FIREWALLS IN AN OSI-ENVIRONMENT

Nils Harald Berge and Jon Ølnes
Norwegian Computing Centre
P.O.Box 114 Blindern, N-0314 Oslo, Norway
Nils.Harald.Berge@nr.no - Jon.Olnes@nr.no

ABSTRACT

A firewall is a well established security measure for connecting to the Internet (TCP/IP protocol suite). Government procurement profiles for data communication products (GOSIPs - Government OSI Profile) usually demand use of official international standards, as defined by the International Standards Organization (ISO). ISO has defined a framework for Open Systems Interconnection (OSI), and develops protocol specifications (ISO-protocols) to fit in this framework. It can be assumed that firewalls are going to be an important security measure also when using ISO-protocols. But firewall products for ISO-protocols remain still to be seen, and little research has been done regarding firewalls in an OSI-environment.
This paper discusses the consequences of introducing ISO-protocols from a firewalls point of view. As one might suspect, it is not trivial to transform from TCP/IP to OSI in this matter. Fundamental problems are presented, and recommendations are given on how to solve them. Use of a firewall to map between internal and external security policies is discussed.

Keywords: Firewalls, ISO-protocols, OSI

INTRODUCTION

A combination of screening routers/bridges and bastion hosts is often recommended for security reasons when connecting a local network to a larger (public) network. This network perimeter defence strategy, called a firewall, is fairly well established when connecting to the Internet (the TCP/IP protocol suite). Two main methods are currently used to establish Internet firewalls:

- **Application gateways**: These are secure, but inefficient, either non-transparent to users and applications or hard to set up and manage. Only a limited set of applications is supported and special tailoring is needed for each one.
- **Packet filtering**: This method is insecure but more efficient. It is comprehensive and transparent to many protocols and applications. Traditional packet filters are stateless, each packet is examined individually and no context information is kept. They have only a low level protocol understanding, and are difficult to set up and verify.

The two methods are often combined, using a screening router for packet filtering and a dedicated host on the internal network for application level security. The two approaches can also be combined in one physical host, using special firewall software.
When discussing firewalls, there is often confusion with terminology. Related terms such as screening routers, hybrid gateways, proxy gateways, dual-homed gateways and bastian hosts are often used. Later examples will use a more “layer oriented” approach, discussing the functionality of a firewall in terms of what protocol level is considered. The results should be applicable to screening routers when considering the lower layers, and to the other firewall strategies when considering the upper layers.

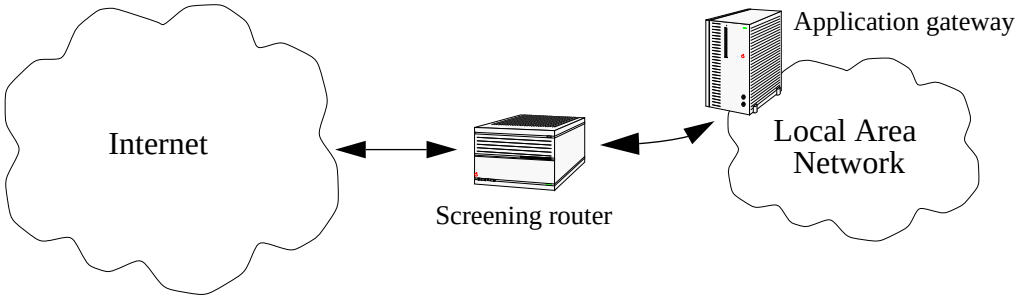

Figure 1: A firewall is often a combination of a screening router - for packet filtering, and an application gateway - for application level security.

A firewall is a relay, which all traffic to and from the internal network passes through. The firewall applies filtering criteria to the traffic, passing on only permitted traffic.

For mainly political reasons, government procurement profiles for data communication products (called “GOSIPs” - Government OSI Profile) usually demand use of official international standards, as defined by the International Standards Organization (ISO). ISO has defined a framework for Open Systems Interconnection (OSI), and develops protocol specifications (ISO-protocols) to fit in this framework. Use of the TCP/IP protocols, which are not ISO specifications, may be acceptable as an interim solution in many GOSIPs, but usually not as a permanent solution.
Contrary to the TCP/IP-world, standards for secure communication exist in the ISO-environment. Still there is no reason to believe that security measures like firewalls are obsolete when ISO-protocols are used. For example:

- the public sector will often have high security demands, including demands for controlling external communication;
- as is always the case, the ISO protocol specifications, and in particular the implementations available, must be expected to contain security relevant weaknesses and bugs.

