content of our article is shown in Fig. 1. First, on the user side, to satisfy different communication anonymity needs of users, we present and analyze various identity anonymity. However, some malicious users use identity anonymity to attack the network, which affects the normal operation of the network. Therefore, it is necessary to trace back the user's network behavior. Second, on the transmission side, we concentrate on the hierarchical anonymous transmission. Aiming at traffic classification attacks based on machine learning during transmission and the corresponding defenses against traffic classification attacks are provided and analyzed. Finally, there are still security issues in tracing user's network behavior and defending against traffic classification attacks. Therefore, the challenges and future research trends of privacy protection and network security in user on-demand anonymous communication are discussed in detail.

The main contributions of the paper are as follows:

1) We survey various identity anonymity on the user side and elaborate abusing of anonymous communication in identity anonymity. By analyzing the characteristics of attacks, we provide the balancing mechanism of anonymous communication and behavior tracking to prevent anonymous abuse.
2) We elaborate and summarize the hierarchical anonymity transmission technologies based on their advantages and disadvantages. We find a kind of serious attack on the transmission side, i.e., traffic classification attacks based on machine learning. By surveying traffic classification attacks based on machine learning, we present and compare traffic camouflage technologies to defend against traffic classification attacks.
3) We discuss open challenges in the research of privacy protection and network security in user on-demand anonymous communication and give future research suggestions.

The rest of this paper proceeds as follows. In Part II, we make a detailed survey of users' privacy protection and network security from the user side. In Part III, we focus on

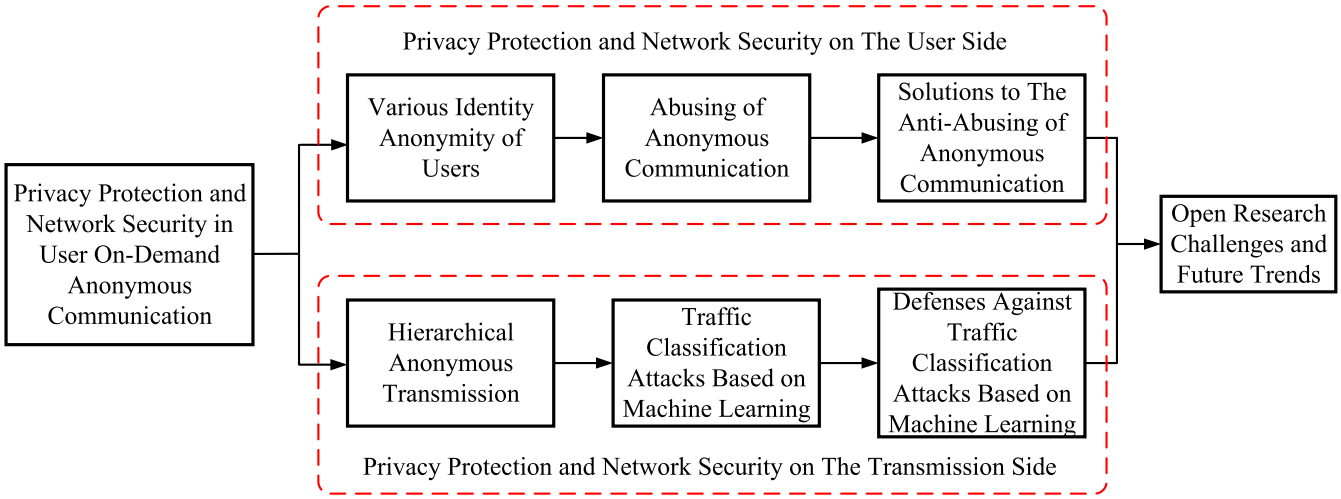

FIGURE 1. Overview of the main content of this article.

privacy protection and network security on the transmission side. In Part IV, we outline open challenges for research on user on-demand anonymous communication. In Part V, we point out future research directions for the user on-demand anonymous communication. Finally, in Part VI, we provide concluding comments.

II. PRIVACY PROTECTION AND NETWORK SECURITY ON THE USER SIDE

On the user side, to achieve various anonymous needs of users, we present various identity anonymity of users. However, some malicious users use identity anonymity to attack the network. Therefore, we provide solutions to the anti-abusing of anonymous communication.

A. VARIOUS IDENTITY ANONYMITY OF USERS

Current encryption technology only protects the content sent by users, and the header of a message still contains the user's identity and other related information, such as the IP address, which reveals the identity information of the communication partners and the communication relationship between users.

Based on the above problems, the identity information of the communicating parties is required to be anonymously protected on the user side [15]. In an untrusted Internet environment, different users have different needs for communication anonymity. To satisfy the different communication anonymity needs of users, a network user can randomly use multiple sources and destination addresses instead of the real user identity to communicate, which not only guarantees the privacy of users, and also meets the various anonymous needs of users.

Narten *et al.* [16] propose an IPv6 extension mechanism to configure multiple IPv6 addresses for a host by designating multiple unique values to its interface identifier (that is, the lower eight bytes of the IPv6 address), which enhances the privacy of users. However, the number of IPv6 addresses generated by the lower eight bytes of the IPv6 address is limited, which is not considered in this paper. Therefore, Han *et al.* [17] propose a pseudonym-based IPv6 addressing architecture in which Each pseudonym is an IPv6 address. However, the mapping between the pseudonym and the IPv6 address is obtained by the attacker, the privacy of the user will be revealed. To solve this problem, Sakurai *et al.* [18] propose to use a hash chain to generate a one-time receiving address, which has high security. Even if the hash address of the user is revealed, attackers still cannot know the true identity of the user.

B. NETWORK SECURITY ON THE USER SIDE

In user on-demand anonymous communication, some malicious users and attackers [19] hide their own identity to attack the network through anonymous communication, which affects the normal communication services of the network. Therefore, it is necessary to find the solutions to anti-abusing of anonymous communication.

1) ABUSING OF ANONYMOUS COMMUNICATION

In an untrusted Internet environment, some malicious users hide their true identity by anonymous communication [20] to launch a DDOS attack on the Internet, which causes the paralysis of the whole network. If the abuse issue of anonymous communication is not solved, the network environment will be filled with a large amount of malicious network behavior, which has a large negative impact on users' privacy and national network security.

On the Internet, the service provider often uses a blacklist to resist attacks. For source node, once it is found to generate a large amount of illegal traffic, the IP address of the source node is put on a blacklist, and data traffic from the source node is not accepted. For relay node, when abusive behavior from relay node occurs, the IP address of the relay node is blacklisted, and legal attempts of subsequent users to

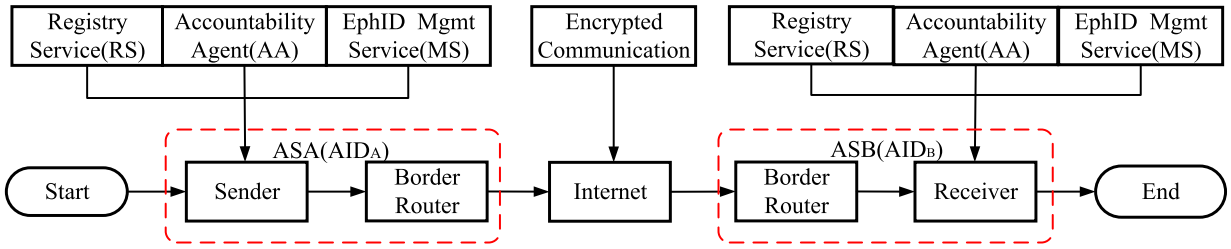

FIGURE 2. The balancing mechanism of anonymity and behavior tracking based on self-verifying identifiers.

utilize the relay node are not accepted. Under this mechanism, the most serious consequence is that a large number of relay nodes of the anonymous communication system are placed on the blacklist so that the system can no longer provide anonymous communication services for users.

For normal users, they protect their identity information from being leaked through anonymous communication. But for malicious users, they use anonymous communication for some illegal network activities. Therefore, to solve the above problem, we must find solutions to the anti-abusing of anonymous communication.

2) SOLUTIONS TO THE ANTI-ABUSING OF ANONYMOUS COMMUNICATION

To resist abusing of anonymous communication, we can trace users' network behavior to guarantee the security of communication.

Mallios *et al.* [21] present a source address translation mechanism. This mechanism can not only achieve anonymous protection of user's identity, but also trace back user's network behavior to prevent anonymous abuse. To ensure that the identity of the sender is anonymous, the source address in the data packet is converted to another address, the sender's address cannot be known from the source address information of the data packet, and the identity of the sender is anonymous. To track network behavior of users, when a sender behaves maliciously, the routers sequentially display their conversion to the sender's address starting from the router directly connected to the destination. This achieves the identity tracking of the sender. However, this method of identity anonymity includes IP address information of the sender and the receiver. Since the IP address can uniquely identify the user, an attacker can use the IP address information to discover the anonymous identity.

