

Review of Security Research on Address Resolution Protocols

Song Guangjia and Ji Zhenzhou

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 150001

35059899@qq.com; jizhenzhou@hit.edu.cn

Abstract

Address resolution protocols (ARPs includes ARP, NDP, SEND, etc.) play an important role in network communication; the security of the ARPs is the premise and guarantee of network security. ARPs consists of three phases: acquisition of the mapping of the target address; duplicate address detection; data structure maintenance. In this paper, we analyze the existing ARPs according to these three phases, analyze security threats and the corresponding attack methods; classify and describe the existing studies on ARPs security. Our analysis shows that the main factors that restrict the development of ARPs are the contradiction between efficiency and cost and the lack of theoretical support in protocol design. Finally, the development of ARPs is prospected.

Keywords: *Network security, Address resolution, ARP, Neighbor discovery, SEND*

1. Introduction

With the continuous development of human society, the network has become part of the infrastructure in most countries from a small-scale private facility. People can access the network to exchange information and obtain resources by wired or wireless anywhere, anytime.

From scientists, teachers, and other professionals to everyone on the earth, network users have also obviously changed. Users changed from credible small-scale crowd to the untrustworthy large-scale crowd. From a sociological point of view, network is an extension of the human senses. The change of network users brings in a concomitant challenge to the network security.

One of the major functions of the network is to allow information exchange; packets flow from one node to another routed by routers or switches. In the network, packets are delivered by two methods: direct delivery and indirect delivery. If two communicating parties are in the same local area network (LAN), a switch can deliver the packets by searching a table which includes the mapping of each port (<Port, MAC> table) so that the switch can deliver the packets to the target node through the port directly; this process is a direct delivery. In indirect delivery, both communicating parties are not in the same LAN, thus the packets need to be forwarded by routers until these reach the local network of the target node, and then the packets are sent to the target node by a local switch using direct delivery[1, 2]. Therefore, no matter what kind of delivery, the target MAC address needs to be initially determined before the packets can be delivered to the target. The process by which the target MAC address can be obtained from its IP address is called address resolution [3]. This process is mainly completed by address resolution protocols (referred to as ARP in IPv4). The consequences caused by attacks against the address resolution process are very serious. For example, typical

man-in-the-middle attacks can intercept or tamper data, and even interrupt network communication. Attacks against the address resolution process are the main threats to the LAN security [4].

The network in this article refers to Ethernet. The packet format is also standard to Ethernet. The abbreviations used in this paper are listed in table 1.

Table 1. Abbreviations

ACD	Address conflict detection
ARP	Address Resolution Protocol
CGA	Cryptographically generated address
DAD	Duplicate address detection
DES	Discrete event system
ECC	Elliptic curve cryptography
FCFS	First come first service
IMCP	Internet control message protocol
IID	Interface identifier
MITM	Man in the middle
NA	Neighbor advertisement
NS	Neighbor solicitation
SAVI	Source address validation implementation
SEND	Secure neighbor discovery
SLAAC	Stateless address auto configuration

2. Development of Address Resolution Protocols

2.1. Overview of the address resolution process

The general address resolution process is described as follows: each host in a LAN has an address cache (cache that stores the IP and MAC addresses of other hosts). Assume that the IP address of host *A* is 10.0.0.3 and that of host *B* is 10.0.0.4. When host *A* wants to send packets to host *B*, the packets can be delivered directly because hosts *A* and *B* are in the same IP segment. Therefore, host *A* checks whether the IP address of host *B* exists in the address cache. If it does, host *A* sends the data packets to the corresponding MAC address. Otherwise, host *A* needs to send an address resolution broadcast packet requesting host *B* for an answer with its MAC address. All the hosts in the LAN can receive the address resolution broadcast packet, but only host *B* will answer a reply packet, which contains its IP and MAC addresses. Host *A* will update its cache table after receiving the reply packet and then send packets to the MAC address of host *B* in the cache [5]. The specific process is shown in Figure 1.

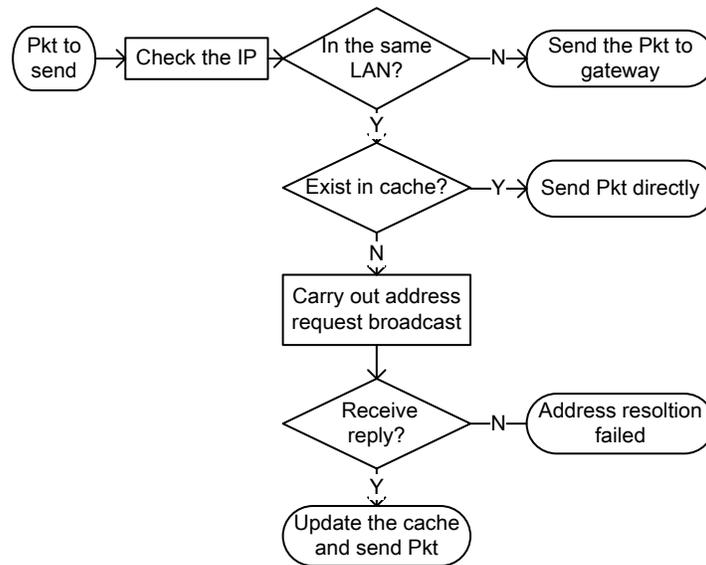


Figure 1. Flowchart of the Address Resolution Process

2.2. Frame Analysis of Address Resolution Protocols

In IPv4, ARP is used for address resolution. In IPv6, neighbor discovery protocol (NDP) is used for address resolution. NDP is an important basic IPv6 protocol. It combines and improves on ARP, internet control message protocol (ICMP), router discovery protocol, and ICMP redirection protocol of IPv4. In addition, as a basic IPv6 protocol, NDP also performs the following functions: prefix discovery, neighbor unreachability detection (NUD), duplicate address detection, and stateless address auto configuration (SLAAC) [6]. Later, owing to security and other reasons, Internet Engineering Task Force (IETF) proposes secure neighbor discovery (SEND) [7]. However, from the analysis of protocol framework, ARP, NDP, and SEND carry out the following three phases at the least:

- <IP, MAC> mapping acquisition
- Duplicate address detection
- Data structure maintenance

(1) <IP, MAC> mapping acquisition. This phase involves the process of obtaining the corresponding MAC address of the destination IP address. Figure 2 shows the ARP message format. The main fields are described as follows:

Type: protocol type in layer 2. The protocol type of ARP is 0x0806.

Hard type: hardware type. The hardware type of Ethernet is 1.

Port type: protocol type. The protocol type of IP is 0x0800.

Hard size: Protocol address length (in bytes). The protocol address length of Ethernet is 6. The protocol address length of IP is 4 [8].

Op: operation code. The Op of ARP request is 1. The Op of ARP reply is 2.