This paper assumes use of a full ISO protocol stack, with ISO application layer services, connection-oriented transport protocol, class 4 (TP4), and the connection-less network protocol (CLNP). These protocols are outlined as the main options in most GOSIPs. TP4 and CLNP are functionally very similar to TCP/IP, but use of an ISO protocol stack also implies some major differences.
These differences cause problems when TCP/IP firewall solutions are applied in an ISO environment. This paper identifies such problems and gives some suggestions to possible solutions.

RELEVANT DIFFERENCES BETWEEN TCP/IP AND OSI

The application, presentation, and session layers of the OSI-model are collectively referred to as the *upper layers*. These layers are *application oriented*. The remaining layers (transport, network, data link and physical) are referred to as the *lower layers*. The lower layers are *communication oriented.* This division complicates the situation compared to the simpler TCP/IP-protocol stack (see figure 2).

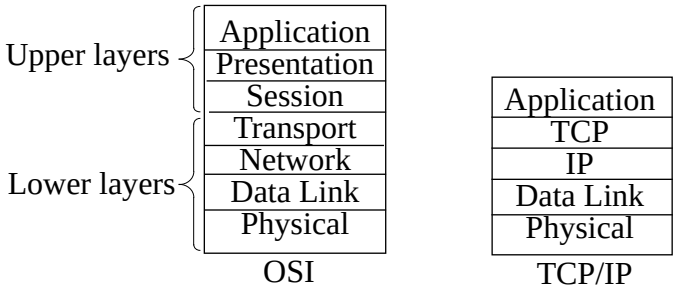

Figure 2: The OSI and TCP/IP protocol stacks

A firewall needs access to both application and communication oriented information to screen traffic in a secure manner. This is simple in the TCP/IP-world because all packets exchanged comprise both types of information. In the OSI-world this is not necessarily the case. Lower layer connections are established prior to upper layer connections, and protocol information concerning the upper layers are not necessarily carried in all lower layer PDUs (Protocol Data Units).
The TCP/IP-stack fits well with the lower layers of the OSI-model. On layer 1 and 2 the same protocols are often used regardless of whether TCP/IP or ISO-protocols are used on higher layers. The most important difference between OSI and TCP/IP is the OSI presentation layer. The presentation layer isolates applications from concerns about the representation (syntax) of the data exchanged, allowing them to deal only with the meaning (semantics). This is achieved by means of the ISO standards ASN.1 (Abstract Syntax Notation One) and BER (Basic Encoding Rules). If a firewall needs access to application layer information, it will have to decode ASN.1, which adds a lot to the processing overhead.

A connection is a much wider term in OSI than in TCP/IP. In OSI all connection oriented protocols have the concept of a logical connection between adjacent layers. Logical connections are in principle established and released independently of connections on other layers. Of special interest are three types of connections:

- **Associations** - between communicating application layer entities. ACSE (Association Control Service Element) is responsible for establishing and releasing an association.
- **Sessions** - between communicating session layer entities.
- **Transport connections** - an end-to-end connection between adjacent transport layer entities.

These connections are important in a filtering context for two reasons:

1. Filtering that depends on application specific information will have to wait until the transport connection and session are established.
2. Sessions and transport connections can be established and released independently of each other, making it possible for different applications to use the same transport connection.

PROBLEMS ASSOCIATED WITH AN OSI-FIREWALL

This section identifies the most important problems associated with an OSI-firewall, and the

next section suggests some possible solutions. Some of these problems were originally identified by [Lazear][7].

