Mobile IP and IPSec in Enterprise use

Markku Rantala
Helsinki University of Technology
Department of Electrical and Communications Engineering
Markku.Rantala@spiritco.com

Abstract

A modern enterprise IT infrastructure consists of increasing number of constantly moving mobile devices. These devices often require a secure and seamless connection to the enterprise intranet resources. This connection setup and maintenance should be transparent from the user's point of view. Problems occur when the mobile device is moving through the enterprise intranet boundaries or if the network point of attachment suddenly changes while the mobile device is roaming in the untrusted network. The current standards do not provide a direct solution for the introduced efficiency and security issues.

There has been proposals on how the enterprise mobility requirements could be satisfied by using the current standards and thus preserving the existing entities at the same time. The proposed solutions fix the connectivity issues but on the other hand fall short on performance and security issues.

This paper concentrates in optimizing the dual home agent model [1]. This model has pitfalls on the performance and security issues and therefore several enhancements to it have been introduced [13]. The first part of the paper studies mechanisms for improving the home network detection algorithm. The second part of the paper discusses and analyses different issues related to overhead optimization. In the end, there is a general wrapup.

KEYWORDS: Mobile IP, IPSec, Enterprise, Overhead, Optimization

1 Introduction

When a mobile device is accessing the enterprise intranet resources from a network point of attachment located inside the enterprise intranet no encryption is required. More frankly speaking, no encryption is a must because of the typically used high bandwidth links and the vast amount of the traffic.

When the mobile device's network point of attachment is outside the enterprise intranet boundaries it gains access to the intranet resources only by using a VPN tunnel. As the usual enterprise deployments have firewalls that block access from the untrusted network to the trusted network, using the VPN tunnel is practically the only feasible method for gaining access to the enterprise intranet. The biggest problem with the current IPSec [2] implementation is that is does not support mobility. Therefore, every time when the tunnel's endpoint IP address changes the security associations have to be renegotiated. This is a costly process as the mobile device's network point of attachment can change rapidly in some situations; for example if the device is moving rapidly or when there are two or more equally "good" aerial links seen by the device.

While the mobility is given by the Mobile IP [3] and IPSec resolves the security issues there is no standardized way how to integrate these two to work together. However, the IETF allows the extension of the current standards as long as the extensions conform to the existing standards [4]. The main reason for not allowing any changes to the present standards is that, especially the IPSec, is already widely deployed and the value for the investments must be preserved.

The IETF has, in their work [1, 4], concluded that the most usable solution for this secure mobility problem is to use two home agents. This model includes one home agent located inside the enterprise intranet and one home agent located in the untrusted network. By using the two home agents the costly untrusted network roaming problem describer earlier is avoided as the end points of the IPSec tunnel remain the same (between the home agents). However, this solution has security and excess overhead problems.

The security problems arise when the mobile device is moving outside from the trusted enterprise intranet. In this situation the mobile device must detect when to start encrypting the data. The proposed solution [1] introduces a rather robust home network detection algorithm that may lead to some plaintext leaks to the untrusted network.

The overhead problems are caused by the rather excessive use of different protocols while the mobile device has registered within the external home agent and is accessing the enterprise's intranet resources through an IPSec VPN tunnel. This encapsulation overhead comes particularly problematic when proportioned to the mobile device's usually restricted link bandwidth.

In this paper, it is assumed that a few of the initial assumptions set up by the IETF can be relaxed in order to find new solutions for the problems. The paper is structures in two rather separate sections of which the first one has emphasis on the security issues and the second one discusses mostly performance issues.

In the first section, an enhanced algorithm for the home network detection is introduced. This algorithm preserves the compatibility with the proposed algorithm [1] and introduces a messaging scenario between the internal and external home agents. The goal of the algorithm is to find a good balance between security and performance.

In the latter section, the focus is on discussing and comparing various methods for the overhead optimization. There are two approaches on how optimizations can be done - either by making some custom tweaks to the existing implementations [5] or by formal methods like the ROHC [6].

In the end, there is a short wrap-up.

2 The home network detection problem

The mobile devices are usually restricted in the terms of the link bandwidth and battery capacity. This causes problems because the requirement is that the data must be encrypted when transferred over the untrusted network. Encryption causes packet overhead that leads to ineffectiveness in the link usage. Encryption also consumes batteries, as it requires relatively expensive calculations. These two characteristics of a mobile device are later, in this text, referred as mobility requirements.

