Interworking between IP Security and Performance Enhancing Proxies for Mobile Networks

Nadim Assaf, Jijun Luo, Markus Dillinger and Luis Menendez, SIEMENS AG

ABSTRACT

This article proposes the interworking between performance enhancing proxies (PEPs) and IPsec in mobile networks. The low-throughput problem due to TCP/IP in a radio access network is illustrated. Performance comparison among different PEPs implemented in the RAN is carried out in order to optimize spectrum efficiency. By using PEP, end-to-end security is compromised, and we propose a concept to circumvent this problem. Furthermore, we propose a way to utilize PEP for different network loads. In our article we suggest a scheme that allows the coexistence of IPsec and PEP over mobile networks, through adding an intelligent module in the node where the PEP is implemented.

INTRODUCTION

With the exponential growth of the Internet and expected huge success of future mobile systems, the convergence of Internet and wireless multimedia applications and services is of great interest. It will further lead to wireless Internet access with high business penetration in mobile communications. With our ongoing activities in third-generation (3G) mobile networking, we envisage the importance of IP-based services over 3G networks. TCP is tuned to perform successfully in wired networks due to transmitting packets over a stable network, where the major cause of packet losses is congestion in the network. But random bit errors on the wireless link are not properly handled by TCP and cause degradation of TCP performance with respect to throughput. In this case TCP will respond to losses caused by link errors by invoking unnecessary congestion control and avoidance algorithms, resulting in reduction of window size and leading to reduction of overall throughput.

Regarding the problem of degradation of TCP/IP performance when operating over wireless links, there are many proposals to increase IP stack performance. In this article we focus on the concept of performance enhancing proxies (PEPs), which were introduced in a working group of the Internet Engineering Task Force (IETF) [1]. We model a conventional IP-based mobile network in Fig. 1 to explain the implementation approach of PEPs in it. Figure 1 illustrates a typical 3G Universal Mobile Telecommunications System (UMTS) network with an IP-based core network connected to many radio access networks (RANs), where PEPs are implemented. The mobile network is connected to the IP-based backbone through gateways, such as a serving General Packet Radio Service (GPRS) support node (SGSN). TCP connections can be established between core network and mobile hosts/user equipment (UE), whereas the performance of TCP connection is much improved due to PEP deployment in the communication chain. In the remainder of this article we discuss the centralized and distributed modes of PEP implementation. In addition, the observed security concerns and their potential solutions are discussed. After introducing the schematic models of different PEP approaches, we compare the performance of these approaches and give an adaptive solution to utilizing them. The problem of security compatibility and its solution are given, and then we offer our conclusions.

OPERATION MODES OF PEP

Different approaches are proposed to improve the protocol performance over wireless networks, which can be classified by implementation mode (centralized or distributed), as shown in Fig. 2.

In the distributed mode, the TCP connection is implemented end to end between the fixed host and the mobile host without the intervention of a PEP module in the intermediate node, such as a radio network controller (RNC) at the transport layer for performance enhancement.
This mode can be applied where channel conditions are acceptable and do not greatly affect protocol performance.

In the centralized mode, the TCP connection goes through an intermediate node equipped with PEP functionality for performance enhancement. This mode can be configured when channel conditions are more severe (e.g., a system is highly loaded and a mobile host is badly shadowed).

The two modes are based on different approaches with different protocol stack penetrations improving TCP performance in "lossy systems" (i.e., in wireless networks) [2].

In the centralized mode, the PEP is implemented in the RNC. The PEP proxy hides any non-congestion-related losses from the TCP sender and therefore requires no changes to the protocol implementations of existing senders. The intuition behind this approach is that since the problem is local, it should be solved locally (i.e., on the wireless link). In the distributed mode, losses on the wireless link are handled through the use of selective acknowledgment (SACK) or explicit loss notification (ELN), explained later. The table in Fig. 2 shows the advantages and disadvantages of both modes.

One example of a centralized mode approach is a reliable link layer protocol that is TCP-aware, such as SNOOP developed by UC Berkeley [3]. Simulations have shown that it improves performance, especially for relatively high bit error probability. The following describes the main features of this type of protocol.