- **Connection establishment** - The transport layer establishes an end-to-end transport connection prior to the exchange of higher layer PDUs. In order to filter on application relevant information, a transport connection request (CR TPDU) must therefore always be accepted. This involves a security risk because the destination system will be confirmed to the initiator.
- **Reuse of a transport connection** - The session protocol specification [ITU-T X.225][6] permits reuse of a transport connection. A transport connection can be left open as a “permanent circuit” that can be used again and again by the session layer (for different applications). The concern is that an attacker might compromise a system by first entering via a legal service, and later use the same transport connection for unauthorized access.
- **Error reporting** - Error reporting functions can be used to confirm the existence of systems, and to discover characteristics about a system. This has been done by attackers using ICMP (Internet Control Message Protocol), and there is no reason why error reporting functions should not be exploited when using ISO-protocols. In fact CLNP provides more accurate error reporting than ICMP does, and will therefore provide an attacker with more information to play with.
- **PDU segmentation** - Each protocol layer has the concept of splitting an SDU (Service Data Unit) into suitably sized PDUs. Under IP is this process called *fragmentation* while ISO prefers the term *segmentation*. Segmentation can greatly complicate screening of protocol information. It may force the recognition and reassembly of PDU fragments at different layers before meaningful comparisons can be made. Segmentation can also be used to fool a network or transport layer relay, because the relay will be unable to “poke” in higher layer protocol information.
- **Performance** - Reassembly on different levels can have a serious impact on the performance of the firewall. But other factors add on to the filtering overhead as well. For TCP/IP each packet may be examined individually, while for ISO-protocols a firewall may have to keep track of several consecutive packets in order to gather enough context information for a filtering decision. If application layer information is used for filtering, it is necessary to decode ASN.1. Headers generally have more fields than in corresponding TCP/IP-protocols, OSI-addresses are longer etc. In short, performance of an ISO-protocol firewall must be expected to be poorer than in the TCP/IP case.

RECOMMENDATIONS AND SOLUTIONS

Use of an application layer relay

We expect the traditional screening router/bastion host setup to be used in an OSI-environment as well. However the functionality of a screening router will be strongly limited compared to the TCP/IP counterpart. IP routers may extract information from TCP and application headers, but with ISO-protocols this is virtually impossible to do in general (except for some transport layer information). A bastion host solution is therefore even more important when using ISO-protocols than in a TCP/IP-environment. Application layer solutions are recommended for several reasons:

- Filtering on application layer information is virtually impossible at other layers, especially if segmentation is used on the upper layers.
- Control of reuse of transport connections should be centralized to the firewall, ensuring that

the underlying transport connection is released when the application using the connection terminates. This needs to be done at or above the session layer.
- There are also other reasons, like the ability to offer tailored services from the firewall to the outside, or to make use of standardized security functions at the application layer.

Using the Connection Request TPDU to block incoming calls

The Connection Request (CR) TPDU is used to establish an end-to-end connection. By blocking all incoming CR TPDUs, or CR TPDUs from specific addresses, the transport connection establishment phase may be controlled. Note the difference between blocking CR TPDUs and filtering on NSAP-addresses. If traffic from a specific address is blocked, it is impossible to establish outgoing connections to this address as well (since acknowledgement-packets etc. will be blocked). By filtering on CR TPDUs you can block incoming calls and at the same time allow outgoing calls to the same address.
The same is achieved under TCP/IP by filtering on the ACK and SYN flags in a TCP-packet, but filtering on PDU-type is safer than filtering on a flag that can easily be manipulated with. Filtering on CR TPDUs can be done by a screening router unless segmentation is a problem.

Filtering on service prior to transport connection establishment

A complete address (PSAP-address) that will identify a particular application in an end system has the following structure:

NSAP-address; **T-selector**; **S-selector**; **P-selector**

The NSAP-address identifies a transport entity in a specific end system. The selector values identify the session, presentation and application layer entities. These selector values are carried by the corresponding protocol (i.e. the P-selector is carried by the presentation layer protocol). The P-selector is used to identify a specific application layer entity, and will hence identify the service being accessed. P-selector values are the ISO counterpart of TCP port numbers, however there are no “well known” P-selector values. The values are selected as a local matter to the system offering a service. This implies:

- Addresses in outgoing traffic do not reveal the requested service. Filtering on service can therefore not easily be employed on outgoing traffic.
- Identification of the service can generally not be done until after the transport connection has been established.

Since selector values are assigned as a local matter, these should be chosen in a way that facilitates screening. When composing a PSAP-address, identical T- and P-selector values should be used. The PSAP-address will typical be registered in a catalogue (X.500) or distributed by other means to potential partners. All incoming CR TPDUs will carry the T-selector value for this connection. If that value is identical to the P-selector revealed later, the service will be identified before establishment of the transport connection, and the connection request may be accepted or refused accordingly. A filter can easily be configured as a list of acceptable NSAP-address/T-selector tuples.
Obviously this kind of filtering cannot be applied to outgoing CR TPDUs, unless it is known that the destination system assigns selector values in the specified way.