Src MAC is the link-layer address of the source node. Src IP is the IP address of the source node. Dest MAC is the link-layer address of the destination node. Dest IP is the IP address of the destination node.

Ethernet Header	Dest MAC Src MAC Type
IPv4 Header	Hard type Port type Hard size Port size op
IPv4 Data	Src MAC Src IP Dest MAC Dest IP

Figure 2. ARP Message Format

The <IP, MAC> mapping acquisition process in ARP is relatively simple compared with NDP. In ARP, mapping may be obtained in two ways: active acquisition and passive acquisition. Active acquisition is the normal address resolution process previously mentioned. Suppose the MAC address of host *A* is AA-AA-AA-AA-AA-AA, and the MAC address of host *B* is BB-BB-BB-BB-BB-BB. When *A* wants to communicate with *B*, *A* will send an ARP broadcast if *A* does not know the MAC address of *B*. If the IP addresses of both *A* and *B* are in the same address segment, the packets can be delivered directly. Otherwise, *A* will broadcast an ARP request packet. This specific format is shown in Figure 3. The Dest MAC field value is FF-FF-FF-FF-FF-FF. Routers (or switches) will broadcast this ARP packet to require *B* to answer with its MAC address. All hosts in the same LAN can receive the ARP broadcast packet (If VLAN is configured, the hosts in a different VLAN cannot receive the broadcast packet), but only *B* will answer with ARP reply packet, which contains the IP and MAC addresses of *B*. The reply is shown in Figure 3(b). *A* obtains the MAC address of *B* after receiving the reply packet and then sends the data to *B* using the corresponding MAC address.

ARP Broadcast		ARP Reply	
Ethernet Header	FF:FF:FF:FF:FF:FF AA:AA:AA:AA:AA:AA 0x0806	Ethernet Header	AA:AA:AA:AA:AA:AA BB:BB:BB:BB:BB:BB 0x0806
IPv4 Header	1 0x0800 6 4 1	IPv4 Header	1 0x0800 6 4 2
IPv4 Data	AA:AA:AA:AA:AA:AA 10.0.0.3 00:00:00:00:00:00 10.0.0.4	IPv4 Data	BB:BB:BB:BB:BB:BB 10.0.0.4 AA:AA:AA:AA:AA:AA 10.0.0.3
(a)		(b)	

Figure 3. ARP Request and Reply

In passive acquisition, ARP does not use finite-state machine, thus even the node does not initiate the address resolution process but instead obtains the <IP, MAC> information of the sender and updates its cache according to the message content when it receives ARP request or ARP reply.

The mapping acquisition process in NDP is similar to that in ARP. It uses neighbor solicitation (NS) and neighbor advertisement (NA) message to complete the address resolution process. The NDP message format is shown in Figure 4.

Ethernet Header	Dest MAC Src MAC Type
IPv6 Header	Src IP Dest IP Next header
IPv6 Data	Flags Type Target address Option

Figure 4. NDP Message Format

NDP and ARP have significant differences, including the following:

(1) NDP no longer uses link-layer broadcast; it uses multicast.

(2) An NDP packet has two additional fields: Target address and Option. The Target address field provides the target IP address in address resolution process. The Option field has different meanings based on the “Type” field, but generally, it is filling in with the link-layer address. The type field refers to the packet type, the type of NS is 135, and the type of NA is 136.

If SLAAC is used, according to the EUI-64 and their own interface identifier (IID), hosts *A* and *B* generate their own link-local addresses, which are FE80: A8AA: AAFF: FEAA: AAAA and FE80: B9BB: BBFF: FEBB: BBBB, respectively [9]. When *A* wants to communicate with *B* but does not know the MAC address of *B*, *A* will multicast an NS first. *B* will use an NA to reply when *B* receives the NS. *A* obtains the MAC address of *B* when *A* receives the NA, thus the two parties can communicate [10]. The NA message and NS message used in the communication are shown in Figure 5(a) and Figure 5(b).

NS		NA	
Ethernet Header	33:33:FF:BB:BB:BB AA:AA:AA:AA:AA:AA 0x0806	Ethernet Header	AA:AA:AA:AA:AA:AA BB:BB:BB:BB:BB:BB 0x0806
IPv6 Header	FE80::A8AA:AAFF:FEAA:AAAA FF02::1:FFBB:BB:BB 0x3A	IPv6 Header	FE80::B9BB:BBFF:FEBB:BBBB FE80::A8AA:AAFF:FEAA:AAAA 0x3A
IPv6 Data	135 FE80::B9BB:BBFF:FEBB:BBBB AA:AA:AA:AA:AA:AA	IPv6 Data	136 S=1,O=1 FE80::B9BB:BBFF:FEBB:BBBB BB:BB:BB:BB:BB:BB
(a)		(b)	

Figure 5. NS and NA in Address Resolution

(2) Duplicate address detection (hereafter DAD, it includes the duplicate address detection process in NDP). This process aims to determine whether an IP address is repeated or conflict with other nodes in a LAN. It is a necessary step before using a new IP address.

Gratuitous ARP is essential to complete DAD. The gratuitous ARP process is as follows: if host *A* decides to use 10.0.0.3, it needs to send an ARP request broadcast before using the IP address. *A* fills its own IP address, which is 10.0.0.3, in the destination IP field. This request is different from the general ARP request. Its main purpose is to detect whether another node has already used 10.0.0.3. The specific format of gratuitous ARP message is shown in Figure 6. This method is mainly used when the host interfaces has just started or the MAC address has changed.

Ethernet Header	Dest MAC Src MAC Type
IPv6 Header	Src IP Dest IP Next header
IPv6 Data	Flags Type Target address Option

Figure 6. Message Format of Gratuitous ARP

To improve the detection process, RFC5227 proposed a new method called address conflict detection (ACD) [11]. ACD adds two new packets: ARP probe and ARP announcement. ARP probe is similar to an ARP request, but its Src IP field is filled in with 0.0.0.0 to reduce pollution. As passive <IP, MAC> mapping acquisition is in effect in ARP, the other hosts will update their caches according to the IP and MAC information obtained after they receive the ARP request broadcast. The ACD process is as follows: when a host wants to use a new IP address, such as 10.0.0.3, the host initially sends an ARP probe message, whose format is shown in Figure 7(a). When a conflict is determined not to exist, the host sends an ARP announcement message, whose format is shown in Figure 7(b). Generally, the announcement needs to be sent thrice. At this time, the Src IP and Dest IP fields of the ARP announcement message are filled in with the new IP address of the host to declare that the host is ready to use 10.0.0.3. If a conflict occurs, ACD provides three options: stop using; resend an ARP announcement, and if a conflict occurs again, stop using; continue to use this address, regardless of the conflict.