A PEP module caches incoming packets. In case of packet losses on the wireless link, the PEP module retransmits the lost packets from its cache, and blocks all duplicate acknowledgments and prevents them from reaching the fixed host; hence, congestion control algorithms are not invoked and throughput is preserved.

Although it increases throughput, a number of drawbacks cannot be neglected. If the PEP is

<table>
<thead>
<tr>
<th>Comparison between the modes</th>
<th>Centralized mode</th>
<th>Distributed mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>• High throughput • Works with legacy fixed hosts</td>
<td>• Better security • Less signaling during inter-RNC handover</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>• Signaling overhead during inter-RNC handover (shifting context) • Context is required to transfer for inter-RNC handover • Security is compromised</td>
<td>• Fixed hosts need to understand PEP protocols</td>
</tr>
</tbody>
</table>

Figure 1. Network topology.

Figure 2. A comparison of centralized and distributed modes.
In order to optimally utilize the given spectrum, we need to have knowledge on the performance of different PEP approaches. Figure 3 shows results of simulations presented in [2]. They give a general idea of the relative performance of different approaches in a typical wireless environment with respect to end-to-end TCP connection without PEP, shown by the dashed line [2]. The solid line shows the typical performance of the centralized mode, which gives much higher rate of information throughput (RIT) than end-to-end protocols. The other two protocols (E2E-SACK and E2E-ELN) are the candidates implemented in the distributed mode. The line with the circle shows that the performance of end-to-end TCP using SACK gives the best throughput, and may be the strongest protocol for the distributed mode.

**Adaptively Utilizing PEP**

The RNC of the network has the ability to choose connection modes (i.e., select centralized or distributed mode). We define the switch between centralized and distributed mode as intermode handover in this article. The RNC works as follows to dynamically utilize the PEP algorithms. While in the distributed mode, the RNC continuously monitors channel conditions such as fading and interference levels. Based on the available information, the RNC decides whether to move to the centralized mode or maintain the current mode. As long as the channel conditions are good and interference is low, the radio link control (RLC) can handle the losses, and there is no need to switch to the centralized mode.

**Distributed to Centralized Mode** — The steps of intermode handover from distributed to centralized mode can be carried out as follows:

1. An intermediate node (RNC) monitors the channel of the current mobile connection.
2. If the channel condition is good and the data load is low, repeat step 1; otherwise, the RNC triggers the mobile host to get into the switching procedure.
3. The mobile host has the option to accept the new settings and benefit from the performance enhancement or reject it due to some constraints:
   - **Security constraints:** The security requirement might be violated if the connection is switched back to centralized PEP mode. Since the PEP needs to be part of the security association between the mobile and the fixed host, it must be able to break the end-to-end semantics of the connections by disabling end-to-end use of IPsec in order to act on the encrypted TCP header. Otherwise, the centralized PEP cannot be implemented.
   - **Protocol constraints:** Due to the reality of mobility, the handover process is inevitable. Handover procedures add considerable overhead to connection due to the transfer of the connection state between the intermediate nodes. If the mobile host is compatible with the changes, it will be reconfigured accordingly.
4. The PEP function for the connection will be activated/integrated in the RNC.

Once the negotiation is terminated, the connections that accept the performance enhancement and the involved mobile host that agrees to the above constrains will switch to centralized mode.

---

**Figure 3.** Performance of the four existing protocols.
Centralized to Distributed Mode — Inter-mode handover from centralized to distributed can be carried out as the following steps:

1. The RNC estimates the improvement of the wireless link.
2. The RNC prompts the mobile host to switch back to distributed mode.
3. If the mobile host accepts, it will be reconfigured accordingly and will download the end-to-end PEP functionality, and so will the fixed host. Therefore, the connections that experience improved channel quality will switch back to the distributed mode. One realization of a message sequence chart is shown in Fig. 4.

As can be seen from the simulation results, the implementation of PEP in the intermediate node (base station) conserves high throughput with high bit error rates in the channel, compared to other approaches. But a problem arises when the connection uses IPsec.

