Secure Candidate Access Router Discovery

Eunsoo Shim, Jens-Peter Redlich, and Richard D. Gitlin

NEC USA, Inc., C&C Research Labs
4 Independence Way, Princeton, NJ 08540 USA

Abstract--The Candidate Access Router [CAR] discovery protocol is designed for use in wireless IP networks to dynamically collect information about neighboring access routers and their capabilities. This capability enables mobile nodes to dynamically execute low-latency handoffs and to intelligently select a target access router. This paper presents possible approaches for CAR discovery, a security analysis of those approaches, and our novel security CAR discovery mechanisms.

I. INTRODUCTION

A key requirement for high levels of Quality of Service (QoS) in wireless networks is low latency of handoffs. MobileIP [14] is likely the Internet standard for mobility support. The MobileIP working group of the IETF has proposed a low-latency handoff mechanism [11] that can reduce handoff latency significantly compared to base MobileIP. However, the low-latency handoff mechanism requires mapping of the Layer-2 ID of the Base Station (BS) to the IP address of the target Access Router (AR) during the handoff procedure, unless we assume soft handoff or multiple rf interfaces working simultaneously in the Mobile Node (MN).

A MN is often in a location where signals from multiple BSs are above the reception threshold. Those BSs could be based on the same or different wireless technologies. An example for the latter would be a wireless overlay network that consists of GSM/CDMA cellular networks and WLANs. Selecting a BS just based on signal strength does not make much sense in such a case. There are many attributes of a link to a BS, or the associated AR, that can affect the preference such as price, available bandwidth, account requirements, and certain protocol supports. The MN may talk to the AR or the network access server associated with each BS one by one and collect all such attributes itself. But this process may introduce disruption of the application traffic if the MN’s wireless interfaces do not support simultaneous access to multiple BSs and the MN has to go through a complete L2 association procedure for a short message exchange on each wireless link. In this case, the information exchange is not power-efficient compared to the case where the necessary information is given to the MN through the currently active wireless link. Hence, it is desirable for the MN to be able to collect the information about the available BSs and the associated ARs via the currently active air link.

To save power, a MN with, for instance, GSM and WLAN dual interfaces, will likely turn off the WLAN interface in a GSM-only coverage area. Consequently, the MN cannot know the availability of the WLAN coverage even if it enters a WLAN coverage area. In the case, it is desirable for the AR to inform the MN about the availability of WLAN.

When a BS or AR is overloaded because too many MNs are attached, the network may pursue load-balancing by guiding some of the MNs to hand over to other attachment points. This requires that the network knows the congestion states or resource status of every BS or AR to which the MNs can hand over without moving.

To simplify our presentation, we will denote the layer-2 ID of the BS associated to a certain AR as the AR’s L2 ID. Obviously, an AR can have multiple L2 IDs if it serves multiple BSs. One solution for all the abovementioned problems is to statically configure every AR or even MN with the IP addresses and the L2 IDs of all the ARs that can be a target. Such static configuration is inflexible and may not be a viable solution where the target ARs belong to different administrative domains. This leads to a mechanism for dynamic collection of information about possible target ARs.

The CAR (candidate access router) discovery protocol has been proposed for this application, and [22] describes the scope of the protocol and related issues.

A CAR is an AR to which a MN can hand over from the current access router, that is, a candidate for the target access router in the next handoff. We define the AR coverage area as the union of the coverage areas of the BSs associated with the AR. Further, we use the AR’s IP address to represent the IP-level identity of the AR that enables IP-based communication to the target AR. Generally, the coverage of the CAR overlaps at least partially with that of the current AR to which the MN is attached. A neighboring AR can be a candidate AR of a MN that is currently being served or can be served by the current AR. Hence, the set of the neighboring ARs of an AR includes all of the candidate ARs of every MN the AR can serve. Also, as the term ‘neighboring’ indicates, the neighboring access routers can be said to be as the ARs whose geographical coverage overlap, at least partially, or are adjacent to that of the current AR.