ARP probe		ARP announcement	
Ethernet Header	FF-FF-FF-FF-FF-FF AA-AA-AA-AA-AA-AA	Ethernet Header	FF-FF-FF-FF-FF-FF AA-AA-AA-AA-AA-AA
IPv4 Header	1 0x0800 6 4 1	IPv4 Header	1 0x0800 6 4 1
IPv4 Data	AA-AA-AA-AA-AA-AA 0.0.0.0 00-00-00-00-00-00 10.0.0.3	IPv4 Data	AA-AA-AA-AA-AA-AA 10.0.0.3 00-00-00-00-00-00 10.0.0.3
(a)		(b)	

Figure 7. ARP Probe and ARP Announcement

As in IPv4, DAD is also a necessary process before an address may be used in NDP. The detection process mainly replies to NS and NA messages. For example, host A needs to send an NS message to check for conflict before using FE80: A8AA: AAFF: FEAA: AAAA. The NS and NA used in DAD are shown in Figure 8.

entry states include “stale,” “incomplete,” “reachable,” “delay,” or “probe” and may transition according to the own lifetimes and usages of the entries. Stale is a steady state, whereas incomplete, reachable, delay, and probe are non-steady states. Entry state transition may be ascribed to three reasons: active exploration, timer information, and feedback of upper-layer protocols. An incomplete state will become a reachable state after the entry receives the confirmed information. Only the mapping in the reachable state can be used directly, whereas the entries in the other states need to be confirmed until their states become reachable or be removed from the cache. The specific state transition process is shown in Figure 9. The characteristic feature of this process is that NDP does not believe the address resolution packet in a broadcast directly. The cache state does not become reachable directly after the node receives an NA broadcast. Only after the node tests and confirms an entry will its state be updated to reachable. Tests and confirmation refer to having received the NA with *S* bit equal to 1 in the Flags field after sending the NS or having received the confirmation of the upper-layer protocol. For example, TCP has just received a message from the other party.

3. Security Threats and Technologies

3.1. Security threats

In the address resolution process, safety problem has long been very challenging both in ARP and NDP. Not all nodes in the network are credible, thus a cheating node usually exists; these malicious nodes can break (listen or intercept) network communication by sending messages containing false IP or MAC information [12-13]. Security attacks can be classified into three types: mapping acquisition process attack; DAD attack; and cache attack.

(1) Mapping acquisition process attack. This attack method is common. For example, in ARP, when host *A* wants to communicate with host *B* but does not know the MAC address of host *B*, host *A* will broadcast, but when host *A* receives a reply packet, *A* would not check whether the ARP reply is true. As long as the destination MAC address of the ARP packet is AA-AA-AA-AA-AA, then host *A* will accept it and update its ARP cache, making it convenient for a spoofing attack. In a typical attack, such as a man-in-the-middle attack, when host *A* broadcast an ARP request to require host *B* to reply, host *C* quickly sends an ARP reply to pretend to be host *B*. After receiving the reply, host *A* does not know if it is a false packet and will mistakenly update its cache table to regard *C* as *B*, then send the packets which should be sent to host *B* to host *C*. *C* can also continue to use the ARP packet to deceive *B* and lead *B* to believe *C* is *A*, thus double deception, that is, a man-in-the-middle attack, is perpetrated [5]. The process adopted in a man-in-the-middle attack is shown as Figure 10. Of course, *C* can also use the invalid MAC address to block the communication between host *A* and other hosts, thereby committing a denial-of-service attack. Both attacks, which disrupt the mapping acquisition process, also occur in NDP by a similar method as in ARP.

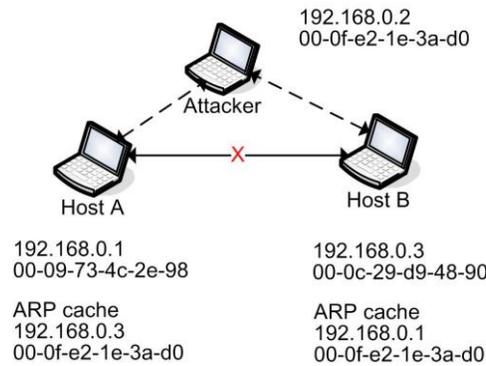


Figure 10. The Man-in-the-middle Attack

(2) DAD attack. This kind of attack is obvious in NDP. When host A has configured a new IP, host A will carry out NS multicast to check for conflict. During this process, an attacker may send a false reply declaring that the target IP is already in use, thus the host can only configure another IP to perform another detection. If this attack is constantly repeated, then host A will not obtain a usable IPv6 address and fail to connect to the network, resulting in a DoS attack [14]. Particularly in IPv6, SLAAC allows a node to configure a local IPv6 address automatically according to EUI-64 rules. Therefore, in SLAAC, DAD attack causes a considerable effect.

(3) Cache attack. ARP lacks a status mechanism, thus as long as the host receives an ARP reply, it will apply the first-come first-served (FCFS) scheme to update its own cache. The attacker can take advantage of this mechanism to commit cache pollution. For example, the attacker can constantly broadcast to pretend to be the default gateway, so the attacker can intercept numerous outer net packets. This method can also interrupt outer net communication. The attacker can also send false <IP, MAC> information, leading to the cache pollution of other nodes in the LAN and consequently to an increase in the packet loss rate in the LAN [15]. Of course, the FCFS principle in ARP also has its advantages, such as the firewall can take advantage of this feature to sustain its authority in the network. If a firewall is deployed in the LAN (assuming the firewall is the default gateway), the firewall can apply the FCFS scheme when a host IP address conflicts with the firewall and broadcast ARP reply quickly to take the cache of the other host; therefore, its gateway function is not affected.

Although the NDP cache maintenance uses status mechanisms, attacks, such as neighbor unreachability detection attacks, still exist. When a node in the local link is unreachable, its neighbor nodes repeatedly send unicast NS packets several times. If a reply is not provided, the related entry will be deleted in the neighbor cache. However, if an attacker forges the IPv6 address of the node and constantly sends NA message to the neighbor nodes, as no authentication mechanism is implemented, the neighbor nodes will mistakenly believe that this node is still reachable, thus the neighbor cache is polluted [16].

3.2. Address Resolution Security Technology Research Status

At the onset, security research focused on ARP security; however, with the promotion and deployment of IPv6, research was directed to the study of NDP, focusing mainly on attack behavior detection, cache inspection and protection, and encrypted communications and protocol improvement. In practical application, switches can use IP-MAC binding to prevent ARP spoofing, but this method needs manual maintenance, which is unsuitable for wide-range, dynamic networks. The VLAN division can be adopted to divide large a LAN into

several small LANs logically, reduce broadcast range, and limit the damage of an ARP attack; however, it changed the logical structure of the network.

