A Reconstruction Forest-Based Interest Flooding Attack Detection Method in Named Data Networking

Guanglin Xing
College of Computer Science
South-Central Minzu University
Wuhan, China
glxing@scuec.edu.cn

Xiaoqi Li
College of Computer Science
South-Central Minzu University
Wuhan, China
lixiaoqi_stu@163.com

Rui Hou*
College of Computer Science
South-Central Minzu University
Wuhan, China
hourui@mail.scuec.edu.cn

Abstract—Named data networking (NDN) is considered one of the most viable architectures in information-centric networking (ICN) and a promising candidate for future internet architecture. However, interest flooding attacks (IFAs) in NDN can seriously threaten the security of a network. An IFA is one type of distributed denial of service (DDoS) attack in NDN. IFA attackers consume the bandwidth, router cache and computing resources of a network by introducing a large number of malicious interest packets into the network, and the router’s ability to receive and forward packets is reduced or can even fail, thus causing the network to crash. Therefore, a highly efficient detection method is urgently needed to mitigate the harmfulness of IFAs. However, existing IFA detection methods have limitations in terms of detection accuracy and anti-interference capability, especially for traffic fluctuations that may produce false positives, thus reducing IFA detection efficiency. In this paper, a novel IFA detection method named the reconstruction forest-based detection method (RFDM) is proposed to achieve highly efficient attack detection and traffic fluctuation determination. A RecForest is used to calculate the reconstruction error through the reconstruction process, detect malicious interest packets, and mitigate IFAs by restricting malicious interest packet forwarding. The simulation results show that the RFDM outperforms other typical IFA detection methods in terms of detection speed, accuracy, and resistance to network fluctuations.

Index Terms—Named data networking, interest flooding attacks, IFA detection methods

I. INTRODUCTION

Today’s internet traffic and data rates are reaching unprecedented peaks and growing rapidly over time. Users obtain data, and information service providers provide data. The existing method of obtaining data on the internet requires knowing the server where the data are located and accessing the data through the address or host, although users are more concerned about the data content itself. To adapt to the changing needs of users, many researchers have started to study future network architectures, and ICN [1] proposes a data content-centric approach in which users do not need to know where the data are located, changing the communication model of the network from host-centric to data content-centric and changing the way data are accessed. A scheme called NDN [2] has been proposed that may complement the current internet architecture based on the internet protocol (IP). Compared with current internet architectures, NDN has advantages in terms of mobility, security, and scalability. Therefore, NDN is considered the most viable potential architecture for ICN.

In the current internet, distributed denial of service (DDoS) is one of the most damaging network attacks; in this scheme, the attacker requests a large number of host resources to exhaust the network resources, crash the network, and block the access of other users. However, the IFA [3], an evolution from DDoS, suffers in NDN, as it is a resource-related attack in which the attacker forges a large number of interest packets, also known as malicious interest packets, by requesting content data that does not exist and sending them to the network; the information in the interest packets then added to the router's pending interest table (PIT). The attacker attempts to exhaust the router's PIT cache with quickly sending a large number of malicious interest packets. This causes the router to receive an abnormal distribution of requests and a lower satisfaction rate of received interest packets; consumes link bandwidth, computing power, and other network resources; prevents normal users from requesting legitimate data; and potentially causes massive network packet loss, which can
In this paper, we first review the current research on countermeasures for IFAs and then introduce the idea of the RecForest [4], which is used in the field of data mining. Based on the theory of this method, we propose a RecForest-based detection method (RFDM), which can quickly and accurately identify malicious interest packets from attackers and protect the interest packets of normal users while avoiding the problem of misclassification caused by traffic fluctuations. When an IFA is activated, the number of malicious interest packets in the network increases, and since the characteristics of malicious interest packets are different from those of legitimate interest packets, the RFDM constructs a RecForest to detect attacks by collecting the characteristics of interest packets through routers and distinguishing legitimate interest packets from malicious ones by reconstructing the characteristics based on the characteristics of interest packets. This approach can improve the precision and accuracy of detection. The main contributions of this paper are as follows.

1. An IFA detection method based on the RecForest concept is proposed, which could improve the timeliness of identifying attacks.
2. The proposed RFDM can accurately distinguish attacks that are misclassified because of traffic fluctuations and thus promote IFA detection accuracy.
3. The simulation results show that the RFDM has advantages in terms of IFA detection efficiency over other typical related works.