The proposed home network detection algorithm [1] is a compromise between the security and mobility requirements. The proposed solution requires the mobile devices to constantly send re-registration messages to both home agents in order to detect the attachment to the home network. This algorithm is always a compromise between the security and the performance. Shorter intervals between the re-registration messages lead to better security but conflict with the mobility requirements. On the other hand, longer intervals significantly increase the risk that the detection of the movement from the trusted to the untrusted network is not detected and some plaintext data is leaked to the untrusted network. Finding the optimum value for the re-registration interval is not throughoutly discussed in the proposal and here the discussion is left to the end of this section.

2.1 Solution guidelines

The mobile devices usually connect to the Internet using much slower links than the home agents. Combined with the other typical characteristics of the mobile devices (low processing capacity, restricted battery life) it is easy to see that the main bottleneck resides in the mobile device end of the network. Of course, the problems rise on the core network side also when the number of mobile devices accessing the enterprise intranet using encrypted tunnels is extremely high but still, from the mobile device user's point of view, the main optimizations from which he/she benefits the most are the ones done in the mobile device end of the network.

In the proposed model [1] the mobile device must send periodical re-registration messages to the network. The messages must be sent to both external and internal home agents. This approach has a few disadvantages discussed in the proposal in more detail. The most expensive problem, in terms of efficiency, from the mobile device's point of view is that if, for some reason, the response from the internal home agent is delayed the mobile device will start a costly IKE negotiation with the external home agent. Here is assumed that the internal home agent is not reachable from the untrusted network. If the home agent response then arrives later, after the IKE negotiation has been initialized, the algorithm forces the mobile device to drop the IKE negotiation (and there is no standardized way to do so) and to start registration process with the internal home agent. This is very undesirable situation as it is extremely resource consuming.

The inter agent messaging scenario, introduced here, proposes a solution for this particular problem. It also addresses the mobile device low bandwidth problem by reducing the number of registration response messages the mobile device receives. This solution also assumes that the internal home agent is not reachable from the untrusted network. This is usually the case as the enterprise firewalls are commonly configured to block inward connections. The external home agent is usually reachable from the trusted network, as the firewalls usually do not regulate outwards traffic. It is assumed that the external home agent is able to communicate with the internal home agent - this requires firewall or VPN server configuration in most of the cases.

The Mobility Agent Advertisement Extension [3] (chapter 2.2.1) contains reserved space in the end of the flags. According to the standard, this unused space should be filled with the zeros and disregarded while received. We add one flag to this reserved space. The flag new flag is defined as:
Flag A - Forces Mobile Device (or Node) to immediately respond to the advertisement

2.2 Inter agent messaging

The mobile device sends registration requests to both home agents. When sending the registration requests the mobile node must use same identification number in the both of the requests. When the external home agent receives the registration request it does not directly send a reply to the mobile device. Instead, the external home agent sends a message to the internal home agent and waits for the response. The message is encapsulated inside the UDP and is delivered to internal home agent by normal routing. There is also a possibility that there is a permanent IPSec VPN tunnel connecting the external home agent to the intranet - in this case the message travels inside the VPN tunnel and benefits from the increased security.

The inter agent messages contain the following fields:

- External home agent IP
- Internal home agent IP
- Mobile device IP address
- Mobile device MAC address
- Identification (acquired from the rrq message)
- Flag D

When the internal home agent receives the message, it first checks whether it has already received an rrq with the identical mobile device identification (IP and identification fields). If so, it sends a message to external home agent with the D (drop) flag activated which tells the external agent to drop the registration process - the internal home agent must do this any time it receives an rrq already acknowledged by the external home agent.

If it has not received an rrq message it then sends a unicast agent advertisement message to the mobile device's current IP address with the A flag set. If the mobile node does not respond in some reasonable time, the internal home agent determines that the mobile node is not located in the trusted network and sends message without D bit set to the external home agent. If the mobile node responds by rrq message, the internal home agent sends two messages to the external home agent with D flag set - one message per identifications used.

If the external home agent does not receive a response from the internal home agent in reasonable time then it sends a registration reply to the mobile device.

From the mobile devices point of view the registration algorithms is very similar to the proposed [1] one. Fundamental difference is that using the inter-agent messaging the mobile device will receive only one reply from the home agents. From the agent side the external home agent's actions are largely controlled by the internal home agent.

2.3 Other issues

The proposed algorithm [1] for the home network detection does not define, in detail, how the interval between re-registration messages should be defined. The length of the interval is both a security and performance issue and therefore should be carefully determined.

The optimum interval for the re-registration is defined by many factors of which the roaming situation, enterprise security policy and link utilization being the most relevant. The enterprise security policy defines the required security level in the enterprise intranet. Analysis of this factor is omitted as being out of the scope of the paper. The link utilization consists of many sub factors of which the link capacity and traffic load being the most relevant ones. It has been proposed [1] that the mobile device could stop sending the re-registration messages when there is no data to send. This approach is otherwise good but falls in short if the mobile device is simultanously moving and not transmitting data.