3.2.1. Node Behavior Inspection Technology: Node behavior inspection is based on fact that each NIC (network interface card) only has one unique MAC address. Particularly in IPv4 network structure, a host usually has one IP address, then the IP and MAC addresses of a node form a one-to-one mapping, which generally neither change nor change frequently.

Therefore, by checking the correspondence between the IP and MAC addresses in the packets that nodes send out, a deceitful behavior can be detected. References [17] and [18] use a discrete event system (DES) in a intrusion detection system to monitor all the ARP messages in the LAN. If the system finds an ARP request, it will use the active probe message to resolve the source address of the request packet. Then, according to the resolution results, it will determine whether a deception exists. If it does, the DES will receive more than one reply, thus leading an event timing different from normal conditions. This method needs a trusted host to run the DES system in the LAN and monitor all the network traffic, so that a single point of failure exists.

References [3] and [4] also do not trust ARP packets, especially when broadcast. Both references use ICMP message for reverse exploration to test whether the source address of the ARP message is true. According to the test results, they decide whether to trust ARP message. The fundamental motivations for using ICMP packet are as follows:

- NIC only accept the message whose destination MAC address is consistent with the NIC hardware, or broadcast message and specific multicast. Other types messages are discarded.
- The attack node cannot prevent the normal node from making a reply to an ARP request and ICMP request. The normal node will not forward ICMP echo request with a destination IP address not consistent with its own IP.

Node behavior may be checked in different ways. First, check whether the IP and MAC addresses in the reply packet are consistent with the original received ARP reply packet; if not, a deception exists. If the node replies to two different IP address tests, a deception exists. Reference [3] can detect nodes that have considerable ability to deceive in a network (can respond to any message), but cannot prevent their attacks. Reference [4] requires a host to maintain a permanent level-2 cache, which should ensure that the IP and MAC map one for one. However, if the firewall of the host shields port 8, this method may not work properly.

In IPv6, the method of checking node behavior becomes difficult, because a node can have multiple IP addresses at the same time in IPv6. A node can have either a link local address or other types generated according to the router prefix information. Thus, a one-to-one correspondence between the IP address and the MAC address no longer exists, making the node behavior detection method more difficult to implement.

Source address validation architecture (SAVA) adopts a new security method. The main idea is to filter packets based on their source address information. The advantages of this method include the prevention of attacks directly coming from the source, source address tracking and traceability, and network diagnosis and management [19-20]. SAVA can be deployed in different layers, such as the access layer, the entrance of the autonomous system, and in between different autonomous systems. Being very flexible, SAVA overcomes the deficiency of the single entrance security. When SAVA is deployed in the access layer, its main function is to prevent a host from using a false address to cheat in a LAN, as shown in Figure 11.

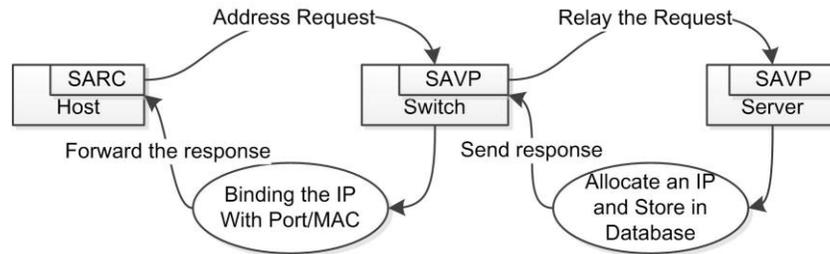


Figure 11. Deployment of SAVA in the Access Layer

Suppose a host needs to use the IP_X . First, the source address request client (SPRC) of the host sends an IP request to the source address validation proxy (SAVP). SAVP records the MAC address of the host and the port number where the request comes from and then forwards the packet to the source address management server (SAMS). SAMS makes a judgment and sends a DAD packet for the IP_X . If it finds no conflict, SAMS will assign the IP and store the IP_X in the database at the same time, and then issue a permit to SAVP, which in turn will bind the IP_X with the port and the MAC addresses. After the binding, SAVP transmits the IP_X to the SARC. According to the binding information, the switch can filter the packet sent by the host.

Source address validation implementation (SAVI) is one of the implementations of SAVA. Based on the network topology, SAVI first divides the network into safe area and unsafe area. To save on cost, the secure domain communication adopts the trusting method, that is, internal communication is not filtered, whereas insecure domain communication is filtered according to the binding information [21-23].

So far, different solutions have been proposed to address configuration, such as DHCP SAVI for DHCP, FCFS SAVI for the ordinary IPv6 network, and SEND SAVI for SEND. In FCFS SAVI, the switch port maintains a database to record information, such as IP addresses of unsafe ports and port numbers. The FCFS mechanism is adopted, that is, recording of information is based on the address resolution behavior of the first observed node, such as when a node undergoes DAD test for the first time. If the test result shows that conflicts with other nodes and with the existing binding information do not exist, the switch will bind the access port number and the DAD address. However, once the information attempting to enter the port is not consistent with the binding information, the switch will prevent it from entering.

Thus, the deployment of SAVI in the access layer can partly prevent address resolution spoofing. However, as SAVI does not inspect the MAC address; therefore, if a node sends a packet with valid source IP address but erroneous MAC address, then SAVI cannot filter these packets. Reference [24] proposes that MAC address be used a binding information to enhance the effectiveness of FCFS in the link layer. SAVI is still currently in the experimental stage. Its domain source address verification function requires a router for support, and the router needs to complete considerable centralized computing, which will affect the network performance. Moreover, different equipment manufacturers implement simple network management protocol (SNMP) differently, thus the deployment of SAVI can be quite difficulty [25-26].

3.2.2. Cache Protection and Inspection Technology: The main principle of this technology is to prevent the attacker from destroying the caches of other hosts. Majority of attackers need to pollute the caches of the victim hosts, making these caches store wrong information. To prevent attacks, the cache of a node must be protected from pollution, an effective method to inspect the cache must be adopted, error entries should be promptly corrected.

Different from the traditional cache update method, status mechanisms have been added to the cache in reference [27]. If no ARP request has been sent before, the cache will not update after receiving an ARP reply or ARP request. At the same time, if more than one ARP reply have been received during the address resolution process, then a deception exists. Reference [28] recommends the use of fuzzy logic controller to find out the reply with the highest credibility from the multiple replies to complete the resolution process; however, if only one ARP reply is received, regardless if the reply is authentic or not, the host will accept it. Reference [28] carries out long-term inspection on a cache and suggests that when multiple IPv4 addresses correspond to the same MAC, then a deception exists. The attack node can be identified according to the MAC address or by manually conducting cache clearance. This method is passive, if an exception is found in the cache, the attack has existed for a certain time, and thus previous communication is no longer safe.