Therefore, based on the above problems, some researchers turned their attention to the study of self-verifying identifiers. Self-verifying identifiers differ from traditional IP addresses in that they are generated by a single hash chain and have a verification function of their own [22]. Even if some identifiers are intercepted and stolen by a malicious attacker, the attacker will not have the corresponding private key to achieve the user's true identity.

Lee [23] implement a balancing mechanism of anonymous communication and behavior tracking on the basis of the self-verifying identifier EphID. The mechanism can satisfy the different anonymous needs of different users and trace the network behavior of users in the autonomous domain. The autonomous domain in the balancing mechanism includes 3 parts, as shown in Fig. 2: the registration server (RS), the EphID management server (MS) and the accountability agent (AA). The registration server (RS) authenticates the legal identity of the host in the autonomous domain (AS). The EphID management server (MS) issues different identifier EphIDs to the legitimate host to meet the different anonymous needs of users, and the accountability agent (AA) tracks users' behavior in the network when abnormal traffic or attacks occur. Moreover, the user can select the usage mode of the identifier EphID according to his own anonymous needs. For a user without an anonymity requirement, only one EphID issued to the user by the EphID management server is applied in the whole communication process. For users with a low anonymity requirement, each data stream is used one EphID for communication. For users with a high anonymous demand, each packet is issued an EphID.

The balancing mechanism of anonymous communication and behavior tracking is based on self-verifying identifiers. The self-verifying identifiers are generated by a single hash chain. The hash chain is generated by iterating a selected secure hash function N times. Therefore, the computational complexity of the balancing mechanism is related to the selected hash function and the number of iterations. The greater the calculation cost of the selected hash function and the number of iterations, the greater the computational complexity of the mechanism.

The balancing mechanism of anonymous communication and behavior tracking replaces the IP identity information with self-verifying identifiers for anonymous communication on demand. Further, when abnormal traffic or attacks occur in the network, the self-verifying identifiers can be used to trace back the attack. This mechanism not only satisfies the users' anonymous communication on demand, but also guarantees the security of anonymous communication, achieves a balance between anonymous communication and behavior tracking, and protects the privacy of communication users. Finally, we summarize and list the corresponding solutions to the anti-abusing of anonymous communication, which are shown in Table 1.

TABLE 1. Solutions to the anti-abusing of anonymous communication.

Typical Solutions	Core Ideas	Advantages	Disadvantages
User behavior tracking based on source address translation [21]	When a sender behaves maliciously, the routers sequentially display their conversion to the sender's address starting from the router directly connected to the destination.	● Convert source IP address in the data packet to another IP address. ● Track the network behavior of the sender by IP address.	● Exposure of IP address information, vulnerable to attack ● Recipients are not protected anonymously.
User behavior tracking based on self-verifying identifier [23]	Accountability agent (AA) tracks users' behavior based on self-verifying identifiers when abnormal traffic or attacks occur.	● Representing user identity information with self-verifying identifiers. ● Attackers cannot decrypt the user identity, have higher security.	● Security risks due to the third-party proxy.

TABLE 2. Abusing of anonymous communication and corresponding solutions.

Security Issue	Description	Solutions
Abuse of anonymous communication	Attackers use anonymous communication to hide their true identity, abuse anonymous communication services, attack and destroy the network, and send virus information.	● User behavior tracking based on source address translation. ● User behavior tracking based on self-verifying identifiers.

In this section, we survey privacy protection and network security of user on-demand anonymous communication on the user side, and we summarize security issues and corresponding solutions of user on-demand anonymous communication on the user side in Table 2.

III. PRIVACY PROTECTION AND NETWORK SECURITY ON THE TRANSMISSION SIDE

Generally, users have different anonymity needs during communication. In order to meet users' different anonymous requirements, we usually employ hierarchical anonymous transmission. However, users' communication information is vulnerable to be monitored and performed traffic classification attacks based on machine learning by attackers in the transmission process. Therefore, this section will focus on hierarchical anonymous transmission and traffic classification attacks on the transmission side. As attackers usually attack anonymous communication on the transmission side, we will concentrate on this part.

A. HIERARCHICAL ANONYMOUS TRANSMISSION

Based on the transmission form of anonymous communication systems, they can be divided into two categories: anonymous communication systems based on *broadcast/multicast* and anonymous communication systems based on *rerouting*.

In anonymous communication systems based on broadcast or multicast, the user sends messages to the receiver by broadcasting, which protects the identity information of the user and makes it impossible for the attacker to infer who the real sender is. Typical anonymous communication systems based on broadcast and multicast include DC-Net [24] and Horders [25]. Anonymous communication systems based on rerouting achieve anonymity primarily through the form of rerouting and guarantee the anonymous security of a user's communication by establishing a communication transmission path between the user and the receiver. Typical anonymous communication systems based on rerouting include Mix-nets [26], Onion Routing [27], Crowds [28], and Tor [29].

The current anonymous communication systems only implement a single form of anonymous transmission. They cannot flexibly adjust the communication transmission mode, dynamically change the anonymity of communication transmission, or meet the different anonymous requirements of different users for communication transmission. To solve the above communication transmission problems in anonymous communication systems, some current researchers have proposed methods of hierarchical anonymous transmission. Hierarchical anonymous transmission can be broadly classified into those based on *broadcast or multicast* and those based on *rerouting*.

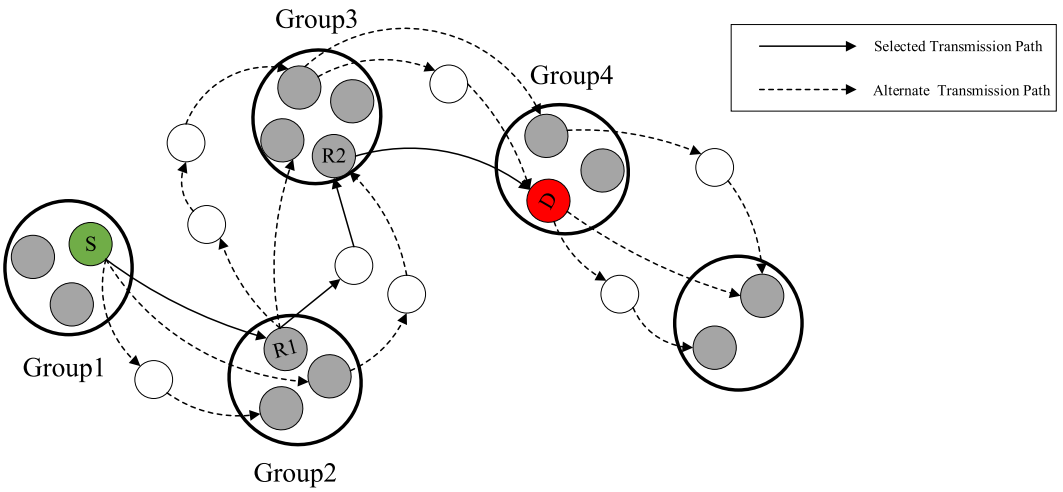

FIGURE 3. Anonymous transmission based on a partitioning mechanism.

1) HIERARCHICAL ANONYMOUS TRANSMISSION BASED ON BROADCAST/MULTICAST

Xu *et al.* [30] propose a hierarchical anonymous communication system based on the DC-Nets, which can achieve three different anonymous levels. The number of members in an anonymous group is related to the strength of anonymity. The more members there are, the stronger the anonymity. Different levels of anonymous communication can be achieved by setting the number of members in the anonymous group. The system assigns a user to a corresponding anonymous group according to the user's anonymous needs. On the other hand, this mechanism has several shortcomings. For example, when a node in an anonymous group is malicious or compromised, it may send errors or interference information that will cause communication to fail. At the same time, the mechanism is vulnerable to DDoS attacks. If malicious nodes collude with each other and continuously send redundant information, it will cause network paralysis.

Based on the above problems, a hybrid DC-Net protocol is proposed in [31], which dynamically adjusts a parameter that controls the degree of anonymity. However, to protect transmitted information, other users on the Internet are required to confuse the attacker by sending some useless messages. Clearly, this is a great waste of network resources and is vulnerable to DDoS attacks.

2) HIERARCHICAL ANONYMOUS TRANSMISSION BASED ON REROUTING

Xu *et al.* [32] propose a SAS hierarchical anonymous communication model based on Crowds. Users choose different anonymity levels to achieve a trade-off between user anonymity and communication performance according to their requirements for communication anonymity and communication performance. In real-world network applications, it is difficult for Crowds to accommodate a large number of users because it uses a centralized management mechanism. Therefore, Shi *et al.* [33] and Aseez and Mathew [34] introduce a partitioning mechanism. As shown in Fig. 3, users in the system are managed through partitioning, which ensures the anonymity of the sender and receiver. Meanwhile, forwarding nodes in the region are randomly selected by users as relay nodes, which can tolerate many users transmitting simultaneously. There are multiple transmission paths between S and D. When S communicates with D, a random path is selected for transmission. An attacker cannot determine which path the user is communicating through, which improves the anonymous security of communication. However, the forwarding node may be compromised by attackers, which will leak the user's information transmitted.