The CAR discovery protocol is one of the charter items of the SeaMoby working group in the IETF and two proposals have been submitted [2] [21]. Still the work is in its early stage and security issues have not been thoroughly analyzed and addressed yet. In this paper, we review the security issues and measures for protocols used for dynamic discovery in the Internet, in particular, the Internet routing protocols. Then we review possible CAR discovery mechanisms including the two proposals under discussion in the IETF, and we analyze security threats for the mechanisms and investigate the requirements for secure CAR discovery. We present secure CAR discovery protocols and simulation results for the...
discovery time. A brief conclusion of the paper is given at the end of the paper.

II. RELATED WORK

There are many protocols in the Internet for dynamic discovery of certain network entities. To list a few well-known examples, those are the Internet routing protocols such as OSPF [8] and BGP [16], the ARP protocol [15], the service location protocol [23], and the IPv6’s neighbor discovery protocol [12]. Whenever information is to be discovered dynamically, security becomes an important issue. The Internet routing protocols are examples for protocols that are similar to the CAR discovery protocol. Their integrity is critical for the correct operation of the Internet.

Kumar [6] analyzed the security requirements of network routing protocols and identified two sources of attacks: subverted routers and subverted links. He proposed neighbor-to-neighbor digital signatures of routing updates, the addition of sequence numbers and timestamps to the updates, and the addition of acknowledgements and retransmissions of routing updates for distance-vector routing protocols. Kumar and Crowcroft [7] analyzed the security threats and security measures of the IDRP and particularly link-level encryption on inter-domain links. Digital signatures using public keys on routing messages were proposed by Perlman [13], Murphy and Badger [9][10], Kent, et al [5], and Smith and Garcia-Luna-Aceves [19], and others to provide authentication of the routing messages. Sequence numbers and/or timestamps in the routing messages are used to prevent replay attacks. In particular, [10] lets a router be authorized to advertise a certain address range with the certificate by the parent organization owning the address range since it was required to verify whether the router was supposed to advertise a certain set of reachable prefixes and the authorization is done at each level of domain along the address hierarchy, with the ICANN as the root authority. Authorization prevents a router from advertising an address space illegitimately, but cannot prevent a router from omitting a certain address space in its UPDATE messages. So authorization is crucial for secure routing even though the infrastructure requirement and computation overhead for verifying the certificates or signatures are substantial. However it cannot still be verified whether the advertisement topology or route information is 100% correct and complete.

III. APPROACHES FOR CAR DISCOVERY

CAR discovery can be divided into two steps: discovering the IP address (or IP-level identity) of the neighboring AR and then finding the capability of the AR. Once two neighboring ARs know each other’s IP address, they can exchange information about their capabilities. In this paper, we focus on how to discover the IP address of the neighboring AR. We pursue distributed and scalable mechanisms, and thus we require each AR collect the information of its own neighboring ARs.

A. Handoff-Based Discovery

This idea is shown in [17][18][21] and depicted in Figure 1. The MN hands over from AR to AR, and thus it will know the IP address of neighboring ARs. In the most straightforward simple form, the MN remembers the IP address of the AR it attached to previously with (old AR) and relays this information to the AR to which it is currently attached after the handoff (current AR). This way the current AR learns the IP address of the old AR. The current AR then informs the old AR of the current AR’s IP address so that the old AR also gets to know the current AR as a neighbor. A variation of this mechanism is that the MN informs the old AR of the current AR’s IP address directly via the wired network [to accommodate those situations where the rf link to the old AR rapidly deteriorates]. Then the old AR gets to know the IP address of the current AR. It is a delicate distinction to differentiate who discovers whom in this approach but it affects the security requirements and the details of the protocol.