3.2.3. Encryption Communication: Encrypted communication is a method of using encryption techniques. Two parties use ciphertext to communicate, thus even if the packet is intercepted by other nodes, it cannot be read or understood. Timestamp or digital signature technologies can be used to prevent replay attacks.

For security purposes, IETF proposes SEND as an enhancement mechanism for NDP, and cryptographically generated address (CGA) is the main characteristic of SEND [29]. CGA is the unique address format of SEND. CGA address is derived using a subnet prefix, public key, collision count, modifier (via multiple hash algorithms to find a suitable modifier value); carrying out Hash2; and then taking the first 59 of the 160-bit hash value, coupled with the Sec (security level) values and other parameters. The specific computation process is shown in Figure 12.

However, the CGA process requires considerable calculation [30] and also increases the encrypted NDP packet volume, thus leading to communication bandwidth increase. In addition, the time consumption of CGA is positively related to the Sec bit: the bigger the Sec, the longer the time. Each time Sec increases by 1, time consumption increases nearly 2^{16} times. CGA address should also be subjected to DAD. Aiming to solve issue on computation cost of CGA, reference [31] proposes a time-stopping algorithm for CGA, which reduces the time used for CGA calculation by determining the running time limit to set a suitable Sec value so as to ensure CGA address generation within the schedule time. Reference [32] proposes a parallel computing algorithm for CGA to reduce time consumption. Reference [33] proposes the use of elliptic curve cryptography (ECC) key to replace RSA key to reduce the computation time but still achieve the same level of security. ECC key is shorter, thus it causes a shorter NDP message. In Reference [34], a special high-performance server for key calculation is proposed in the network, and DHCP server is recommended to manage CGA. DHCP protocol is improved to broadcast CGA parameters in the network, thus the DHCP server becomes an important node, leading to single point of failure. In CS-CGA, ECC is used to replace RSA to speed up calculation and encrypt prefix at the same time, with an aim to refute the time-balancing algorithm [35]; however, this method undoubtedly affects the router forwarding speed.

CGA is difficult to deploy, most operating systems support NDP, but does not support SEND, and primary communication equipment manufacturers only partially achieves the full function of SEND, thus SEND remains in the experimental stage [36]. Furthermore, the mixed use of SEND and the NDP can cause routing problems [37]. In summary, the current research on SEND mainly focuses on the following three aspects:

(1) Optimize CGA process or use an alternative algorithm to achieve the same security but reduce the complexity of the algorithm.

- (2) Add MAC address as a parameter to the CGA calculation process to prevent link-layer spoofing [38].
- (3) Use time strategy according to the time requirements to determine the security level of CGA and to complete CGA calculation within the scheduled time.

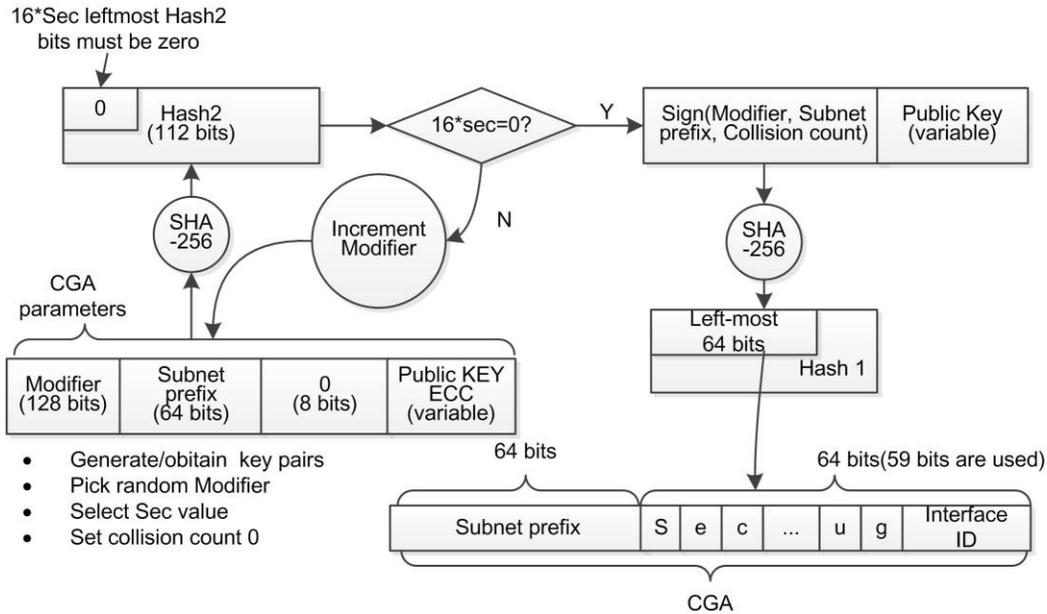


Figure 12. Schematic of CGA address generation

3.2.4. Improvement of Resolution Process: Improvements on the protocol mechanism of the three phases, particularly the address mapping acquisition process, have been proposed. One such improvement is improving the finite-state machine or the authority server or the middleware.

Gouda et al. propose a new address resolution framework, which includes a secure server and two new protocols, namely, invitation–accept and request–response. The invite–accept protocol is employed so that the host registers its <IP, MAC> mapping on a secure server. Request–response protocol is used so that the host requests the MAC addresses of other LAN hosts from the security server [39]. However, this architecture is not compatible with the standard ARP, and the server is a single point of failure.

Brushi et al. propose S-ARP protocol as an improvement of ARP. The method mainly uses asymmetric encryption technology to authenticate the ARP message to prevent ARP spoofing. S-ARP needs to deploy in the network an authoritative distribution of key server (AKD), which stores the IP address and public key of each host in the LAN. The S-ARP host needs to connect to the AKD server to get the public key of the host who sends an ARP message to validate the just received ARP message. The host can also cache the public key to improve authentication speed [40]. The advantages of this method include the use of asymmetric encryption technology to verify the legitimacy of ARP packets and compatibility with ARP. However, encryption and decryption consume a certain amount of CPU resources, thereby impacting the network bandwidth. Furthermore, the safety of the AKD server should be secured. Issac et al. propose a unicast protocol called S-UARP, which combines ARP with DHCP [41]. S-UARP uses the DHCP server and extends it to support the address resolution process. During the process, as the DHCP server has the address configuration information of

all hosts in the LAN, the source host does not need to broadcast an ARP request for address resolution but instead uses unicast to send resolution request to the DHCP server. Then, the DHCP server sends the address configuration information (IP and MAC) of the destination host to the source host, and digital signatures are adopted in the communication process to prevent spoofing. This method requires the concurrent change of ARP and DHCP. Furthermore, the DHCP server should always be safe, thus a single point of failure exists.

4. Comparison of Existing Research and Prospective Research

Previous research has shown that two main factors restrict the development of ARPs: implementation cost and lack of theoretical support.