Therefore, the trust value of the node [35] is added to the partitioning mechanism, which achieves the user's hierarchical anonymous communication. Each node has a different level of anonymity. When an anonymous communication is performed, the user only uses trusted nodes with the highest degree of anonymity to perform encryption and decryption operations and uses ordinary routing nodes to forward information, which ensures the security of anonymous communication and avoids the delay problems of a long path.

In recent years, with the widespread use of Tor networks, Chen *et al.* [36], [37] propose a Tor-based anonymous communication architecture. Users can choose different anonymity levels for transmission according to their own needs. Different anonymity levels determine different lengths of the transmission path and sizes of the transmission routing set. As the number of users increases, more secure anonymity will be obtained. However, due to the probabilistic random forwarding, the forwarding path can be very long, increasing the system load and delay overhead. To solve the communication delay caused by probabilistic forwarding, Zhou *et al.* [38] propose a Tor-based controllable anonymous communication system, which can download information from the relevant relay nodes according to the user's anonymity needs and give users the ability to balance the anonymity and transmission efficiency of transmission links. Anonymous links with strong anonymity are established for data with high anonymity, and anonymous links with

TABLE 3. Comparison of hierarchical anonymous transmission methods.

Type	Reference	Advantages	Disadvantages
Based on broadcast or multicast	[30]	● Meets different anonymous needs of users based on the DC-Nets. ● Has good scalability.	● Vulnerable to DDoS attacks
	[31]	● Dynamically adjusts a parameter that controls the degree of anonymity based on the hybrid.	● Sends some useless messages. ● Vulnerable to DDoS attacks
Based on rerouting	[32]	● Achieves different anonymity levels based on Crowds. ● Achieves a trade-off between anonymity and communication performance	● Has difficulty accommodating a large number of users.
	[33][34]	● Meets different anonymity needs of users based on a partitioning mechanism. ● Accommodates a large number of users.	● Vulnerable to DDoS attacks.
	[35]	● The trust value of the node is added to the partitioning mechanism, improving transmission security. ● Avoids the delay problem of a long path.	● Vulnerable to DDoS attacks
	[36][37]	● Achieves different anonymity levels based on Tor. ● Uses the probabilistic random forwarding and has good security.	● Large system overhead and communication delay.
	[38]	● Balances anonymity and transmission efficiency. ● Enhances the user's autonomous controllability in the process of establishing an anonymous communication path.	● Vulnerable to traffic analysis attacks
	[42][43][44]	● Shares information among nodes. ● Provides access to information at any time in any place.	● Needs improvement in security.

high transmission efficiency are established for data with great timeliness [39], which enhances user's autonomous controllability in the process of establishing an anonymous communication path.

With the rapid development of the Internet of Things (IoT), big data and cloud computing, Stergiou *et al.* [40] and Memos *et al.* [41] propose a mutual integration among them, which can not only share information with each other, but also provide access to information at any time in any place. Moreover, more intelligent, more efficient and more secure transmission technologies [42]–[44] applicable to the above-integrated systems have been extensively researched.

Through the above survey of hierarchical anonymous transmission, we conduct a detailed comparison of hierarchical anonymous transmission technologies from the advantages and disadvantages, which are listed in Table 3.

B. NETWORK SECURITY ON THE TRANSMISSION SIDE

In recent years, with the development of artificial intelligence technologies, some attackers have used artificial intelligence technologies to analyze user's information transmitted in anonymous communication. By observing the traffic features [45] of transmission information, attackers can classify user's information transmitted, obtain private information about users [46], and launch effective network attacks [47]–[50]. Therefore, we will give a detailed overview of traffic classification methods based on machine learning, the traffic classification attacks based on traffic classification methods, and the defenses against the traffic classification attacks.

1) TRAFFIC CLASSIFICATION METHODS BASED ON MACHINE LEARNING

Machine learning is a method of realizing artificial intelligence. By learning existing data, gaining experience and discovering rules, a system's performance can be improved, and new data can be assessed and predicted [51]. Machine learning is divided into traditional machine learning (shallow learning) and deep learning. Fig. 4 describes the process of traffic classification based on traditional machine learning. First, the encrypted traffic that will be classified is collected. According to the features selected by a person,

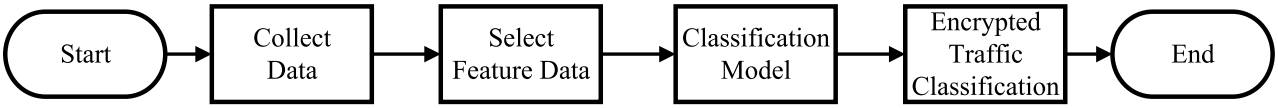

FIGURE 4. Flow chart of encrypted traffic classification based on traditional machine learning.

feature data is extracted and achieved from the encrypted traffic. Then, the related algorithms of machine learning are used to establish a classification model for the feature data. Next, the unknown encrypted traffic is classified using a classification model.

Encrypted traffic classification based on traditional machine learning is performed based on traffic features. The features used by the method are typical characteristics of the data stream, such as stream duration, number of streams per second, and so on. The statistical characteristics of the data stream are unique and can be used to distinguish it from other applications [52]. According to the way data stream features are used, encrypted traffic classification based on traditional machine learning is divided into supervised learning, unsupervised learning, and semisupervised learning types. Supervised learning [53] means that the training samples are labeled data. During the training process, the network parameters are adjusted by comparing the output of the network and the label of the data. Typical algorithms for supervised learning include support vector machines, neural networks, Bayesian algorithms, and decision trees. The training samples of unsupervised learning [54] need not be labeled, and samples with similar characteristics are determined to be in the same class in the learning process. Unsupervised learning algorithms include K-NN, K-means, GMM, HMM, and other clustering algorithms. The main concern of semisupervised learning [55] is how to obtain a good classifier when part of the training data is missing. Semisupervised learning includes semisupervised algorithms based on K-means and GMM. According to different application scenarios, traffic classifications are divided into *service classification*, *application classification,* and *user action classification*.

- *Service classification*: Service identification refers to the classification of the application type (e.g., website browsing, email, social software, video, file transfer, and voice calls) to which traffic belongs. At present, machine learning is widely used to improve the accuracy of service type identification for encrypted traffic. In [56], a Bayesian neural network is proposed to classify the famous P2P protocols, such as Kazaa, BitTorrent, and GnuTella, with an accuracy of 99%. In [57], the K-NN and C4.5 decision tree algorithms are used to classify traffic types according to time-related features such as traffic duration, bytes per second, and forward and backward arrival times, with an accuracy of 92%. The main disadvantage of these methods is that the stage of feature extraction and feature selection is largely completed manually with the help of experts. Therefore, these methods are time-consuming, expensive and prone to human error.
- *Application classification*: Application identification refers to distinguishing which application the traffic comes from. AppScanner [58] collects encrypted traffic by automatically running many popular apps in the Google App Store on mobile devices. Through a series of processing strategies, burst traffic is divided and segmented to obtain data characteristics. Finally, the two lightweight supervised learning algorithms, SVC and random forest, are used to classify streams to facilitate online deployment. AppScanner can capture and process network traffic in both online and offline modes, with a classification accuracy of over 99%.
  When a new application appears in the network, identifying the traffic of the new application poses a significant challenge to the robustness of classification. Zhang *et al.* [59] propose a new statistical traffic classification scheme that combines random forests and k-means for both supervised and unsupervised learning, identifies new application traffic, and accurately differentiates predefined application categories. The scheme consists of the following three parts shown in Fig. 5: an unknown discovery module, an application traffic classification module, and a system update module. The unknown discovery module is designed to automatically find traffic samples for the new application from a set of untagged traffic randomly collected from the target network. There are two steps to extract traffic samples from a set of unlabeled network traffic. The first step is the k-means based identification of traffic clusters. The second step is sample extraction by using a random forest model. The application traffic classification module takes premarked training samples and new application traffic samples as inputs to build a robust classifier. For robust traffic classification, a new classification method is proposed that considers flow correlation in real-world network traffic and classifies correlated flows together rather than in individual flows. To achieve fine-grained classification, the system update module can analyze the traffic of the new application and supplement the knowledge of the system by building new classes.
- *User action classification*: User action recognition refers to distinguishing which user's action the traffic comes from. Although network traffic is now transmitted in encrypted form, we can still classify the sub behavior of the application based on the statistical characteristics of the traffic. The traffic generated by the action sequences