In such a situation the user has to sacrifice one service for the sake of the other, because centralized PEP and IPsec are mutually exclusive, as explained later.

**PROBLEMS OF PEP WITH IPSEC**

**THREATS**

To motivate the necessity for security in communications in wireless networks, some attacks are outlined that could be performed by malicious core software if no protection mechanisms are in place.

**Attacks Against the User** — Mobile hosts could be used to compromise the user’s private sphere. Data stored in the terminal, such as address book, date book, and called numbers, could be changed or made available to third parties. The effectiveness of functions for secure e-commerce, such as payment or computation of digital signatures for a user, could be affected. Furthermore, the mobile host could be changed into a surveillance device by installation of manipulated core software that transmits environmental sounds to an attacker. Since the core software controls the user interface, the terminal could appear to be switched off or in standby mode, and the attack would therefore not be noticed by the user. The user could also be harmed by being billed for telephone calls set up by malicious software (especially calls to premium-rate numbers).

**Attacks Against the Wireless Network** — The mobile host contains security functions that protect the communication at the air interface. Malicious software could circumvent these functions, for example, by sending security parameters to external entities and creating a temporary clone (rogue shell). Thus, the attacked user could be impersonated and the mobile network could be used in his expense. Malicious software could also launch denial-of-service attacks against the network and other users. For example, the terminal could transmit with maximum power, interfere with other users’ signals, or generate huge signaling traffic. Depending on the capabilities of the terminal, other radio systems could also be affected. These attacks could be very effective when distributed attacks involving several hosts are launched.

**Attacks Against the Host** — Future networks will meet an environment composed of heterogeneous networks. Therefore, the reconfiguration of a mobile host through an over-the-air method is a crucial problem that needs to be considered. During the reconfiguration process, the manufacturer (and network operators) cannot test the equipment in test laboratories once it has been shipped. The mobile host could be modified in such a way that it does not comply with requirements for type approval and regulation. It could
also be made nonfunctional or, depending on the actual hardware, damaged permanently (e.g., a mobile host downloading a power management function by IP transport). By misusing the power management functions and exhausting the battery, the operating time of a terminal could be dramatically reduced.

PROBLEMS OF INTERWORKING BETWEEN PEP AND IPSEC

In most cases PEP can handle traffic applying security protocols above the transport layer such as Secure Socket Layer (SSL), since the SSL protocol runs above TCP/IP and below higher-level protocols such as HTTP or IMAP, and therefore the TCP header is not encrypted and PEP can manipulate acknowledgments. However, only a limited number of applications include support for the use of transport layer security nowadays. On the other hand, network (IP) layer security, such as IPsec, can generally be used by any application, transparent to the application, offering access control, data integrity, authentication, and confidentiality. Providing all these services guarantees protection for IP and upper layer protocols.

In fact, IPsec not only offers “information” data integrity but also “control” data integrity, such as acknowledgments on the transport layer, so attackers cannot use information such as network statistics (i.e., flow of acknowledgments) to make an attack. An example could be that hackers attack the TCP header and change the window size to zero, forcing the host to go into persist mode and stop sending data. Another is to regenerate the same acknowledgment number, making the host interpret it as duplicate acknowledgments, causing reduction in window size and consequently in throughput. Also, the connection could be made to shut down.

IPsec works in two modes, transport and tunnel, explained briefly here.

Transport mode is used for remote access and site-to-site security, including virtual private networks. The entire packet (header and payload) is encrypted and given a new IPsec header that is divided into an encapsulating security payload (ESP) header field and an IP header field with an authentication trailer, thus hiding the topology of the protected sites (Fig. 5).

In these circumstances, PEP employed in the centralized mode (i.e., implemented in an intermediate node) cannot act on traffic protected by the use of IPsec because it cannot examine the TCP headers of IP packets. Due to encryption of IP packets via IPsec’s ESP header in either transport or tunnel mode, a TCP header is rendered and the payload is unintelligible to the PEP [5].