Making decisions based on the roaming situation requires the use of Layer 2 [7] information and therefore raises up significant vendor interoperatibility issues.

2.4 Link layer information

The length of the re-registration interval while mobile device roams in the company intranet is always a balance between the performance and security. Yet, it should be defined by the real situation, not by intuition. The real situation can be extracted, in most cases, easily from the link layer. The network interface of the mobile device has information about the current link state and available networks. The other kind of information that can be obtained is link quality metrics, connection state, throughput etc. The amount of metrics available is strongly a vendor dependent issue and by now, there is no uniform standard how to do transfer layer 2 information upwards to layer 3.

2.5 Roaming situation awareness

The re-registration interval could initially be set to some empirically obtained value, which compromises with the enterprise security policies and link utilization factors. The interval could then be changed dynamically by the mobile device based on the roaming situation. Typical cases where the re-registration interval should be altered are:

1) The home network's point of attachment link signal is becoming weak and there are other stronger aerial links available. Here the mobile device should set the re-registration interval shorter to avoid plaintext leaks if the point of attachment suddenly changes.

2) If the home network's point of attachment link signal is strong and there are no other aerial links seen by the mobile device's network interface then the re-registration interval should be longer.

IETF has recently formed a workgroup: Detecting Network Attachment (DNA) [8]. This workgroup concentrates on researching methods for more accurate layer 2 and layer 3 network state information exchange. The work focuses on IPv6 and therefore is not directly applicable to IPv4.

3 Overhead optimization

When the mobile device is accessing the enterprise intranet resources outside the intranet boundaries the data must be encrypted. The use of external home agent and IPSec VPN tunneling creates a significant packet overhead. The payload compression is not reasonable because the payload, in usual applications, is already compressed at the application layer. Therefore, the focus is on how the header overhead could be reduced. A wide range of sources [9, 10, 11] indicate that in sequential packet flows exists only scarce amount of differences between the packets. In other words, the headers contain a considerable amount of already known information.

The amount of the overhead is highly dependent of the access method [1] (chapter 3). The significance of the overhead optimization is also closely related to the mobile device's network link characteristics. There are many different type of links available which have more or less common characteristics [12]. As the mobile devices are in subject of the discussion, the links used are usually of a wireless type. The wireless links can be further categorized to a short range, high bandwidth and long range, low bandwidth subtypes. The latter subtype is the one in the focus here because the low link capacity makes the overhead optimization very essential. The preceding case is discussed briefly at the end of this section.

There are two approaches to the header optimization. The other, preferred by the IETF, is to use header compression. The IETF has worked on the header compression issues from the beginning of the 90s [9]. The current trend in the networking is towards wireless networks [12]. The current trend in header compression takes in account the special features of wireless links. The method currently under development in the IETF is called ROHC [10] (RObust Header Compression). The ROHC is designed to be easily extendable to new

protocols and it includes an error correction protocol - crucial due to wireless link characteristics.

The other approach to overhead optimization is to use custom tweaks. The proposed optimization [13] reduces the overhead by reducing encapsulation complexity in the typical use cases. The proposal adds some restrictions to the messaging for example not allowing fragmentation of the packets.

The header optimization in the high bandwidth, low range networks is not a crucial issue. This is because the devices that are capable of using the full bandwidth of the network do not significantly suffer from the packetization overhead. A solution developed to the worst case situation also serves the better cases well.

3.1 Standards vs. tweaks

With the ROHC, a nearly perfect compression is achieved. The major downside is that deploying ROHC requires significant changes to existing deployments. The use of ROHC in the proposed solution conflicts with the the interoperability, standard preservation and deployability issues. When comparing the tweaked solution to ROHC it is clearly a more lightweight solution: easier to deploy and preserves the existing deployments and configurations better. The vendor interoperability is a challenging question as these kind of tweaks are not guaranteed to achieve a de facto standard status. One benefit of the standardized solution is that it is more interoperable. Nevertheless, when the deployment is relative challenging operation, is the standard going to be widely adopted? One motivation for the deployment of ROHC is that the careful use of radio resources is becoming ever more important [14]. The financial matters suggest that ROHC will probably be widely adopted in the future.

If not, then there is not so big difference, in terms on interoperability, between using the tweaks. History has also proved that sometimes custom tweaks can be quite long living and widely adopted, as is the case for example in the Network Address Translation (NAT) which was born when it was realized that the Internet was running short of IPv4 addresses.