The rest of the article is organized as follows. Section II describes the existing solutions. Section III describes the proposed solution in detail. Section IV describes experimental simulations and discusses the simulation results. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORKS

In NDN, data names are structured like domain addresses separated by “/”, e.g., “/google/video1”, and each piece of data content is assigned a globally unique name. A data content requester, i.e., a consumer, finds the desired data content by creating an interest packet with the desired data name and sending it to the named data network. The data content holder, i.e., producer, encapsulates the data content into a data packet and sends it back to the network, to be returned to the consumer according to the characteristics of the path record of the interest packet in the router, and then the data packet is returned hop-by-hop along the reverse path of the interest packet. Data structures can include a content store (CS), a PIT, and a forwarding information base (FIB). The CS caches the data packets obtained from the producer. When a data packet is obtained, the corresponding interest packet entry is removed from the PIT. The FIB records the forwarding rules between NDN routing nodes and forwards the incoming interest packets to the appropriate interface to search for the producer of the corresponding data packet.

Since NDN relies on interest packets and data packets for forwarding and transmission, IFA attackers can forge a large number of malicious interest packets because there is no corresponding data packet to satisfy the malicious interest packet, and the malicious interest packet will stay in the PIT on the router for some time until the interest packet expires and is then removed from the PIT. When a large number of malicious interest packets enter the network, the router’s PIT cache is quickly filled up, exhausting the cache and network bandwidth and preventing new interest packets from being added, causing network traffic fluctuations and even network crashes, which seriously affects the security of the network. Therefore, the significant threat posed by IFAs has attracted the attention of an increasing number of researchers.

Solutions have been offered for IFAs. Afanasyev et al. [5] proposed satisfaction-based pushback methods to mitigate IFAs. Based on the idea of one interest packet acquiring at most one data packet, we can calculate the interest satisfaction ratio (ISR) compared with a preset threshold and determine whether the router is suffering from an IFA. The ISR is the ratio between the forwarded interest packets and the corresponding data packets on the router. Compagno et al. [6] proposed an IFA coping mechanism called Poseidon, which continuously monitors the ISR and the PIT size for each interface. When these two values exceed the preset threshold, the rate of incoming interest packets is limited. Xue et al. [7] proposed a detection mechanism to identify malicious interest packets by calculating the ISRs of edge routers. When the ISR exceeds a preset threshold, the forwarding of malicious interest packets is restricted. However, the above approach relies on the ISR to detect attacks and is prone to misclassification.

Yi et al. [8] proposed a NACK mechanism for interest packets, where the producer receives an interest packet requesting the nonexistent data name and sends a NACK message to instruct the router to request nonexistent data, and the router releases the PIT entry after receiving the NACK packet. Although the above methods can release PIT entries and reduce PIT occupancy, no other measures, such as rate limiting, are taken to mitigate the attack, resulting in PIT
Zhi et al. [9] proposed an IFA detection mechanism based on an entropy-SVM and Jensen–Shannon divergence by collecting IFA data, using the SVM for feature extraction, detecting whether an IFA occurs according to the features, and finally identifying malicious interest packets by Jensen–Shannon divergence. Zhou et al. [10] proposed a mechanism for detecting an IFA using deep reinforcement learning by collecting IFA features, and setting and training an agent to perform IFA detection; if an IFA is detected, the agent performs relevant restriction operations on the router. The two above methods utilize an entropy-SVM. With deep reinforcement learning, both need to collect IFA data and train offline to perform IFA detection and cannot be trained online in real time; IFA detection will fail if the IFA features are not updated in time.

Benmoussa et al. [11] proposed a congestion-aware-based IFA detection method by detecting the ISR, interface receive and send rates and network congestion conditions and combining the values of the three to determine whether an interface is malicious. The above methods are susceptible to changes in one of the values, which can affect the results.

Xin et al. [12] proposed an IFA detection method based on information entropy, where cumulative entropy is used to calculate the distribution of interest packet names and relative entropy is used to identify malicious interest packets. Zhi et al. [13] proposed a Gini impurity-based IFA defense mechanism, which determines whether an attack is occurring by calculating whether the cumulative entropy and relative entropy have reached a preset threshold. Hou et al. [14] proposed a Theil-based IFA defense mechanism, which randomly divides the interest packets within a PIT into intragroup and intergroup packets and detects IFAs by comparing the differences between groups. The above three methods are all more advanced detection methods, based on the mechanism of the statistical name distribution; when a large number of normal interest packets enter the router and the name distribution changes rapidly, these methods can cause attack misclassification. Vassilakis et al. [15] proposed an IFA mitigation scheme that is based on the number of expired PIT entries for each user, and then take appropriate mitigation measures based on the identity of the user.