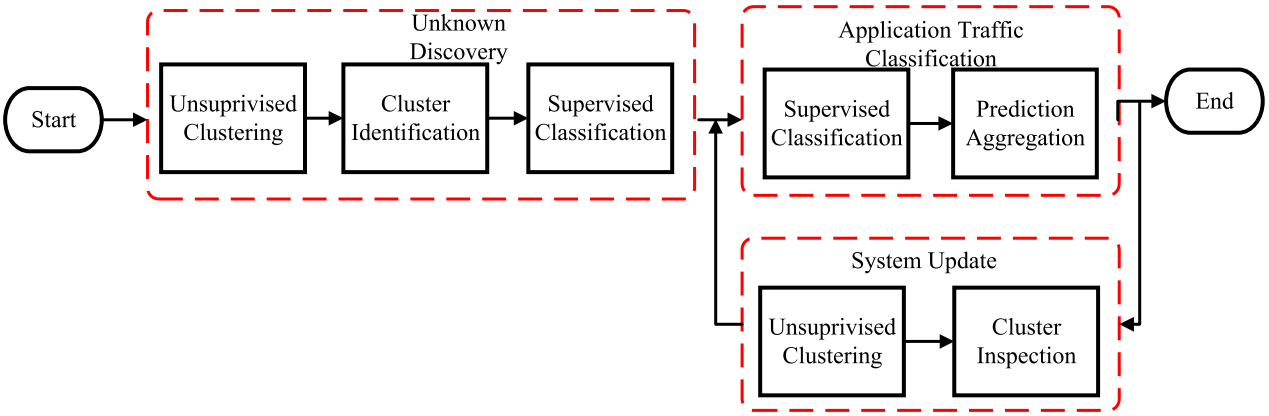

FIGURE 5. Application traffic classification based on random forests and k-means.

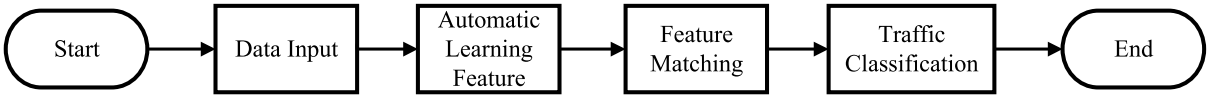

FIGURE 6. Flow chart of traffic classification based on deep learning.

of applications such as Gmail, Twitter, and Facebook is collected and analyzed [60]. First, the data are preprocessed, that is, useless interference packets are removed on the basis of the characteristics of the encrypted data packets. After data is preprocessed, the random forest model is used for offline training. The accuracy of the classification of sub behavior can be higher than 95%. Instant messaging services are rapidly becoming the dominant form of communication among consumers. The behavior of users [61] is analyzed by only observing the TCP payload length and packet direction when users send information. Using the above classification analysis of user actions, the user's communication behavior and communication relationship can be accurately known, which poses a serious threat to the user's privacy and security.

A set of traffic features that can accurately reflect traffic characteristics must be designed for traffic classification based on traditional machine learning. Moreover, the quality of the feature set directly affects classification performance. Therefore, how to design a suitable traffic feature set is an important problem for traffic classification based on traditional machine learning. Although many researchers have been working on this problem in recent years, it is still unsolved.

Deep learning [62] is an effective way to solve the problem of feature design in traditional machine learning. Deep learning is a machine learning technology based on the idea of characterization learning. Characterization learning [63], also known as feature learning, refers to the automatic learning of features directly from the original data to avoid the problem of artificial feature design. It is a branch of machine learning that has arisen in recent years. Deep learning is a typical representation of characterization learning, which is a method using deep neural networks.

As shown in Fig. 6, traffic classification based on deep learning can automatically acquire traffic features through layer-by-layer learning from the original input data, which is different than machine learning. Feature matching is performed on the unknown encrypted traffic by the learned traffic characteristics, and the traffic that meets the characteristics is identified and finally classified.

Encrypted traffic is classified based on the deep learning method of convolutional neural networks [64]–[66]. The specific process shown in Fig. 7 includes three stages: data preprocessing, model training, and model testing. For the preprocessing phase, the input raw data is processed and sorted into the format required by the CNN model for the input data. In the training phase, the steps of feature design, feature extraction, and feature selection are integrated into the minimum batch random gradient descent method. The test phase uses the fine-tuned CNN model constructed in the training phase to predict the labels of traffic data generated in the preprocessing phase and ultimately acquires the classification results. Traffic characteristics can be automatically learned directly from the original traffic, and the advanced traffic features learned layer-by-layer are classified directly in the softmax layer. The feature learning based on deep learning [64]–[66] is completely transferred to the neural network to learn. Compared with traffic classification based on traditional machine learning, it has a synergistic effect, and it is more likely to obtain a better global solution. However, network traffic is combined by the structure of traffic bytes, data packets, and network flows. The structural information of network traffic, such as the timing interval of data packets of different network flows are significantly different.

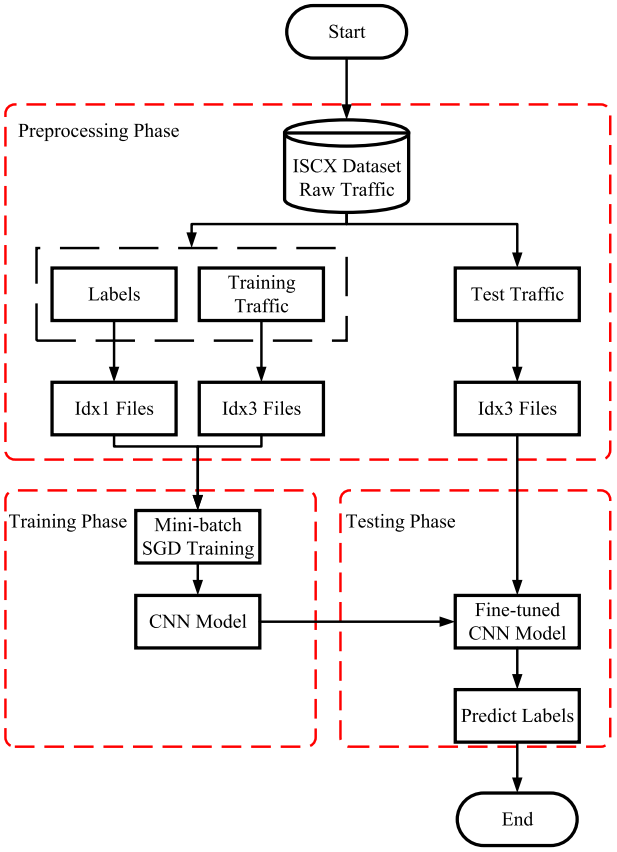

FIGURE 7. Encrypted traffic classification based on a convolutional neural network.

When we extract the timing characteristics of two different data packets, it is best to extract them independently. The above studies [64]–[66] do not make full use of the structured information of network traffic and have a negative impact on classification results.

Through the introduction of encrypted traffic classification methods based on traditional machine learning and deep learning, we have compared and summarized the differences in the core ideas, typical methods, advantages, and disadvantages, which are shown in Table 4.

2) TRAFFIC CLASSIFICATION attacks BASED ON MACHINE LEARNING

With the maturity of traffic classification technology based on machine learning, some attackers use machine learning to identify and classify the traffic transmitted by communication users, thus ultimately obtaining the user's private information or launching malicious attacks on the network, which will have a huge impact on the security of network users. Biggio *et al.* [67] propose to construct aggressive data and put a small amount of aggressive data into training samples for SVM training based on traditional machine learning, which leads to a significant increase in the error rate of SVM in the detection samples. Rimmer *et al.* [68] use deep learning methods to extract and analyze network traffic to identify the communication relationship between the sender and receiver, which eliminates the anonymous security of users.

By collecting and monitoring the original information, attacks based on machine learning do not need to discover the content of the data package, but rather use machine learning technology to classify the traffic and identify relationships between communication parties. This kind of attack is very difficult to detect, and attack tools constructed by machine learning can easily evade existing defense facilities. In addition, some attackers identify the communication relationship between the sender and the receiver by monitoring the traffic of their neighbors. For example, Levine *et al.* [69] monitor the traffic information of neighbor nodes by the characteristic of the fixed HTL value when the sender initiates communication, and it uses Bayesian methods to judge whether a user is the sender.

Attacks based on machine learning increase the difficulty of detection and defense. There are few means of addressing the type of attacks in the current research. The problem about how to defend against attacks based on machine learning must be studied more effectively.

3) DEFENSES AGAINST TRAFFIC CLASSIFICATION ATTACKS BASED ON MACHINE LEARNING

With the rapid development of machine learning technology, machine learning has been used to extract and classify users' data to achieve the deanonymization [70] of anonymous communication, which poses a new challenge to the security of anonymous communication systems.