4.1. Implementation Cost

4.1.1.C: Many studies have proposed the need to improve and assign a dedicated server in the LAN, such as references [4, 40]. To some extent, these methods are difficult to implement, because of hardware cost. Regardless if the scale of the network can be large or small, equipping any network with a dedicated server is impossible. For example, only the network with high-security requirements and whose provider has financial capability may be equipped with a firewall. To solve the single point of failure problem, the hardware and software reliability of the server must be also very high, undoubtedly increasing implementation cost.

4.1.2. Operation Cost: Operation cost is related to implementation and maintenance costs. For example, most switches and operating systems support IP-MAC binding. Binding can effectively resist spoofing attack, but the operation complexity is too high. A plurality of devices needs to be configured to support binding. When the IP or MAC addresses of the node change, manual maintenance is required. For large-scale or dynamic networks, the workload is too large.

4.1.3. Computation Cost: Computation cost is mainly reflected in encryption technology, such as the implementation of SEND. Using encryption, digital signature, among others can effectively prevent eavesdropping interception, tampering, and replay attack. However, the encryption and decryption processes require sizeable computation, which leads to communication cost increase, which can affect user experience. For example, user experiences with VPN and common network are completely different.

4.2. Lack of Theoretical Support

At present, many studies on ARP lack theoretical support. The proposed improved methods are mainly based on experience and the purpose of resisting certain attacks, resulting in shortcomings. This situation has also led to the lack of a unified framework in the current research. Comparing various studies horizontally is difficult. We propose the application of game theory, mechanism design, and security protocol theory to the design of ARPs.

4.2.1. Address Resolution is Undecidable:

Prerequisites:

- (1) The MAC address is unrepeatable.
- (2) If mapping $w = \langle \text{IP}, \text{MAC} \rangle$ does not exist in the network (the host configured as w does not exist), then w is illegal.

Proof:

Given that algorithm p can determine the legitimacy of each mapping $w = \langle \text{IP}, \text{MAC} \rangle$, an arbitrary $w = \langle \text{IP}_x, \text{MAC}_x \rangle$ is selected as input. If the output of algorithm p is false, then w is illegal; if the output is true, then w is legal. The following are performed.

First, host Z is randomly selected in the network. The IP address of host Z should not be in conflict with other nodes. Then, its mapping is recorded as $w' = \langle \text{IP}', \text{MAC}' \rangle$, w' is the input for algorithm p , and the output is observed. If the output is F, then this output is clearly false because w' is a mapping owned by a real host. If the output of the algorithm p is T, then host Z is completely discarded. Then, mapping w' no longer exists in the network (because the MAC address is unrepeatable), that is, w' is illegal.

Given that p is an algorithm, only a fixed result can be an output for a fixed input w ; the output can be inconsistent with a fact at any time. Thus, the algorithm p that can determine whether an arbitrary w is legal does not exist. This conclusion shows that any system that attempts to determine address mapping is imperfect. Conversely, if the address resolution is a decidable problem, the ARP is not needed. The host only needs to decide on all the possible w s one by one to find the MAC address of the target host.

4.2.2. Application of Game Theory and Mechanism Design: Game theory has been widely applied to network security. Such applications include optimal defense of firewall and IDS, security assessments, information security technology, network attack prediction, optimal active defense, and intrusion response [42-45]. By constructing a game model, according to the offensive and defensive strategies of both sides, we can determine the utility matrix to quantify the utility of both offensive and defensive strategies. After which, we can determine the Nash equilibrium and then improve the system based on the optimal strategy.

Mechanism design theory focuses on how to design the rules of the game to achieve the purpose of the designer, which is to maximize the utility of the designer or as fair as possible for each participant. The definitions of incentive compatibility and participation constraint in network attacks are as follows:

(Incentive Compatibility, *IC*): For the attacker, the expected utility of not attacking (normal participation) is higher than attacking.

(Participation Constraint, *IR*): For the attacker, the expected utility of attacking is higher than the maximum expected utility of not attacking (reservation utility).

A real-world problem, called wallet problem, may be taken as an example. You (police) pick up a wallet containing cash (x dollars), and an ID card (full name is y). If you have to return the wallet to the owner, what should you do given the following alternative mechanisms?

Mechanism 1: First, issue a notice saying "I picked up a wallet containing x dollars and the ID card of y ." Then, welcome the owner to claim. The claim rule is that the wallet will be given to the first one who comes to claim.

Mechanism 2: Issue a notice saying "I picked up a wallet. The owner may come to claim it. The claim rule is to answer questions. If you can tell the items inside the wallet (cash count and ID information), then you can take away the wallet, else you may take a negative record (for deceit).

Which one would you choose? Most people would choose the second mechanism because the first is obviously unreasonable. If giving the wallet to the person who first comes to claim it, more and more dishonest impersonators may be encouraged.

We formulate the wallet problem as a three-stage game. The first stage: The police (A) designs a mechanism and then sends a signal. The second stage: a false owner (C) chooses whether to accept the mechanism. If accepted, the third stage, that is, the two sides play the game in accordance with the game mechanism.

In the second stage, if C chooses not to participate, then C will get an exogenous reserve r . If C successfully claims the wallet, the utility is money x and the cost in the process of claim (toll) is y . If cheating behavior is found, C will be fined money g .

If the wallet is given to a fake owner, A is required to pay the true owner, so the utility is $-x$. If A gives the wallet to the true owner, the true owner would reward z .

The probability of C getting the wallet is p , the probability of C failing (getting nothing) is $1 - p$, thus the expected utility for C is

$$Eu_C = p(x) + (1-p)g$$

(IR_C) is

$$p(x) + (1-p)g > r$$

The expected utility for A is

$$Eu_A = p(-x) + (1-p)z$$

(IR_A) is:

$$p(-x) + (1-p)z > 0$$

To facilitate the analysis, we assume that the toll is $0.01x$, the reward is $0.2x$. In the wallet claim process, one true owner and one impersonator exist. According to the game rule of mechanism 1, due to the absence of any inquiry, mechanism 1 is "first come, first served" (FCFS) in the claim process. For both owner and impersonator, the probability of getting the wallet is the same, thus $p = 0.5$. If the impersonator is found, he/she is not punished, that is, $g = 0$. The exogenous reserve of C is toll ($0.01x$). The expected utility for C is

$$Eu_C = 0.5x$$

Thus, the expected utility for C is greater than the toll $0.01x$, satisfying the participation constraint.

$$Eu_A = 0.5 \cdot (-x) + 0.5 \cdot (0.2x) = -0.4x < 0$$

Eu_A does not satisfy the participation constraint if A is rational, that is, A will choose not to send a signal.