Since the existing methods use the router interface received interest packet rate, ISR and name distribution as the basis for classification, the received interest packet rate will lead to misjudgment when the interface rate increases suddenly because of a burst in normal traffic in the network. The ISR indicator is relatively singular, and when the network experiences a burst in normal traffic, the network may become congested, and the ISR will decrease, which leads to misclassification. The above two methods do not identify the real malicious packets of interest and make accurate restrictions, and the network quality for normal users will remain at a low level until the attack stops. The name distribution method can identify true malicious packets of interest under certain conditions, but the name distribution in the router also changes rapidly when the network experiences a burst in normal traffic; therefore, this method may misclassify the burst of normal traffic as a malicious attack and cause the network quality of normal users to be affected.

To overcome the problem of misclassification caused by traffic fluctuations, reduce misjudgments, and improve the accuracy of attack detection, we introduce the RecForest concept in this paper and propose the RFDM. The RFDM uses the states of the interest packets in the PIT (i.e., the number of interest packets sent and the number of packets received) as the basis for classification, and it identifies malicious interest packets through the RecForest method. The RFDM is used to detect malicious interest packets based on the following scenario considerations: in normal burst traffic, normal users send a large number of legitimate interest packets, which may cause network congestion; however, network congestion does not cause legitimate interest packets to fail to obtain data packets but prolongs the time needed to obtain data packets. Therefore, by using the different status information of legitimate and malicious interest packets and setting a reasonable detection period, the RFDM can reduce the misjudgment of normal burst traffic and set accurate restrictions on malicious interest packets, improving the network quality of normal users during the attack. Compared with previously proposed schemes, the RFDM can better address the problems of misclassification and poor identification accuracy due to traffic fluctuations.

III. THE RECONSTRUCTION FOREST-BASED DETECTION METHOD

In NDN routers, each router can record the received interest packets and the outgoing data packets. When an IFA occurs, malicious interest packets will stay in the PIT until they expire because there is no corresponding data packet, while normal interest packets will receive the corresponding data packets and therefore will not stay in the PIT for a long time. The resulting difference allows us to detect these malicious interest packets by collecting the relevant data.
A. The RecForest Principle

The RecForest reconstructs the data with a bounding box, which is a rectangular box of an interest packet data sample \( x \). It limits the neighboring region of sample \( x \). Given a bounding box, the reconstructed point of \( x \) is defined as the center point of the bounding box. The bounding box has an upper bound and a lower bound, and multiple bounding boxes are obtained by traversing the RecForest. The smallest bounding box can be obtained from the upper bounds and lower bounds of multiple bounding boxes; the center point of the smallest bounding box is the reconstruction point of the sample, and the reconstruction error is obtained by computing the origin of the reconstruction point and the sample \( x \). The reconstruction error shows how much the sample deviates from the sample origin. Additionally, since the outlying attributes of the RecForest are identifiable, the error of each attribute of the reconstruction error can be calculated, and misclassification can be reduced by comparing each attribute’s error.

B. Data Collection

In a given detection period, the router collects the number of received interest packets (InInterestNum) and the number of outgoing data packets (OutDataNum) in real time. An interest packet dataset is then constructed based on these two statistical metrics as two attributes of the interest packets. For example, during a detection period, if a route receives 2 interest packets with the name “/google.com/video1” and sends out 1 data packet with the same name, then the interest packet with the name “/google.com/video1” can be constructed as follows: InInterestNum = 2, OutDataNum = 1.

C. Constructing the RecForest

When the interest packet data are collected, the construction of a RecForest is initiated, where a RecTree is a completely random tree with a binary tree structure. The construction steps are as follows: \( n \) randomly selected data from the interest packet data are used as the training set, and the dataset is subdivided into two nonempty subsets by a randomly chosen attribute \( i \) and a cutoff value \( c_i \) with a split \( x_i > c_i \). The attributes and values are stored in the nodes of the RecTree, and these attributes and values are used to reconstruct the interest packet data. The training set is recursively randomly selected and split until the interest packet dataset can no longer be split or reaches the maximum height of the set tree \( h \). Following this rule, a RecTree is constructed, and multiple RecTrees are constructed by randomly selecting \( n \) interest packet data as the training set multiple times to form a RecForest.