Transport connection establishment when T-selector ≠ P-selector

If the T-selectors do not reveal the service requested, filtering must be done in two steps (assuming use of protocol information above the transport layer):

- The initiator establishes a transport connection to the firewall. This is done transparently, and as far as the initiator is concerned looks like a transport connection to the destination end system (behind the firewall). The first DT (Data) TPDU transmitted will normally contain information to open upper layer connections. This packet is screened by the firewall, and the requested service is verified.
- After verifying the service, the firewall establishes a transport connection to the destination end system. The connection between the initiator and the firewall is then relayed onto this connection. This is illustrated in figure 3.

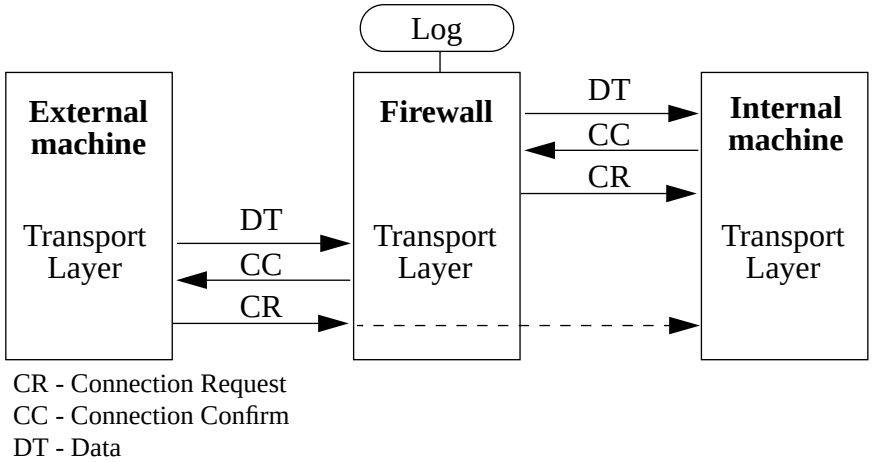

Figure 3: The firewall answers “on behalf of” internal machines, and establishes the transport connection with internal machines.

After the initial screening traffic may be relayed between the two connections by means of a transport layer relay. However an application layer relay is recommended to enable later filtering on higher layer information

All requests are logged by the firewall. This is important not only to detect the origin of an attack, but also to detect the attack itself. If proper logging procedures are not employed it may in fact be very hard to detect that a network has been under attack.

Preventing reuse of a transport connection

The session protocol specification permits reuse of a transport connection. This is illustrated in figure 4.

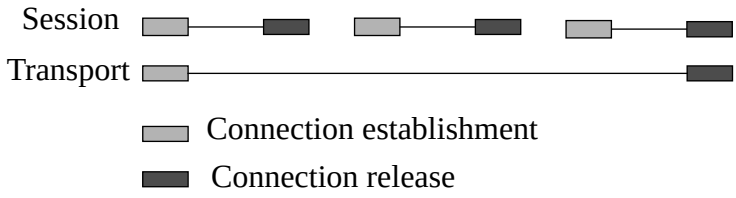

Figure 4: OSI permits multiple sessions on one transport connection.

The efficiency gained is that connection establishment handshaking is avoided for each new session. The security concern is that an “old” transport connection could be used to access an unauthorized service. This may be possible if filtering is applied only during the transport connection establishment phase.

Reuse of the transport connection can only be done if transport expedited data is not available and if:

- the initiator session entity asks to keep the transport connection with parameters in an ABORT SPDU or FINISH SPDU or;
- the initiator session entity receives a REFUSE SPDU or an ABORT SPDU where parameters indicate that the transport connection is to be kept open.

Preventing the ability to reuse a transport connection can easily be done in a firewall working at the session layer or higher, by monitoring the relevant SPDUs and making sure that the underlying transport connections are always terminated whenever a session terminates.
This approach will not work with a screening router or a transport layer relay. Identifying session protocol information concerning specific SPDUs will be virtually impossible due to the amount of context information required.
If the firewall cannot control reuse of transport connections, every system behind the firewall must use session layer implementations that exclude this possibility.