B. L2 Beacon-Based Discovery

The MN receives the L2 beacons of the neighboring ARs and informs the current AR of the L2 IDs included in the beacons of the neighboring AR. Then the current AR sends an inquiry including the L2 ID using multicast in the wired network and the AR having the L2 ID replies to it with its IP address [2]. This mechanism is depicted in Figure 2. It is similar to the ARP protocol. One can notice that it may cause excessive traffic overhead if the multicast inquiry messages cross the domain boundaries. As mentioned above, geographical adjacency is independent of domain boundary and thus inter-domain search is inevitable. One can improve the protocol by having a per-domain discovery agent that handles inter-domain inquiry. That is, the access router sends an inquiry using unicast to the per-domain discovery agent and the agent sends a multicast inquiry within the local domain if it does not have an answer. Also, if the discovery agent does not get a response from the local domain ARs, it sends an inquiry

![Figure 1. Handoff-Based Discovery, A Simplified View](image1.png)

![Figure 2. L2 Beacon-Based Discovery, A Simple Form](image2.png)
to the discovery agents in other domains using multicast. Each discovery agent may answer the inquiry or sends an inquiry to its own domain ARs using multicast. This new mechanism will have much less traffic overhead since a much smaller number of nodes participate in the inter-domain multicast, but it introduces more infrastructure requirements and complexity.

C. Geographical Information-Based Discovery

Since we are dealing with geographical overlappings of AR coverage areas, one may think the information of the location and the coverage area’s shapes and sizes of the ARs could be distributed to all ARs and each AR could figure out its neighboring ARs from this information. The location information and the coverage area shape and size should be configured statically. In this case, the ARs would flood the information among the ARs using multicast much like the link state routing protocol like OSPF does. OSPF can use broadcasting, since it advertises on its local links, but the CAR discovery mechanism should use multicast since the ARs are remotely distributed in the wired network. A problem with this approach is that the coverage shape and area are not easy to define precisely. Frequently, the coverage area is affected by geographical objects, such as buildings or walls. The coverage area may not look like a circle even if we consider only two dimensions. Things become much more complicated if we consider three-dimensional coverage, which is necessary for WLANs in multi-story buildings. Another problem with this approach is that the flooding of the geographical information is not scalable over domain boundaries. This approach is distinguished from the former two approaches in that it does not rely on the MN at all.

IV. SECURITY THREAT ANALYSIS AND SECURITY REQUIREMENTS

What harm could attackers cause, to whom, by exploiting the CAR discovery protocol? First, the attacker may have inserted false information into the CAR table or the neighbor table that includes information about the neighboring ARs. If the table is filled up with garbage, the table will become useless and all the advantages of CAR discovery will disappear. For example, low-latency handoff using the CAR table will fail. If the information in the CAR table is used to guide the MN in search for target ARs, the corrupted CAR table results in misguidance of the MNs. The attacker may pursue blocking of handoff to a certain AR or seduce the MNs to handover to a certain AR from which the attacker can learn the passwords of other mobile users. Or the AR simply does not exist, and the MNs may waste power searching for a nonexistent AR.

Second, the attacker may have its computer inserted as an AR in the CAR table of an AR and collect the information of the AR status.

Third, an attacker may simply flood the ARs with the CAR discovery messages so that they have no time to do other jobs. This is a typical DoS (Denial of Service) attack.

Certainly, the CAR discovery protocol should be designed to prevent all of the above harmful behaviors.

We see that the MN plays a key role in the dynamic discovery in the first two approaches, where the MN reports the ID of a neighboring AR. However a mobile device such as a handset or PDA can be lost or stolen and the entire software can be compromised. Unconditional trust of the MN is almost equal to trusting the humans using the MNs, and trusting such a huge number of mobile users is a security nightmare. Thus we consider MNs as not trustable in general. We consider the MN as an information delivery medium rather than an information originator in the CAR discovery process. That is, we see the handoff-based discovery as a process where the old AR provides its IP address to the MN and the MN returns the old AR through the current AR. Furthermore, the L2 beacon-based discovery is such that the neighboring AR sends the L2 beacon or its L2 ID to the MN, and the MN delivers the information back to the sending AR through the current AR.

One approach for the security is pursuing the correctness of the information, that is, making sure the claimed CAR is a real CAR. Another approach is taking a loose position about the truthfulness and use the information about a discovered CAR only for those MNs that reported the CAR. In the latter case, false information from a malicious MN does not affect other MNs. We can say the former approach is to build a shared CAR table for all the MNs, and the latter is to build a separate CAR table for each MN. We will compare the two security approaches in terms of the benefits and the cost for the case of Handoff-Based Discovery.