Since machine learning uses the characteristics of network traffic to classify and deanonymize traffic between users, the key idea to prevent traffic classification attacks is to remove the characteristics of traffic associated with users, including packet size distribution, packet order, flow rate, and flow time. Traffic camouflage technology obscures the characteristics of traffic and reduces the accuracy of traffic identification. Therefore, we can use traffic camouflage technology to resist traffic classification attacks based on machine learning. Traffic camouflage methods include traffic filling and traffic confusion. Specifically, traffic filling technology can fill the packet size based on a certain strategy to remove packet length features [71]. However, this form introduces additional redundant interference traffic and increases communication overhead and network latency.

To address the overhead and delay problems mentioned above, an anonymous communication system based on traffic confusion is proposed [72] to resist traffic classification attacks. The system can resist traffic classification attacks by batching payload flow from different streams, dividing the payload flow into several paths, and adding an artificial delay or artificial redundant flow to the payload flow. At the same time, the system also incorporates a novel dynamic handover strategy for traffic confusion, which can make use of the temporal and spatial correlation between different clients' payload traffic to ensure the anonymity of users with a low

TABLE 4. Comparison of traffic classification methods based on machine learning.

Type	Core ideas	Typical methods	References	Advantages	Disadvantages
Based on traditional machine learning	Modeling from historical data for classification based on manually selected features	Supervised learning	[53][56][58] [59][60][61]	• Low computational complexity. • Strong expandability • Strong interpretability.	• Artificially designs the traffic feature set. • Weak generalization ability for complex problems. • Poor real-time classification.
		Semisupervised learning	[55]		
		Unsupervised learning	[54][57][59]		
Based on deep learning	Automatic learning about features directly from the original data	Encrypted traffic classification by automatically learn representative traffic features.	[64]	• Automatically learns the features from raw data. • Avoids manual design features.	• Strong requirements for data processing and computing power. • Do not make full use of the structural information of network traffic.
		Semantic segmentation using fully conventional networks	[65]		
		Malware traffic classification using representation learning	[66]		

overhead. However, this defense method only uses batching or dividing payload flow, the defensive effect is bad.

Therefore, Zhang *et al.* [73] propose to shape the transmission flow, that is, to construct multiple virtual network card addresses and dynamically adjust their data packets, change the traffic characteristics of the data packets flowing through them, and interfere with attackers. Consequently, attackers are not able to monitor and analyze the traffic to obtain private information about communication users. Further, this method does not add redundant traffic to interfere with the attackers, so it reduces the additional communication overhead.

With the continuous development of machine learning technology, the ability to resist interference and noise data in networks is constantly improved, which brings many challenges [74] in securing defense technology against traffic classification. To resist the negative effect of artificial intelligence technology on anonymous communication, traffic-mimicking confusion technologies have attracted increasing attention from researchers.

Traffic-mimicking confusion technologies [75] disguise the characteristics of real traffic, reduce recognition accuracy based on the methods of flow statistics features, and protect users' communication anonymity and privacy. In [76], a convex optimization method for modifying data packets in real time is proposed. The packet size distribution of the original traffic is disguised as the packet size distribution of other traffic. The transformed traffic can effectively avoid identification by traffic classifiers such as VoIP and Web. However, this paper only disguises the packet size distribution. We can also disguise other features of data packets.

Wu *et al.* [77] confuse the interval, order, length and other characteristics of data packets to resist traffic classification. The detailed confusion process is shown in Fig. 8. First, the method employs a traffic sniffer on the Android platform to intercept user communication traffic and provide entrance traffic to the obfuscation engine. Second, the obtained user traffic is transmitted to the obfuscation engine as input traffic. There are four basic methods built into the obfuscation engine: traffic distribution fitting, time interval obfuscation, packet order obfuscation, and length obfuscation. The user can adjust the parameters of the four basic methods to adjust the degree of confusion. The engine obfuscates user communication traffic according to the formulated obfuscation strategy, and the traffic output from the engine is used as the input traffic by the traffic generator. The traffic generator randomly inserts unrelated traffic into the input traffic and finally sends the traffic to the destination server IP through the wireless network card of the Android terminal. The confused packets are difficult to classify by existing machine learning methods, which protects user's anonymous securities.

The traffic-mimicking confusion technologies in the above works disguise normal traffic to prevent attackers from classifying transmitted traffic according to the statistical characteristics of the traffic and identifying the relevant private information of a communication user. In contrast, the interference flow [78] is disguised as normal traffic, misleading the attacker into judging incorrectly. In this paper, an anti-network GAN adaptively adjusts the traffic characteristics of a learning target network to generate data traffic that is indistinguishable from the targeted traffic, so that the attacker cannot distinguish which is the real target traffic, to resist classification attacks based on machine learning.

Traffic camouflage technology disguises the characteristics of traffic associated with users, such as packet size distribution, packet order, flow rate, and flow time. Traffic camouflage technology includes traffic filling and traffic

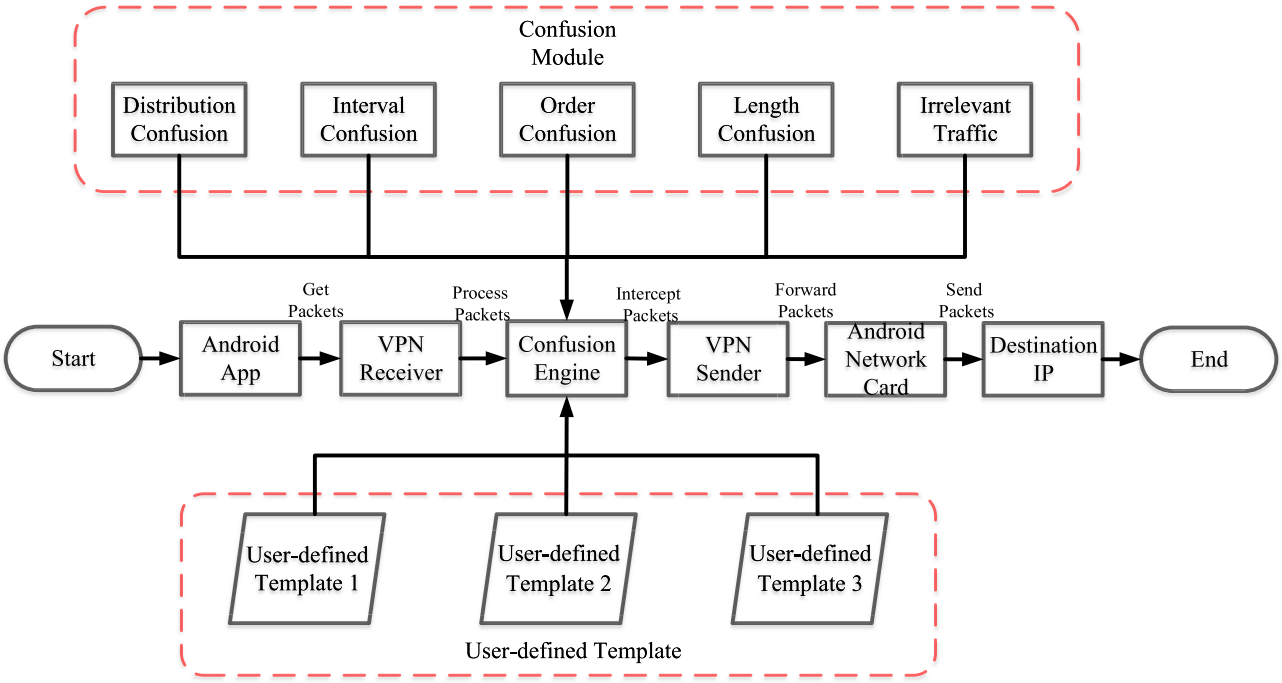

FIGURE 8. Traffic confusion against traffic classification attacks based on machine learning.

TABLE 5. Defenses against traffic classification attacks based on machine learning.

Typical Type	References	Core Ideas	Advantages	Disadvantages
Traffic filling	[71]	Add redundant fake packets to hide real traffic.	• Fills packet size to remove packet length features.	• Introduces additional overhead. • Only for certain types of attacks.
Traffic-mimicking confusion	[72][73] [75][76] [77][78]	Change the distribution, interval, order, length, and other characteristics of data packets.	• Disguises transmission traffic, misleads attackers to judge incorrectly.	• Increases traffic confusion overhead.

confusion. The computational complexity of traffic camouflage technology depends on the method of disguise and the degree of disguise. If the computational cost of the method used for traffic camouflage is greater and the amount of camouflage is greater, the computational complexity of traffic camouflage is greater. Through the above introduction to defend against traffic classification attacks based on machine learning, we summarize their differences based on core ideas, advantages, and disadvantages, as shown in Table 5.