If A chooses mechanism 2, C needs to answer questions. As C does not know the specific money amount in the wallet, C can only guess, thus the probability of success is very low, even if A told C that x is an integer between 0 and 200. The success probability of C is $p = 0.005$. The expected utility for C is $0.005x$, which is lower than the toll $0.01x$.

The expected utility for A is

$$Eu_A = 0.99 \cdot (0.2x) + 0.01 \cdot (-x) = 0.197x$$

Almost every time, A can get 19.7% of the total money as reward. As the IR_C is not met if C is rational, C will give up participating in the second stage (as impersonator).

In the network, the mechanism of address resolution process is similar to mechanism 1 of the wallet problem. The host previously broadcasts the key information in the LAN, the resolution rule is FCFS, and no authentication mechanism is put in place for replies. For deceitful nodes, nothing could be better obviously. To improve the safety of the address resolution process, it can be carried out implementing the following:

- The hidden key information (destination address) in the address resolution process and DAD process should be employed to increase the attack difficulty for the deceitful node.

- The mechanism of the address resolution must be improved so that if the cheating nodes do not abide by the protocol, it cannot obtain much information and more profit.

An improved ARP is designed as follows: host *A* sends out an address resolution request, but the destination address to be resolved is not open. Other hosts reply with their own IP address to *A*. *A* checks all received replies. If an IP matches the required destination address, the address resolution process is successful; otherwise, the address resolution process failed. To reduce the disturbance to the LAN, host *A* can give out the prefix of the destination address to be resolved, only the node whose address matches the prefix needs to respond. This method is a reverse address resolution process.

4.2.3. Cryptography and Security Protocol: Cryptography is used to solve the problem of how to make two parties communicate safely in an unsafe channel [46]. Two parties are generally called Alice and Bob, adversary is usually called Oscar. The security protocol is the message exchange protocol based on cryptography. Its purpose is to provide a variety of security services in the network environment. The security protocol mainly solves such problems as encoding and decoding of information, key agreement, and key distribution; identification of data integrity; identification of the identity of the other party; and realization of digital signature and non-repudiation.

Generally speaking, the solutions to these problems are based on the same premise: a contact has been established between Alice and Bob, Alice can send information to Bob directly, regardless whether the contact is secure or not.

Address resolution is different from problems previously mentioned. In the address resolution problem, how can Alice find Bob needs to be solved. Only after Alice found Bob can they carry out the subsequent encrypted communication. If the first Bob is a fraud, then the subsequent security means is meaningless. Thus, to solve the address resolution problem, a new security protocol must be designed.

The new security protocol should meet the following conditions:

- (1) It is a non-arbitration protocol: it does not require a third party authority, and two parties can complete it.
- (2) Alice cannot disclose that she is looking for Bob, but the protocol should let Bob know that Alice is looking for him.

Condition 1 ensures that a trust server is not required in the network to provide arbitration (identification) service. The purpose of condition 2 is to hide the key information to prevent attacks, which can be realized by a one-way function. For example, if h is a hash function, Alice can announce that she is looking for h (Bob), thus people do not know who Alice is seeking, but Bob will know Alice is searching for him by calculating h (Bob). Here the hash function h is open.

5. Conclusion

As all nodes in the network are assumed reliable, current ARPs are inherently vulnerable. ARPs have three phases: target address acquisition process, DAD, and data structure maintenance. These three phases are subject to many security risks. In the design of an address resolution protocol, efficiency and safety need to be considered. According to the existing research, a balance between safety and efficiency is difficult to achieve. Implementation cost, single point of failure, and communication overhead restrict the development of ARPs. The studies on ARPs security differ in focus and research methods, thus comparing these studies is difficult. Consequently, ARPs lack a unified theoretical foundation and research framework. However, the research gaps show that ARPs research still has much room for improvement and wide-ranging development prospects.

Acknowledgements

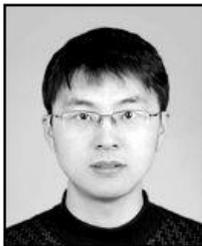
This paper is supported by National Nature Foundation of China under Grant No. 61173024.

References

- [1] Fall, Kevin R., and W. Richard Stevens. TCP/IP illustrated, volume 1: The protocols. Addison-Wesley, (2011).
- [2] Peterson, Larry L., and Bruce S. Davie. Computer networks: a systems approach. Elsevier, (2007).
- [3] Pandey, Poonam. Prevention of ARP spoofing: a probe packet based technique. Proceedings of IEEE 3rd International Conference on Advance Computing (IACC), (2013).
- [4] Tripathi, Nikhil, and B. M. Mehtre. Analysis of various ARP poisoning mitigation techniques: A comparison. Proceedings of IEEE International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), (2014).
- [5] Plummer, David. RFC 826: Ethernet Address Resolution Protocol: Or converting network protocol addresses to 48. bit Ethernet address for transmission on Ethernet hardware, (1982).
- [6] Narten, Thomas, et al. RFC4861: Neighbor Discovery for IP version 6 (IPv6). Standards Track, <http://www.ietf.org/rfc/rfc4861.txt>, (2007).
- [7] J. Arkko, ED, J. Kempf, B. Zill, P. Nikander. RFC 3971: Secure neighbor discovery (SEND). March, (2005).
- [8] Braden, R. RFC 1122: Requirements for Internet Hosts—Communication Layers, (1989) October.
- [9] Hinden, Robert M., and Stephen E. Deering. IP version 6 addressing architecture, (2006).
- [10] Matsumoto A, Fujisaki T, Hiromi R. K. Kanayama. Problem Statement of Default Address Selection in Multi-prefix Environment: Operational Issues of RFC3484 Default Rules.draft-ietf-v6ops-addr-select-ps-01 (work in progress), (2007).
- [11] Cheshire, Stuart. RFC 5227: IPv4 Address conflict detection, (2008).
- [12] Kumar, Sumit, and Shashikala Tapaswi. A centralized detection and prevention technique against ARP poisoning. Proceedings of IEEE International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), (2012).
- [13] Nam, Seung Yeob, Dongwon Kim, and Jeongeun Kim. Enhanced ARP: preventing ARP poisoning-based man-in-the-middle attacks. Communications Letters, IEEE 14.2 (2010), pp.187-189.
- [14] Yang, Xinyu, Ting Ma, and Yi Shi. Typical dos/ddos threats under ipv6.Proceedings of IEEE International Multi-Conference on Computing in the Global Information Technology (ICCGI), (2007).
- [15] Carl-Mitchell, S., and J. S. Quarterman. RFC 1027-using arp to implement transparent subnet gateways, (1987).
- [16] Thomson, S., T. Narten, and T. Jinmei. RFC 4862: Ipv6 stateless address autoconfiguration, (2007) September, Status: Draft Standard.
- [17] Barbhuiya, Ferdous A., Santosh Biswas, and Sukumar Nandi. "An active des based ids for arp spoofing. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC), (2011).
- [18] Neminath, H., et al. A DES approach to intrusion detection system for ARP spoofing attacks. Proceedings of IEEE 18th Mediterranean Conference on Control & Automation (MED), (2010).
- [19] Wu J, Bi J, Li X, Ren G and Xu k. RFC 5210: A source address validation architecture (sava) testbed and deployment experience, (2008).
- [20] Wu, Jianping, Gang Ren, and Xing Li. Source address validation: Architecture and protocol design. Proceedings of IEEE International Conference on Network Protocols (ICNP), (2007).
- [21] Andreasen, F., M. Baugher, and D. Wing. RFC4568: Session description protocol (SDP) security descriptions for media streams, (2006) July.
- [22] García-Martínez, Alberto, and Marcelo Bagnulo. An Integrated Approach to Prevent Address Spoofing in IPv6 Links. Communication letters, IEEE, (2012), Vol. 16, No.11.
- [23] Bagnulo, Marcelo and Alberto García Martínez. SAVI: The IETF standard in address validation. Communications Magazine, IEEE, (2013), Vol. 51, No. 4, pp. 66-73.
- [24] Bi, J, Yao G, Halpen J and Levy-Abegnoli, Ed. SAVI for Mixed Address Assignment Methods Scenario. RFC-draft, (2014).
- [25] XIAO, Pei-yao, and Jun BI. OpenFlow Based Intra-AS Source Address Validation. Journal of Chinese Computer Systems, (2013), Vol. 34, No. 9, pp.1999-2003.
- [26] LI, Jie, Wu jianping, Xu ge and Chen wenlong. An Hierarchical Inter-Domain Authenticated Source Address Validation Solution. Journal of software, (2012), Vol. 35, No.1, pp.85-100.
- [27] Trabelsi, Zouheir, and Wassim El-Hajj. Preventing ARP attacks using a fuzzy-based stateful ARP cache. Proceedings of IEEE International Conference on Communications, (2007).