A. Security Analysis for Handoff-Based Discovery with a Shared CAR Table

The handoff-based discovery depends on the IP address of the old AR being delivered by the MN to the new AR. The key is whether the information that arrives at the new/current AR is true. The old AR is in the best position to judge this. Eventually the current AR should rely on the old AR’s confirmation of the received information. Therefore, the old AR has a dominant position in the discovery mechanism.

We consider two attacks by the MN when delivering the IP address of the old AR: the delayed delivery attack and the third party delivery attack.

In the delayed delivery attack, the MN the MN reports to the current AR the IP address of a remote AR to which the MN was attached quite a while ago prior to several handoffs. Then the remote AR will confirm that the MN was there and as a result the remote AR will consider the current AR as a CAR even though though this is not actually the case. To defeat this attack, we introduce a ticket with a timestamp, which is called an AR ID ticket. That is, now the old AR provides the MN (1) an AR ID ticket with timestamp, (2) the old AR IP address, (3) the old AR L2 ID, (4) the MN home IP address, and (5) the ticket ID. Inserting a magic string and a random string in the ticket and encrypting the whole ticket prevents any possibility of replay of the ticket and enables the authentication of the ticket by the old AR. The old AR discards too old tickets delivered by the MN through the current AR.

Actually, relying on a timestamp is tricky. Since there is no rule about how long a MN should stay with an AR before
handoff, just checking time does not tell whether the current AR is the current AR after one handoff or several handoffs. Therefore the old AR should record two moments: when the ticket is issued for the MN and when the L2 source trigger [11] indicating the MN’s handoff is issued. That is, the old AR should know when the MN hands off from the old AR to detect the delayed relay attack of the MN.

The third party delivery attack is the following. The malicious MN sends the ticket to its prox MN (perhaps over the wired network) and has it delivered to the target access AR. To pass the MN home IP address check, the two MNs switch their home IP addresses immediately after the ticket is delivered to the proxy MN. They can do this right after the MN hands over from the old AR to another AR and right before the proxy MN enters the coverage area of the target current AR. This attack means that we need to verify that the MN delivering the ticket to the current AR is the same MN that received the ticket from the old AR. Furthermore we see that the home IP address of the MN is not really useful to identity the MN in this case. Therefore we need a physical identity of the MN that is not duplicable, compromisable or transferable by software commands or communication to defeat the third-party relay attack. One example of a physical ID satisfying the requirements is the SIM card used in the GSM networks. Authenticating a MN’s physical identity using the SIM card in the GSM network is described in [3][4]. The Neighbor AR Identity message from the current AR includes the physical ID so that the old AR can compare it with the physical ID of the MN in its locally stored record or in the ticket. Figure 3 depicts a handoff-based discovery mechanism described so far.

Now we consider a man-in-the-middle attack that can be done for the Neighbor AR identity message transferred in the wired network. A middle man can capture the Neighbor AR identity message and change the current AR IP address. To defeat this attack, the messages between the old AR and the current AR should be authenticated. This can be done by establishing a security association --- for example, by using IPsec. If the claimed current AR is also malicious and cooperates with the malicious MN, the old AR cannot detect the third-party delivery attack since the claimed current AR is going to lie to the old AR. So any AR participating in the CAR discovery process should be authorized and only verification of the authorization allows the old MN to trust the information from the claimed current AR. A digital signature by the AR attached to the Neighbor AR Identity message can serve as an authorization check.

As the last attack, we consider a denial of service attack by a malicious MN or a malicious AR. A malicious MN can send lots of tickets to the current AR causing the current AR to spend too much time on communication with other ARs. To defeat this attack, the AR should allow only one ticket message from each MN during its attachment and simply ignores any subsequent ticket messages. A malicious node may send lots of Neighbor AR Identity messages to an AR. The AR allows only a limited number of Neighbor AR Identity message from the same node within a certain time period. We risk missing valid Neighbor AR identity messages to defend the AR from the possible denial of service attack.