Therefore, the user or administrator should decide whether to protect data using IPsec or activate the PEP for throughput enhancement, but not both at the same time. From this arises the need to find a way to compromise between data security, integrity, and high throughput.

SOLUTIONS TO THE PROBLEMS

There are some different scenarios the implementation of PEP should be able to support:

- Arrival of nonencrypted packets that require high throughput
- Arrival of encrypted packets where the PEP cannot act on them
- Arrival of encrypted packets where the PEP is part of the security association

In the case of nonencrypted packets, no problems are encountered. The PEP module in the intermediate node can read the TCP headers and hence manipulate the flow of acknowledgments autonomously. In the encrypted IP packet case, there are two options. Either the packets bypass the PEP module and are directly forwarded to the mobile hosts, in which case the connection will not benefit from performance enhancement introduced by the PEP; or the user can trust the PEP in the middle, and IPsec can be used between each end system and the PEP. In general, though, the end system cannot trust PEPs, and this is not as secure as end-to-end security because the traffic might be exposed when it is decrypted for processing. However, there is research underway investigating the possibility of changing the

![Figure 6. A switch in the intermediate node and the flow chart of its decision function.](image)
implementation of IPsec to more easily use PEP. One approach is multi-layer IP security over protocol stacks by partially releasing the freeform IPsec. The TCP header is encrypted as one layer (i.e., the PEP function should get involved with the encryption process), and the PEP should include security associations used to encrypt the TCP header. The payload of TCP is encrypted end to end. This approach still requires trust of PEP but to a much lesser extent. However, it of course increases the complexity of IPsec.

From the user viewpoint, the option of using IPsec with different levels of security requirements is needed. If an application requires a high level of security, IPsec cannot be compromised. If an application does not require high security but high throughput is the greatest demand, security can be compromised to improve the significance of the PEP in the connection. Of course, the terminal should be configured properly to support both encrypted and nonencrypted packet types.

All in all, there must be a smart decision making module in the intermediate node, which will forward IP packets to the PEP in case IPsec is not applied or the PEP is trusted and can decrypt and reencrypt the packets, and let the encrypted packets with IPsec bypass the PEP if it cannot take part in the security association. In order to conceive this, note that in the IP protocol there is a number in the header field. It is valid for both a normal IP header or a new IPsec header, where the header determines the protocol to be used. With this number, an application can easily determine if a packet is protected with IPsec or not and thus make a decision. When the protocol used on top of IP is ESP, the number in this field will have a value of 50 (decimal), as shown in Fig. 5.

In addition, if PEP is trusted (e.g., it has the security association key), the packets will be decrypted, passed to the PEP module, and then reencrypted and sent to their destination. If the PEP is not part of the security association, the packets should bypass the module. Figure 6 illus-

---

**Figure 4.** Message sequence chart of the adaptive usage of PEP.
trates the traffic of IP packets from one terminal to the other through the RNC. It also illustrates the functionality of the decision making module according to security. Moreover, it is possible in most IPsec implementations to have security associations between different IP addresses and switch this security association on or off according to the approach. This will give flexibility to choose to send some packets protected by IPsec and others freely without the use of any protection.

**CONCLUSION**

In this article we explain the adaptive usage of PEPs mainly for security reasons in IPsec connections. An intermediate node in the radio access network is to decide whether a TCP connection should be established directly to the terminal or run over a PEP module in order to benefit from the performance enhancement added by a PEP. Dynamic PEP utilization depends not only on load information, but also on security limitation. If end-to-end security is demanded, any PEP in the communication chain will compromise this requirement. To decide we must consider three communication modes: transmission with nonencrypted packets, with encrypted packets where the PEP does not have the security association key, and with encrypted packets where the PEP has the security association key. We discuss potential security concerns when using PEP and IPsec concurrently, and propose an intelligent switch for interworking between IPsec and PEP. Depending on the chosen communication modes for security we have different approaches to interworking between IPsec and PEP, which is subject to operator policy and user security demands.

**ACKNOWLEDGMENT**

This article is based on our published material from 3Gwireless 2002 by the Delson Group (http://www.delson.org).

**REFERENCES**