- [28] Oh, M. Y G, Hong S and Cha S. ASA: Agent-based secure ARP cache management. *Communications Iet*, (2012), Vol.6, No. 7pp.685 - 693.
- [29] Aura, Tuomas. RFC 3972: Cryptographically generated addresses (CGA), (2005).
- [30] Kukec, Ana, Suresha Krishnan, and Shenga Jiang. The Secure Neighbor Discovery (SEND) Hash Threat Analysis. RFC 6274, (2011) June.
- [31] Alsa'deh, Ahmad, Hosnieh Rafiee, and Christoph Meinel. Stopping time condition for practical IPv6 cryptographically generated addresses. *Proceedings of IEEE International Conference on Information Networking (ICOIN)*, (2012).
- [32] Rafiee, Hosnieh, Ahmad Alsa'deh, and Christoph Meinel. Multicore-based auto-scaling SEcure Neighbor Discovery for Windows operating systems. *Proceedings of IEEE International Conference on Information Networking (ICOIN)*, (2012).
- [33] Qadir, Sana, and Mohammad Umar Siddiqi. Cryptographically Generated Addresses (CGAs): A survey and an analysis of performance for use in mobile environment. *IJCSNS International Journal of Computer Science and Network Security*, (2011), Vol.11, No. 2, pp. 24-31.
- [34] Su Guangxue, Wang Wendong, Gong Xiangyang, Jiang Sheng and Gao Xuesong, Que Xirong. A quick CGA generation method. *Proceedings of IEEE International Conference on Future Computer and Communication (ICFCC)*, (2010), Vol. 1, pp. V1-769-V1-773.
- [35] Alsa'deh, Ahmad, Feng Cheng, and Christoph Meinel. CS-CGA: compact and more secure CGA. *Proceedings of IEEE International Conference on Networks (ICON)*, (2011).
- [36] Alsa'deh, Ahmad, and Christoph Meinel. Secure neighbor discovery: Review, challenges, perspectives, and recommendations. *Security & Privacy, IEEE*, (2012), Vol. 10, No. 4pp. 26-34.
- [37] Hou Yi, Wang Zhen-xing, Wang Yu, Zhang and Lian cheng. Routing attack in the ND and SEND mixed environment. *Proceedings of IEEE Fourth International Conference on Multimedia Information Networking and Security (MINES)*, (2012).
- [38] Oh, Hayoung, and Kijoon Chae. An efficient security management in IPv6 network via MCGA. *Proceedings of International Conference on Advanced Communication Technology*, (2007), Vol. 2, pp.1179 - 1181.
- [39] Gouda, Mohamed G, and Chin-Tser Huang. A secure address resolution protocol. *Computer Networks*, (2003), Vol. 41, No. 1, pp.57-71.
- [40] Bruschi, Danilo, Alberto Ornaghi, and Emilia Rosti. S-ARP: a secure address resolution protocol. *Proceedings of IEEE 19th Annual Conference on Computer Security Applications*, (2003).
- [41] Issac, B., and L.A. Mohammed. Secure unicast address resolution protocol (S-UARP) by extending DHCP. *Proceedings of 13th IEEE International Conference on Networks*, (2005), pp.1-6.
- [42] Jiang, Wei, Fang Binxing, Tian Zhihong and Zhang Hongli. Evaluating network security and optimal active defense based on attack-defense game model. *Chinese Journal of Computers*, (2009), Vol. 32, No. 4, pp. 817-827.
- [43] Shi, Jin, Yin Lu, and Li Xie. Dynamic intrusion response based on game theory. *Journal of Computer Research and Development*, (2008), Vol. 45, No. 5, pp. 747-757.
- [44] FENG Ping-Hui, LIAN Yi-Feng and AI Ying-Xia. An Evaluation Model of Vulnerability Exploitation Cost for Network System. *Chinese Journal of Computers*, (2006), Vol. 29, No.8, pp. 817-827.
- [45] Zhang Yongzheng, Fang Binxing and Chi Yue. Risk propagation model for assessing network information systems. *Ruan Jian Xue Bao (Journal of Software)*, (2007), Vol. 18, No. 1, pp. 137-145.
- [46] Schneier, Bruce. *Applied cryptography: protocols, algorithms, and source code in C*. John wiley & Sons, (2007).

Authors



Song Guangjia is currently a Ph.D. candidate at Harbin Institute of Technology, Harbin, China. His research interests include network security, address resolution protocol, academic credibility.



Ji Zhenzhou, he received the B.E., M.S. and Ph.D. degrees in Computer science and technology from Harbin Institute of Technology, Harbin, China. He is currently a professor and Ph.D. supervisor in Harbin Institute of Technology. His research interests include advanced computer architecture, parallel computing technology, computer network security and the QoS system.