This section proposes a traffic camouflage technology to defend against traffic classification attacks based on machine learning. Traffic camouflage technology changes the statistical characteristics of transmitting traffic to prevent attackers from classifying users' traffic based on statistical traffic characteristics. These changes remove the original statistical characteristics of the traffic associated with the user, which invalidate traffic classification attacks based on machine learning, prevent information transmitted by the user from being monitored and leaked, and protect the privacy and security of the user.

Through the investigation to network security of user on-demand anonymous communication on the transmission side, we summarize security issues and the corresponding defenses of user on-demand anonymous communication on the transmission side in Table 6.

IV. OPEN RESEARCH CHALLENGES

In an untrusted Internet environment, how to design a secure anonymous communication system that meets the different anonymity needs of users and protects user privacy faces many problems and challenges, as discussed below.

TABLE 6. Security issues and corresponding defenses against traffic classification attacks.

Security Issues	Description	Solutions
Traffic classification attacks	Attackers use machine learning methods to classify user's traffic, identify the relationship between communication parties, and launch attacks on users.	• Traffic filling: add artificial redundant flow to the payload flow. • Traffic confusion: change the distribution, interval, order, length and other features of data packets to confuse attackers.
	Attackers use deep learning methods to extract user's traffic for identifying the communication relationship between users.	

A. USER SIDE: INADEQUATE SECURITY CONSIDERATIONS FOR USERS' IDENTITY ANONYMITY

At present, most of the research focuses on users' identity anonymity or network behavior tracking. When users communicate anonymously by using anonymous identities, the communication system should not only ensure an anonymous experience, but also be able to track the behavior of users in the network to ensure the security of anonymous communication. The self-verifying identifier mechanism [23] uses a third-party proxy with authentication capabilities to implement user's various anonymity needs and network behavior tracking. However, due to the introduction of the third-party proxy, it does not take into account the security problems the proxy introduces. For example, if the third party is a malicious attacker or has been compromised, it will pose a serious threat to the user's anonymous security. Therefore, when we use new identifiers to meet users' diverse anonymity needs and track network behavior, we should consider avoiding the security risks brought by a third party.

B. TRANSMISSION SIDE: LACK OF SECURITY CONSIDERATIONS FOR HIERARCHICAL ANONYMOUS TRANSMISSION

In practice, many components of anonymous systems, such as network structures and communication components, change over time [79]. In particular, the communication performance of routing nodes in a network, such as bandwidth, communication delay, etc., will change over time. Existing hierarchical anonymous communication does not consider the impacts of routing nodes over time on communication anonymity, and when multiple frequently used intermediate routing nodes collude, partial transmission paths may be leaked. In addition, the transmission paths are vulnerable to traffic analysis attacks.

C. THREATS BASED ON MACHINE LEARNING TO USERS' PRIVACY

In recent years, with the popularization of smart terminals represented by mobile phones, user's personal will is carried by user's behavior traffic in the applications of user's phones, which becomes a loophole in user's behavior analysis, data mining, and user's privacy leakage. With the continuous development of machine learning technologies [80], [81], the accuracy of traffic classification is improved. An increasing number of attackers have applied machine learning methods to classify and predict user's network behaviors, which causes great threats to the privacy of users.

D. INSUFFICIENT DEFENSE AGAINST TRAFFIC CLASSIFICATION ATTACKS BASED ON MACHINE LEARNING

In anonymous communication, users encounter various attackers with different performances, and the locations of these attackers in the transmission paths are different. The attackers monitor and analyze the transmitted user traffic, which poses a great threat to the anonymous security of users. Current anonymous communication systems typically use encryption algorithms to protect transmitted data. However, traffic classification attacks based on machine learning can analyze traffic characteristics, extract traffic features and classify user traffic, even if the data is encrypted. Moreover, this kind of attack has some degree of immunity against noise interference with traffic, and it has a high classification accuracy for anonymous traffic, which poses a great danger to the privacy of users.

V. FUTURE TRENDS

Through the analysis of open problems in user on-demand anonymous communication, studying the future trends are composed of several aspects as follows:

A. DESIGN OF DECENTRALIZED IDENTITY ANONYMITY AND BEHAVIOR TRACKING

Decentralized anonymous design is the result of mutual authentication of multiple users in the network. Even if some nodes in the network collude with each other, they cannot obtain the private authentication information of all verified nodes, so it is impossible to infer the identities of communication parties, which protects users' anonymous securities

and avoids security risks due to compromise or failure of the third party when using third-party authentication.

B. ANONYMOUS TRANSMISSION BY ADAPTIVE DYNAMIC ROUTING

Many of the recently proposed anonymous communication protocols, such as the anonymous Internet routing protocols Dovetail [82] and HORNET [83], ignore the impact of dynamic changes on important system components over time in their security analysis. Therefore, we must strongly consider the effect of time on anonymity in communication transmission and adaptively adjust the routing transmission path according to the usage frequency of the network components so that the path dynamically changes across time and space and achieves high flexibility, high security, and high scalability. Moreover, the usage characteristics of identical network nodes ensure the secure and anonymous transmission of user information to the greatest extent.

C. ANONYMOUS ENHANCEMENT TECHNOLOGY AGAINST MACHINE LEARNING ATTACKS

If a network user performs different communication services at different times, their corresponding traffic characteristics will be distinct from the traffic characteristics of previous training and learning. Then, if the original method of machine learning is used to classify the user's traffic, there will be deviation [84], leading to errors in judgment. Since traffic classification methods based on machine learning perform sampling and learning of specific traffic at a specific time in a specific scenario, they have certain special characteristics. They can only recognize traffic in specific circumstances and have no universality, which limits the scope of their use. Therefore, with dynamic changes in time and user location, intelligent traffic confusion is used to dynamically process the user traffic transmitted, which counteracts traffic classification methods based on machine learning and effectively protects the user's privacy.

D. OPTIMIZE SECURE DEFENSE STRATEGIES

In the process of anonymous communication, various attacks will inevitably be encountered. The characteristics of various attacks are analyzed in detail, and corresponding secure defense strategies are designed. A security assessment of the defense strategies is carried out, and the defense schemes are optimized to find an optimal secure defense strategy that can effectively resist multiple traffic analysis attacks.

VI. FINAL REMARKS

We review the privacy protection and network security in user on-demand anonymous communication from the user side and transmission side. For the user side, the major security issue is abuse of anonymous communication, and the corresponding solution is a mechanism for balancing anonymous communication and behavior tracking to prevent anonymous abuse. For the transmission side, the major security issues are traffic classification attacks based on machine learning, and the potential countermeasures are traffic camouflage technologies to disguise users' transmitted information. However, some open challenges are remaining for research on user on-demand anonymous communication; thus, we provide some suggestions for future research directions.

With the continuous development of machine learning, greater challenges are posed to privacy protection and network security in user on-demand anonymous communication. The focus of future work may change from passive defense to active defense, and more adaptive and dynamic defense mechanisms should be designed for the user on-demand anonymous communication. Furthermore, we will start to implement dynamic defense mechanisms with random route mutations on our school campus. Finally, we will consider more comprehensive approaches to improve network security while ensuring user privacy in anonymous communication by considering a combination of dynamic security defense, traffic camouflage, and anonymous communication.

ACKNOWLEDGMENT

The authors would like to thank all the reviewers for the valuable suggestions.

REFERENCES

[1] T. Lu, Y. Zhou, and Y. Liu, "Review on problem of Internet user privacy leakage," *Comput. Sci.*, vol. 41, no. 10, pp. 62–67 2014.
[2] A. Ahmad and B. Hawashin, "A secure network communication protocol based on text to barcode encryption algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 12, pp. 64–70, 2015.
[3] A. Johnson, C. Wacek, R. Jansen, M. Sherr, and P. Syverson, "Users get routed: Traffic correlation on Tor by realistic adversaries," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. - CCS*, 2013, pp. 337–348.
[4] P. Mittal and N. Borisov, "Information leaks in structured peer-to-peer anonymous communication systems," *ACM Trans. Inf. Syst. Secur.*, vol. 15, no. 1, pp. 1–28, Mar. 2012.
[5] T. Shu, Y. Chen, and J. Yang, "Protecting multi-lateral localization privacy in pervasive environments," *IEEE/ACM Trans. Netw.*, vol. 23, no. 5, pp. 1688–1701, Oct. 2015.
[6] M. Sherr, H. Gill, T. A. Saeed, A. Mao, W. R. Marczak, S. Soundararajan, W. Zhou, B. T. Loo, and M. Blaze, "The design and implementation of the A3 application-aware anonymity platform," *Comput. Netw.*, vol. 58, pp. 206–227, Jan. 2014.
[7] W. Liu and M. Yu, "AASR: Authenticated anonymous secure routing for MANETs in adversarial environments," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4585–4593, Nov. 2014.
[8] S. Le Blond, D. Choffnes, W. Caldwell, P. Druschel, and N. Merritt, "Herd: A scalable, traffic analysis resistant anonymity network for VoIP systems," in *Proc. ACM Conf. Special Interest Group Data Commun. (SIGCOMM)*, 2015, pp. 639–652.
[9] G. T. K. Nguyen, "Performance and security tradeoffs of provable Website traffic fingerprinting defenses over Tor," Ph.D. dissertation, Dept. Comput. Sci., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 2017.
[10] S. Chakravarty, "Traffic analysis attacks and defenses in low latency anonymous communication," Ph.D. dissertation, Dept. Arts Sci., Columbia Univ., New York, NY, USA, 2014.
[11] T. Baumeister, Y. Dong, Z. Duan, and G. Tian, "A routing table insertion (RTI) attack on freenet," in *Proc. Int. Conf. Cyber Secur.*, Dec. 2012, pp. 8–15.
[12] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," *IEEE Access*, vol. 6, pp. 24411–24432, 2018.
[13] F. Shirazi, M. Simeonovski, M. R. Asghar, M. Backes, and C. Diaz, "A survey on routing in anonymous communication protocols," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–39, Jun. 2018.