B. Security Analysis for Handoff-Based Discovery with Separate CAR Tables

Figure 4 depicts the discovery process modified from Figure 3. Since the MN has much less motivation to provide a false CAR IP address in this mechanism, and since the MN knows the IP address of the current AR, the old AR can use the MN more seriously. The old AR provides a secret key to the MN in addition to the ticket and the MN generates a new ticket which contains the original ticket from the old AR and the encrypted current AR’s IP address and L2 ID with the secret key. We assume there is a security association between the MN and the AR to which the MN is attached. Then the current AR’s IP address is inserted in the Neighbor AR Identity message by the current AR and the current AR’s IP address in the modified ticket are compared by the old AR. It is double-checking.

A new concept, MN Class ID card, is introduced in Figure 4, which is related to the scalability aspect of having separate CAR tables. For example, let say an AR has 200 CARs and the number of MNs is 10,000,000. This is possible for an AR covering several cells of a cellular network in the metropolitan area where many WLANs are deployed. Storing a physically separate CAR table for each MN is a big overhead in the AR’s...
memory space. We can reduce the memory requirement by having a common CAR table in the memory and assigning to each MN a bitmap each of whose bits corresponds to an entry of the CAR table in the sequential order. The maximum memory requirement then is \([\text{total number of the MNs}] \times [\text{size of the bitmap}] + [\text{total number of the CARs}] \times [\text{size of a CAR table entry}]\). As the number of the MNs increases, the memory requirement increases linearly. If the AR must remember which CARs a certain MN discovered, then a linearly increasing memory requirement is unavoidable. A solution is letting the MN remember which CARs it discovered. The MN Class ID card contains the abovementioned bitmap and is digitally signed by the AR. So each AR issues a MN Class ID card to any MN that ever reported a CAR to the AR. Then a MN can have many such cards. The AR can use a simple secret key based encryption to digitally sign the card since it is the issuing AR that will read the card. Every time the MN reports a new CAR, the bitmap in the card is updated. And the MN submits the card to the AR when it wants to get information about the AR’s CARs.

Two important security requirements of the mechanism depicted in Figure 3 are omitted in the new mechanism in Figure 4. Those omitted are the L2 trigger and the MN physical ID. An advantage of the separate CAR table scheme is that it has only few requirements for the underlying network. Furthermore the separate CAR table scheme can produce useable CAR tables without authorization of ARs which is required by the shared CAR table scheme. Still access control is required about how much information to be shared with a certain CAR.

C. Security Analysis for L2 Beacon-Based Discovery

Here we consider attacks only by malicious MNs. We assume all the involved ARs are trusted and legitimate. In addition, we assume the MN can receive only L2 beacons from the neighboring ARs, that is, it cannot establish links to the neighboring ARs and thus it cannot communicate with the neighboring ARs while it has an active link with the current AR, which is a minimum requirement for all L2 wireless technologies.

We can think of the same type attacks as for the handoff-based discovery, that is, delayed relay attack and third-party relay attack. We notice that the L2 beacon-based discovery approach is vulnerable to the third-party relay attack. The key idea of the security measure against this attack is checking the identity of the MN relaying the information. But the L2 beacon is broadcasted typically and the BS transmitting the L2 beacon cannot identify the MN receiving the beacon information. Thus, we think the L2 beacon-based discovery approach is not secure and we do not envision an approach to achieve the necessary level of security.

D. Security Analysis for Geographical Information-Based Discovery

Since the MN is not involved and there is no means among the ARs to verify the information delivered by other ARs; security depends entirely on the authentication of the messages and the trust relationship among the ARs. So each AR should be authenticated and then authorized to participate in the process. One way to do this is via a digital signature using the public key on the advertisement message for the coverage area information of each AR. Then the major issue is how an AR can know the public keys of all the ARs. Since the advertisement messages containing the geographical information contains the sending AR’s IP address, an efficient way to distribute the public keys is attaching the certified public key to the advertisement message. Then each administrative domain should have an authority to sign the public keys of the ARs in the domain. Then the public key of external domain’s authority should be signed again by a central authority like ICANN. Like the routing protocol, the advertisement message should include also sequence number and timestamp to prevent replay attacks and allow expiration of obsolete information.