4 Analysis

4.1 Use of Layer 2 information

The use of Layer 2 information is a very efficient way to increase the performance of the proposed solution. From the layer 2 the mobile device is able to get precise information about the link status changes for example. Although the layer 2 information is usually easily extractable, there is no standardized way of doing so (yet). Therefore, using the layer 2 information is a very vendor specific feature. This is not always a particular issue because the use of layer 2 information does not directly mean interoperability problems with the proposed solution. For example, the introduced Roaming situation awareness scenario uses layer 2 information but requires no modifications to the proposed solution.

4.2 Inter-agent messaging

The introduced solution offers a way to prevent unnecessary IPSec negotiations. The solution requires some minor changes to the current MobileIP implementation that makes it harder to deploy. The modifications are needed to both home agents and mobile devices. In general, the solution developed is not very deployable and in order to work effectively it has high bandwidth and latency demands on the links in between the external and internal home agents.

Justification for moving signaling from the mobile device aerial link towards links connecting the home agents is that the agents are usually binded to higher bandwidth links. The solution also reduces the number of registration replies sent on the low bandwidth link between the mobile device and home agent(s).

This scenario is the most applicable when the external and internal home agents are physically joint. This violates with the multi vendor interoperability requirement as the external and internal home agents must come from a single vendor. If the agents reside in the same device, the messaging can be implemented in the hardware, which makes the solution more efficient.

4.3 Overhead optimization

The interoperability is one of the most essential features in a mobile environment where the mobile devices usually collaborate with wide number of devices from different vendors. This suggests that the standardization is the best way of doing things. However, this does not directly mean that the existing standards are the best way of doing things. The standardized overhead optimization issues are currently concentrated around the header compression mechanisms. On the other side, the vendors are constantly trying to gain a competitive edge on the competitors. Adhering standards is not simply enough nowadays and therefore vendors constantly present custom solutions and tweaks in their products. Some of these vendor specific features may later become de facto (widely adopted) standards.

5 Conclusion

The home network detection algorithm introduced in the proposed solution has a few shortcoming what comes to combining the security and mobility requirements. Fixing these shortcoming by modifying the existing Mobile IP and IPSec standards does not seem reasonable as this leads to a situation where massive redeployments are required. Nevertheless, much can be done by modifying the mobile device's software only. The use of layer 2 information is a very promising method and the IETF is working on the matter. Because the use of layer 2 information does not necessary require changes to the current Mobile IP and IPSec standards the vendor interoperability and deployability of the solution is good.

The mobile device optimization is a matter of mobility and security requirements dominated by the low-bandwidth link restrictions. The static infrastructure optimizations are mostly security, management and scalability issues while the

bandwidth is not as big issue. The scalability in this case is restricted by the processing power not bandwidth.

It is not obvious, that the proposed solution needs much enhancements. It is working at the present and the mobile device computing power and bandwidth is constantly increasing. The increase in bandwidth is a two-sided question as, at the same time, the demands for the VPN servers increase. This implies that the solution must be as scalable as possible.

References

[1] S. Vaarala Mobile IPv4 Traversal Across IPSec-based VPN Gateways Mobile IP. September 29, 2003

[2] S. Kent, R. Atkinson Security Architecture for the Internet Protocol Network Working Group, November 1998

[3] C. Perkins RFC3344, IP Mobility Support for IPv4 Network Working Group, August 2002

[4] F. Adrangi Ed., H. Levkowetz, Ed. Problem Statement, Mobile IPv4 Traversal of VPN Gateways Mobile IP Working Group, February 14, 2004

[5] S. Vaarala, A. Nuopponen, F. Adrangi, Optimized Mobile IPv4 UDP Encapsulation Network Working Group. January 2004

[6] Robert Price, Richard Finking Formal Notation for Robust Header Compression Network Working Group. October 27, 2003

[7] ISO Standard The OSI Reference Model March 3, 2004

[8] IETF Internet Area Working Group Detecting Network Attachment (DNA) March 3, 2004

[9] Network Working Group, V. Jacobson Compressing TCP IP Headers for Low-Speed Serial Links February 1990

[10] IETF Network Working Group RFC3095 RObust Header Compression, Framework and four profiles: RTP, UDP, ESP, and uncompressed July 2001

[11] Michael Singer Effnet Offers VoIP Cure For "Header Overhead" January 14, 2002

[12] Nortel Networks Marketing Publications The Profitable 3G Wireless Internet and The Road To Ipv6, WP July 2000

[13] IETF Network Working Group, S. Vaarala, A. Nuopponen, F. Adrangi Optimized Mobile IPv4 UDP Encapsulation January 2000

[14] Hans Hannu Signaling Compression Requirements and Assumptions December 2002