[14] J. Luo, M. Yang, Z. Ling, W. Wu, and X. Gu, "Anonymous communication and darknet: A survey," *J. Comput. Res. Develop.*, vol. 56, no. 1, pp. 103–130, 2019.
[15] K. Chen and H. Shen, "Fine-grained encountering information collection under neighbor anonymity in mobile opportunistic social networks," in *Proc. IEEE 23rd Int. Conf. Netw. Protocols (ICNP)*, Nov. 2015, pp. 179–188.
[16] T. Narten, R. Draves, and S. Krishnan, *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*, document 4941, 2007.
[17] S. Han, V. Liu, Q. Pu, S. Peter, T. Anderson, A. Krishnamurthy, and D. Wetherall, "Expressive privacy control with pseudonyms," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 291–302, Aug. 2013.
[18] A. Sakurai, T. Minohara, R. Sato, and K. Mizutani, "One-time receiver address in IPv6 for protecting unlinkability," in *Annu. Asian Comput. Sci. Conference.*, vol. 2007, pp. 240–246.
[19] M. Yang, X. Gu, Z. Ling, C. Yin, and J. Luo, "An active de-anonymizing attack against Tor Web traffic," *Tsinghua Sci. Technol.*, vol. 22, no. 6, pp. 702–713, Dec. 2017.
[20] G. Horsman, "The challenges surrounding the regulation of anonymous communication provision in the United Kingdom," *Comput. Secur.*, vol. 56, pp. 151–162, Feb. 2016.
[21] Y. Mallios, S. Modi, A. Agarwala, and C. Johns, "Persona: Network layer anonymity and accountability for next generation Internet," in *Proc. IFIP Int. Inf. Secur. Conf.*, 2009, pp. 410–420.
[22] A. Venkataramani, J. F. Kurose, D. Raychaudhuri, K. Nagaraja, M. Mao, and S. Banerjee, "MobilityFirst: A mobility-centric and trustworthy Internet architecture," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 74–80, Jul. 2014.
[23] T.-H. Lee, "Towards an accountable and private Internet," Ph.D. dissertation, ETH Zürich, Zürich, Switzerland, 2017.
[24] D. I. Wolinsky, H. Corrigan-Gibbs, B. Ford, and A. Johnson, "Dissent in numbers: Making strong anonymity scale," presented at the 10th USENIX Symp. Oper. Syst. Design Implement. (OSDI), 2012.
[25] H. G. L. Junzhou, "G-hordes: A safe anonymous communication system," *J. Southeast Univ. (Natural Sci. Ed.)*, vol. 39, no. 2, pp. 220–224, Mar. 2009.
[26] J. Buchmann, D. Demirel, and J. Van De Graaf, "Towards a publicly-verifiable mix-net providing everlasting privacy," in *Proc. Int. Conf. Financial Cryptogr. Data Secur.*, 2013, pp. 197–204.
[27] A. Panchenko and J. Renner, "Path selection metrics for performance-improved onion routing," in *Proc. 9th Annu. Int. Symp. Appl. Internet*, Jul. 2009, pp. 114–120.
[28] X. Jing, W. Zhenxing, Z. Liancheng, and W. Qian, "Recipient anonymity: An improved crowds protocol based on key sharing," in *Proc. WASE Int. Conf. Inf. Eng.*, Aug. 2010, pp. 60–64.
[29] M. Backes, A. Kate, S. Meiser, and E. Mohammadi, "(Nothing else) MATor(s): Monitoring the anonymity of Tor's path selection," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2014, pp. 513–524.
[30] H. Xu, J. Chen, and S. Chen, "A scalar anonymous system model based on DC-nets," *MINI-MICRO Syst.*, vol. 27, no. 3, pp. 461–465 2006.
[31] H. Corrigan-Gibbs and B. Ford, "Scavenging for anonymity with BlogDrop," in *Proc. Provable Privacy Workshop*, 2012, pp. 1–4.
[32] H. Xu, X. Fu, Y. Zhu, R. Bettati, J. Chen, and W. Zhao, "SAS: A scalar anonymous communication system," in *Proc. Int. Conf. Netw. Mobile Comput.*, 2005, pp. 452–461.
[33] C. Shi, X. Luo, P. Traynor, M. H. Ammar, and E. W. Zegura, "ARDEN: Anonymous networking in delay tolerant networks," *Ad Hoc Netw.*, vol. 10, no. 6, pp. 918–930, Aug. 2012.
[34] F. Aseez and S. Mathew, "Hierarchical partition-based anonymous routing protocol (HPAR) in MANET for efficient and secure transmission," *Int. J. Cybern. Informat.*, vol. 5, no. 2, pp. 417–425, Apr. 2016.
[35] Q. Ren, "Research and realization of trusted scalar anonymous communication system in P2P network," M.S. thesis, Dept. Inf. Eng., Wuhan Univ. Technol., Wuhan, China, 2009.
[36] X. Chen, H. Hu, B. Liu, F. Xiao, and Z. Huang, "ALHACF: An anonymity-level selected hierarchical anonymous communication framework," in *Proc. IEEE/IFIP Int. Conf. Embedded Ubiquitous Comput.*, Dec. 2008, pp. 251–256.
[37] X. Chen, Q.-L. Hu, H. Lu, G.-M. Xia, and J. Long, "ILACF: An incentive-based low-latency anonymous communication framework," in *Proc. IEEE 11th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Jun. 2012, pp. 964–969.
[38] Y. Zhou, Z. Wu, and B. Yang, "Diversity of controllable anonymous communication system," *J. Commun.*, vol. 36, no. 6, pp. 105–115 2015.
[39] S. Remya and K. S. Lakshmi, "SHARP: Secured hierarchical anonymous routing protocol for MANETs," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2015, pp. 1–6.
[40] C. Stergiou, K. E. Psannis, A. P. Plageras, Y. Ishibashi, and B.-G. Kim, "Algorithms for efficient digital media transmission over IoT and cloud networking," *J. Multimedia Inf. Syst.*, vol. 5, no. 1, pp. 1–10, Mar. 2018.
[41] V. Memos, K. E. Psannis, Y. Ishibashi, B.-G. Kim, and B. Gupta, "An efficient algorithm for media-based surveillance system (EAMSuS) in IoT smart city framework," *Future Gener. Comput. Syst.*, vol. 83, pp. 619–628, Jun. 2018.
[42] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient IoT-based sensor BIG data collection–processing and analysis in smart buildings," *Future Gener. Comput. Syst.*, vol. 82, pp. 349–357, May 2018.
[43] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of IoT and cloud computing," *Future Gener. Comput. Syst.*, vol. 78, pp. 964–975, Jan. 2018.
[44] K. E. Psannis, C. Stergiou, and B. B. Gupta, "Advanced media-based smart big data on intelligent cloud systems," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 1, pp. 77–87, Jan. 2019.
[45] A. Panchenko, L. Niessen, A. Zinnen, and T. Engel, "Website fingerprinting in onion routing based anonymization networks," in *Proc. 10th Annu. ACM Workshop Privacy Electron. Soc. (WPES)*, 2011, pp. 103–114.
[46] P. Mittal, A. Khurshid, J. Juen, M. Caesar, and N. Borisov, "Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting," in *Proc. 18th ACM Conf. Comput. Commun. Secur. (CCS)*, 2011, pp. 215–226.
[47] H. S. Anderson, J. Woodbridge, and B. Filar, "DeepDGA: Adversarially-tuned domain generation and detection," in *Proc. ACM Workshop Artif. Intell. Secur. (ALSec)*, 2016, pp. 13–21.
[48] G. Hospodar, R. Maes, and I. Verbauwhede, "Machine learning attacks on 65 nm arbiter PUFs: Accurate modeling poses strict bounds on usability," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2012, pp. 37–42.
[49] W. Xu, Y. Qi, and D. Evans, "Automatically evading classifiers," in *Proc. Netw. Distrib. Syst. Symposium.*, 2016, pp. 21–24.
[50] M. Nasr, A. Bahramali, and A. Houmansadr, "DeepCorr: Strong flow correlation attacks on Tor using deep learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Jan. 2018, pp. 1962–1976.
[51] B. Ma, H. Zhang, Y. Guo, Z. Liu, and Y. Zeng, "A summary of traffic identification method depended on machine learning," in *Proc. Int. Conf. Sensor Netw. Signal Process. (SNSP)*, Oct. 2018, pp. 469–474.
[52] P. Velan, M. Čermák, P. Čeleda, and M. Drašar, "A survey of methods for encrypted traffic classification and analysis," *Int. J. Netw. Manage.*, vol. 25, no. 5, pp. 355–374, Sep. 2015.
[53] Y. Kumano, S. Ata, N. Nakamura, Y. Nakahira, and I. Oka, "Towards real-time processing for application identification of encrypted traffic," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 136–140.
[54] P. V. Amoli and T. Hämäläinen, "A real time unsupervised NIDS for detecting unknown and encrypted network attacks in high speed network," in *Proc. IEEE Int. Workshop Meas. Netw. (M&N)*, Oct. 2013, pp. 149–154.
[55] Y. Du and R. Zhang, "Design of a method for encrypted P2P traffic identification using K-means algorithm," *Telecommun. Syst.*, vol. 53, no. 1, pp. 163–168, May 2013.
[56] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for Internet traffic classification," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 223–239, Jan. 2007.
[57] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and vpn traffic using time-related," in *Proc. 2nd Int. Conf. Inf. Syst. Secur. Privacy (ICISSP).*, 2016, pp. 407–414.
[58] V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, "AppScanner: Automatic fingerprinting of smartphone apps from encrypted network traffic," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P)*, Mar. 2016, pp. 439–454.
[59] J. Zhang, X. Chen, Y. Xiang, W. Zhou, and J. Wu, "Robust network traffic classification," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1257–1270, Aug. 2015.
[60] M. Conti, L. V. Mancini, R. Spolaor, and N. V. Verde, "Analyzing Android encrypted network traffic to identify user actions," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 114–125, Jan. 2016.
[61] S. E. Coull and K. P. Dyer, "Traffic analysis of encrypted messaging services: Apple iMessage and beyond," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 5–11, Oct. 2014.