V. SIMULATION RESULTS

From the perspective of the network administrator, if the CAR tables of the ARs in the network are fully filled, that is, each AR discovers all of its CARs, the network can serve any mobile node’s need about CAR information and also the network can do load-balancing at its best effect. When a CAR table of an AR is in such a state, we call the CAR table converged to its full state. The time taken for an AR to discover all its CARs is also the time taken for its CAR table’s convergence.

In the handoff-based discovery mechanism, whenever there is a handoff of MN from an AR to another AR, the old AR discovers the current AR. Assuming a trust relationship between the two AR, the current AR accepts the old AR as one of its CARs as well. So, the discovery time is the time necessary for a handoff to occur between the two ARs. Using the fluid flow mobility model [1] [12], the handoff rate across

![Figure 5](attachment.png)  
**Figure 5.** Attack using a proxy MN on the L2 beacon-based discovery

![Figure 6](attachment.png)  
**Figure 6.** CAR table convergence time vs. MN density
a boundary is

\[ \lambda = \frac{\rho v L}{\pi} \]

where \( \lambda \) is the handoff rate or cell boundary-crossing rate (1/sec), \( \rho \) is the active mobile node density (1/m²), \( v \) is the mobile moving speed (m/sec), and \( L \) is the cell perimeter (m). So the CAR table convergence time is inversely proportional to the active number of MNs, when the other parameters are the same.

We did a simulation on the ns-2 simulator for the CAR table convergence time. We used the hexagonal cell structure typical for the cellular networks where the micro-cell diameter (distance from a vertex to the opposite vertex) is 200m, the air interface’s bandwidth is 2Mbps, there is an AR for each cell, and each AR is connected to its neighboring ARs with a 10Mbps wired link. We implemented a break-before-make handoff model with L2 beacons every 100ms.

We can see that the CAR table convergence time for the separate CAR table scheme will be very large since the effective MN density is very low for each (logical) CAR table. From the perspective of the MN, the separate CAR table means that the MN cannot achieve a low-latency handoff for every first-time handoff to an AR and the MN won’t be informed of the capabilities of the ARs that the MN has not visited before. The CAR table convergence time does not mean the MN won’t have low-latency handoffs all the time during the period. For example, let say there are three ARs and all of them are neighboring to the others. If a MN roams only between two of them, the MN won’t have a low-latency handoff only once and then all other handoffs will have low latency, even though the MN’s CAR table is not converged. Such a case is not a big problem. However the separate CAR table scheme could be a problem in a wireless overlay network. For example, a GSM/CDMA cell overlaps a WLAN. If a MN is attached to the GSM/CDMA AR and it does not scan the WLAN signals, it won’t be informed of the availability of the WLAN. So a flexible policy is required in applying the separate CAR tables: for example, the CAR entries common to above a threshold number of the CAR tables could be shared for all the MNs, though this is a policy that loosens the security level.

VI. CONCLUSION

We presented three approaches for CAR discovery, and we analyzed the security issues associated each approach. We found that the L2 beacon-based approach is vulnerable to an attack where a malicious MN’s provides a false L2 beacon to the AR. The handoff-based discovery mechanism with shared CAR table can produce a single reliable CAR table, but it has stringent security requirements including authorization of the participating ARs, a mechanism like a L2 trigger to inform the AR the moment when a MN hands off, and the physical identity verification of the participating MNs. On the other hand, the handoff-based discovery mechanism with a separate CAR table can produce separate CAR tables for each MN so that the damage by a malicious MN’s false information does not affect other MNs; this results in reasonably secure mechanism with much less security requirements than the shared CAR table scheme. But this comes with the price of longer CAR table convergence time for each MN. We think the handoff-based approach is most reasonable for CAR discovery considering the security, feasibility and scalability aspects of the approach. Since the SIM card is widely used in commercial cellular networks, the shared CAR table scheme of the handoff-based approach would be most appropriate in large commercial, public wireless networks.

REFERENCES

[3] GSM 03.20: “Digital cellular telecommunications system (Phase 2+); Security related network functions”
[4] GSM 02.17: “Digital cellular telecommunications system (Phase 2+); Subscriber Identity Modules (SIM); Functional characteristics”