D. IFA Detection

After constructing the RecForest by the above method, we select interest packet data according to the selected attributes and values, and by traversing the RecTree, we can obtain the reconstructed range of this interest packet. By calculating the reconstructed points, when traversing all the RecTrees, we can obtain the range of the multiple reconstructed interest packets, and by calculating the distances from all the reconstructed points to the origin, we can obtain the reconstructed error. Since the RecForest can identify the outlying attributes of an anomaly, the error components of each attribute of the reconstruction error can be calculated, which improves the interpretability of the reconstruction error and reduces the misclassification rate. When an IFA occurs, the number of malicious interest packets is large, and no corresponding data packets are received, which makes the reconstruction error small because the received data packet attribute error component is 0, so the reconstruction results are heavily clustered near the reconstruction center.

After the RecForest is constructed, each data point in the interest packet dataset is traversed in the RecTrees, and after obtaining the reconstruction errors of all interest packet data, they are ranked according to the reconstruction errors from smallest to largest, and the top-ranked one can be considered a malicious interest packet.

The computational complexity of the RFDM can be calculated as follows: if \( m \) RecTrees with a maximum tree depth of \( h \) are generated from the training set of \( n \) data points, the time complexity of the construction phase is \( O(nmh) \), which is linearly related to the size of the training data. In the traversal and analysis phase, the time complexity is \( O(tmh) \), where \( t \) is the size of the interest packet dataset.

IV. Simulation and Discussion

To verify the accuracy and efficiency of the proposed RFDM, simulation experiments were conducted on the ndnSIM [16] platform. The simulation experiments used a small mesh structure as the NDN topology, which included one content publisher; eleven NDN routers; nine normal users, four of which were burst traffic users; and five attackers, as shown in Figure 1. The basic parameters were set as follows: the simulation duration was 15 s, the detection period was 100 ms, the survival time of the interest packets was 1 s, the transmission time of the data packets and interest packets
was 10 ms per hop, the number of RecTrees was 100, the normal users used the prefix "/good" to request content, the burst traffic users used the prefix "/burst" to request content, and the attackers used the prefix "/evil" to carry out attacks. In the simulation, the normal users requested legitimate interest packets at a rate of 100 packets/s to request existing data content. The burst traffic users requested legitimate interest packets at a rate of 200 packets/s to request existing data content at the 5th to 8th s, while the attackers requested malicious interest packets at a rate of 200 packets/s to request nonexistent data content at the 10th s to launch the IFA.

A. Accuracy of the RFDM

First, we describe the fluctuating traffic caused by normal users and compares it with four other typical IFA detection methods to verify the detection accuracy of the RFDM. When fluctuating traffic caused by normal users occurs, the name distribution may be misclassified as an IFA by existing IFA detection methods because the router receives more legitimate interest packets in a short period of time, resulting in a change in the name distribution. Since R4 is closest to the producer and the interest packets in the network need to ultimately reach router R4, the graphs in this paper are mainly based on router R4 detection.

Figure 2 shows a comparison of the accuracy of the proposed RFDM and other typical IFA detection methods. As shown in Figure 2, the value “1” indicates that an IFA is detected, and the value “0” indicates that an IFA is not detected; between 5 s and 7 s, i.e., during the occurrence of flow fluctuations, the information entropy-based method, the Gini impurity-based method, and the Theil-based method do not misclassify IFAs at the beginning of the flow fluctuations, but they eventually do. Although the expired-PIT method is not affected by traffic fluctuations in the occurrence of false positives, the attack occurs 1 s after IFA detection because when the PIT entry expires, the expired-PIT method starts to detect the attack, resulting in a slower detection speed. The RFDM shows a value of “0” during the whole period of traffic fluctuation, which resists the misjudgment caused by traffic fluctuation and shows a low misjudgment rate. In addition, as shown in the figure, the IFA starts to be activated in the 10 s phase, and the RFDM quickly detects the IFA throughout the attack period, thus demonstrating its high accuracy. This result shows the superior advantage of the RFDM over previous detection methods in resisting the effects of traffic fluctuations and IFA detection.