[62] S. Rezaei and X. Liu, "Deep learning for encrypted traffic classification: An overview," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 76–81, May 2019.

[63] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[64] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Jul. 2017, pp. 43–48.

[65] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[66] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, 2017, pp. 712–717.

[67] B. Biggio, B. Nelson, and P. Laskov, "Poisoning attacks against support vector machines," in *Proc. 29th Int. Conf. Int. Conf. Mach. Learn.*, 2012, pp. 1467–1474.

[68] V. Rimmer, D. Preuveneers, M. Juarez, T. V. Goethem, and W. Joosen, "Automated Website fingerprinting through deep learning," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, Feb. 2018, pp. 1–16.

[69] B. N. Levine, M. Liberatore, B. Lynn, and M. Wright, "Statistical detection of downloaders in freenet," in *Proc. 3rd IEEE Int. Workshop Privacy Eng.*, 2017, pp. 1–8.

[70] G. He, M. Yang, J. Luo, and X. Gu, "Inferring application type information from Tor encrypted traffic," in *Proc. 2nd Int. Conf. Adv. Cloud Big Data*, Nov. 2014, pp. 220–227.

[71] X. Cai, R. Nithyanand, T. Wang, R. Johnson, and I. Goldberg, "A systematic approach to developing and evaluating Website fingerprinting defenses," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2014, pp. 227–238.

[72] S. Le Blond, D. Choffnes, W. Zhou, P. Druschel, H. Ballani, and P. Francis, "Towards efficient traffic-analysis resistant anonymity networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 303–314, Aug. 2013.

[73] F. Zhang, W. He, and X. Liu, "Defending against traffic analysis in wireless networks through traffic reshaping," in *Proc. 31st Int. Conf. Distrib. Comput. Syst.*, Jun. 2011, pp. 593–602.

[74] Y. Wang, H. Zhou, F. Hao, M. Ye, and W. Ke, "Network traffic classification method basing on CNN," *J. Commun.*, vol. 39, no. 01, pp. 14–23 2018.

[75] Y. Zhou, S. Wagh, P. Mittal, and D. Wentzlaff, "Camouflage: Memory traffic shaping to mitigate timing attacks," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2017, pp. 337–348.

[76] C. V. Wright, S. E. Coull, and F. Monrose, "Traffic morphing: An efficient defense against statistical traffic analysis," in *Proc. NDSS*, 2009, pp. 1–14.

[77] C. Wu, J. Wang, and S. Wei, "Traffic obfuscation process of Android APPS against machine learning detection," *Comput. Appl. Softw.*, vol. 35, no. 11, pp. 301–308 2018.

[78] J. Li, L. Zhou, H. Li, L. Yan, and H. Zhu, "Dynamic traffic feature camouflaging via generative adversarial networks," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2019, pp. 1–9.

[79] R. Wails, Y. Sun, A. Johnson, M. Chiang, and P. Mittal, "Tempest: Temporal dynamics in anonymity systems," *Proc. Privacy Enhancing Technol.*, vol. 2018, no. 3, pp. 22–42, Jun. 2018.

[80] L. He, C. Xu, and Y. Luo, "VTC: Machine learning based traffic classification as a virtual network function," in *Proc. ACM Int. Workshop Secur. Softw. Defined Netw. Netw. Function Virtualization (SDN-NFV Security)*, 2016, pp. 53–56.

[81] H. Chu and X. Zhang, "Application of a hybrid feature selection algorithm in Internet traffic identification," *J. Chin. Comput. Syst.*, vol. 33, no. 2, pp. 325–329, 2012.

[82] J. Sankey and M. Wright, "Dovetail: Stronger anonymity in next-generation Internet routing," in *Proc. Int. Symp. Privacy Enhancing Technol. Symp.*, 2014, pp. 283–303.

[83] C. Chen, D. E. Asoni, D. Barrera, G. Danezis, and A. Perrig, "HORNET: High-speed onion routing at the network layer," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur. $_{CCS}$*, Denver, CO, USA, Oct. 2015, pp. 1441–1454.

[84] H. Zhang, G. Lu, M. T. Qassrawi, Y. Zhang, and X. Yu, "Feature selection for optimizing traffic classification," *Comput. Commun.*, vol. 35, no. 12, pp. 1457–1471, Jul. 2012.

YUN HE received the M.S. degree from the China University of Mining and Technology at Beijing, Beijing, China, in 2016. She is currently pursuing the Ph.D. degree in communication and information systems with the University of Science and Technology Beijing, Beijing. Her research interests include security of network and information, anonymous communication, and privacy protection.

MIN ZHANG is currently an Associate Professor with the School of Computer and Communication Engineering (SCCE), University of Science and Technology Beijing, Beijing, China. Her researches focus on content distribution networking and network security. She has fulfilled more than ten research projects, including the National Natural Science Foundation of China and National Hi-Tech Research and Development Program (863 Program). In her research field, she authored more than 60 articles and holds 30 patents.

XIAOLONG YANG (Member, IEEE) received the B.Eng., M.S., and Ph.D. degrees in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 1993, 1996, and 2004, respectively. He is currently a Professor with the School of Computer and Communication Engineering, Institute of Advanced Networking Technologies and Services, University of Science and Technology Beijing, Beijing, China. He has fulfilled more than 30 research projects, including the National Natural Science Foundation of China, the National Hi-Tech Research and Development Program (863 Program), and the National Key Basic Research Program (973 Program). His current research interests include optical switching and Internetworking and the next-generation Internet. He has authored more than 80 articles. He holds16 patents in these areas.

JINGTANG LUO (Member, IEEE) received the B.Eng. and Ph.D. degrees in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2011 and 2016, respectively. He is currently a Researcher with the State Grid Sichuan Economic Research Institute, Chengdu. His current research interests include congestion control in data center networks and information security. He serves as a Reviewer for international academic journals, including the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and the *Journal of Computer Science and Technology*.

YIMING CHEN received the B.S. and Ph.D. degrees in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2013 and 2018, respectively. He is currently a Researcher with the State Grid Sichuan Economic Research Institute, Chengdu. His current research interests include the smart control method of distributed power systems and artificial intelligence techniques in the digital power grid.