As shown in Figure 3, the RFDM achieves a true positive rate of 0.9042 in detecting malicious interest packets, while for the other three detection methods, the Theil-based is 0.8875, the Gini impurity-based is 0.8708, and the information entropy-based is 0.7958. The results show that the RFDM’s detection of malicious interest packets is more accurate than those of the other three IFA detection methods.
B. Efficiency of the RFDM

Figure 4 shows the comparison of the proposed RFDM with other methods in terms of the number of PIT entries during the simulation. As shown, the proposed RFDM is able to detect an attack quickly after IFA startup and reduce the PIT to a normal level after 0.3 s. Among these methods, the Theil-based and Gini impurity-based methods can also reduce the PIT to a normal level within 0.5 s after an attack occurs, while information entropy-based methods reduce the PIT to a normal level 0.7 s after the attack occurs, while the slowest method, expired-PIT, takes nearly 1.2 s to reduce the PIT to a normal level. The Theil-based and Gini impurity-based methods are similar in speed, and the information entropy-based method is the second slowest. Since the Theil-based, Gini impurity-based, and information entropy-based methods are based on interest packet name distribution statistics, they need some time for calculation, while the information entropy-based method takes longer to detect IFAs because it needs to wait for the calculation of cumulative entropy. Expired-PIT needs to wait until the PIT in the entries expires before starting to calculate the statistics, so it takes the longest time.

Figure 5 shows a comparison of the number of PIT entries in the network with different attack rates used by the attackers. Figure 5 shows that when the attack rate is 200 packets/s, the RFDM can quickly detect the malicious IFA and take measures to limit it. As the attack rate used by the attacker increases, the number of PIT entries increases dramatically. The results show that the RFDM can also detect IFAs well at different attack rates and take corresponding restriction measures to mitigate the impact of IFAs.

Therefore, it can be concluded from Figure 4 and Figure 5 that the proposed RFDM is more stable, more effective in detecting IFAs, and better resists the impact of traffic fluctuations, which can effectively mitigate the impact of IFAs on normal users.

C. Discussion of the Detection Period

The detection period has an important impact on the detection computation overhead and detection accuracy. Therefore, we separately discuss the impact of the detection period on the IFA detection computation overhead in the same attack scenario; i.e., the normal users send legitimate interest packets at a rate of 100 packets/s, the burst traffic users request legitimate interest packets at a rate of 200 packets/s and the attackers launch an attack at a rate of 200 packets/s. We counted five different sets of detection periods, from 10 s to 15 s for a total of 5 s, and determined the time overhead required by the RFDM in each round of the detection period from collecting data to deriving the IFA results. As shown in Figure 6, when the detection period is 10 ms, the time overhead used to calculate the IFA results is 870.275 ms. When the
detection period is 100 ms, the required calculation overhead is 222.606 ms. The number of PIT entries varies with time due to the detection period. The lowest consumption is required when the detection period is 100 ms and the transmission time of the interest packet and data packet per hop in the network is 10 ms. Figure 6 shows the true positive rates for different detection periods in the detection of interest packets, and the true positive rate is 0.7917 in the 10 ms case. Since the transmission time of the packets is set to 10 ms per hop in the simulation experiment, it is likely that the legitimate interest packets have not yet been returned by the data packets when the data are collected, leading to a lower true positive rate. When the detection period was increased to 50 ms, the true positive rate improved to 0.9004. When the detection period was 500 ms, the true positive rate reached 0.9271, but due to the long detection interval, it is likely that the router PIT had been filled with malicious interest packets because of a high-speed IFA, leading to a network crash. Figure 6 shows that a detection period of 100 ms yields the optimal computational overhead and better IFA detection results when the packet transmission time per hop in the network is 10 ms.

V. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

IFAs in NDN can affect the normal use of a network and cause significant harm to the network. In this paper, we propose the RFDM, which uses a RecForest to calculate reconstruction errors by collecting information from interest packets and detects malicious interest packets among normal interest packets according to the reconstruction error. The simulation results show that the proposed RFDM can not only detect IFAs quickly but also address the problem of traffic fluctuations caused by legitimate users, improve the accuracy of attack identification, and reduce the number of false positives. Compared with four other typical IFA detection methods, the RFDM can effectively reduce the problem of excessive PIT usage caused by IFAs and can obtain higher values in identifying true positives. This demonstrates that the RFDM can provide better IFA detection performance. In future work, we plan to conduct research in larger and more complex network environments and incorporate more sophisticated attack models.

ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China, Grant Number 62372479.

REFERENCES