A few words about PDU segmentation

The terms segmentation and fragmentation are equivalent, and refer to the process where a CLNP or IP packet of size N is broken up into smaller pieces if and when it becomes necessary to transmit the packet over a subnetwork for which the maximum packet size is less than N. The OSI-standards use the word segmentation, the TCP/IP-standards prefer fragmentation. The only difference between OSI and TCP/IP in this issue is that OSI segmentation is permitted on every layer, while in TCP/IP fragmentation is a network layer function. For filtering purposes this complicates the situation. Some problems with IP-fragmentation are identified, especially if the firewall consists of a screening router only [Cheswick and Bellovin][1]. Routers normally do not have the ability to reassemble fragmented traffic. This means that TCP protocol information can be distributed over several IP-packets. Different approaches are used to solve this problem:

- In most cases the TCP-header will be present in the first IP-fragment. The method is therefore to apply filtering to the first IP-fragment alone, and let the other fragments flow transparently through the firewall. If the first fragment is blocked, the other fragments will be

reassembled to an incomplete packet and the packet will be discarded anyway.
- Another option is to block all fragmented traffic. This will of course strongly limit the possibilities for external communication.
- Filtering can be applied only to the IP-header. The problem here is that filtering information will be limited (identification of the service is in the TCP-header). There is also a security-risk concerned with trusting IP-header information (it is easy to fake an IP-address).

The same approaches can be used in the CLNP/TP4-case. It is a possible problem that CLNP- and TP4-headers in many cases will be larger that the corresponding IP- and TCP-headers (TP4/CLNP-headers comprise more fields, have longer addresses etc.). It is therefore a higher possibility that the complete TP4-header will not fit in one CLNP-segment.

If the firewall consists of an application level gateway, the gateway will reassemble segments on all layers. The concern in this case is that segmentation on multiple layers can have serious effects on the efficiency of the firewall. In most practical situations it is highly unlikely that segmentation will be used on the upper layers, but there is nothing to prevent an attacker from using this if there is something to gain. It could e.g. be used to “fool” a transport layer relay, or in a *denial-of-service-attack*. To prevent the latter from happening it would be wise to block segmented traffic if the load on the firewall reaches a specific level.

CLNP Error Reporting

Error reporting functions can be used by outsiders to create a map of internal networks, or to gain information about specific hosts. This can e.g. be done by transmitting *echo request* packets, and observing the *echo response*. The response packet will reveal if the adress is legal, if the host is reachable, or if the host is unreacheble but the address is legal. How much error reporting that should be permitted to the outside, through a firewall, is a controversial subject. It is good practice always to have the firewall as source of error-reporting packets, and not internal machines. Table 1 (source [Piscitello and Chapin][8]) compares ICMP (Internet Control Message Protocol) messages and CLNP Error Reports.

Table 1: Comparison of ICMP-Messages and CLNP Error Reports

Category	CLNP Error Report	ICMP Message
General	Reason not specified Protocol procedure error Incorrect checksum PDU discarded-congestion Header syntax error Segmentation needed, not permitted Incomplete PDU received Duplicate option	Parameter problem Parameter problem Parameter problem Source quech Parameter problem Fragmentation needed, but don’t fragmentflag is set Parameter problem Parameter problem
Address-related	Destination address unreacheble Destination address unknown	Network unreacheble Host unreacheble
Source routing (SR)	Unspecified SR error Unknown address in SR-field Path not accepted	SR failed SR failed SR failed

Table 1: Comparison of ICMP-Messages and CLNP Error Reports

Category	CLNP Error Report	ICMP Message
Lifetime	Lifetime expired while data unit was in transit Lifetime expired during reassembly	Time to live exceeded in transit Reassembly time exceeded
PDU discarded	Unsupported option, unspecified error Unsupported protocol version Unsupported security option Unsupported source-routing option Unsupported record-route option	Parameter problem Parameter problem Parameter problem Parameter problem Parameter problem
Reassembly	Reassembly interference	Reassembly time exceeded

The address related messages in table 1 are the most important ones in a filtering context. Other messages may also be exploited by an attacker (like the “Unsupported-messages” to discover specific characteristics about a host). It should be noted that CLNP gives more accurate error reports than ICMP. It can therefore be assumed that CLNP error reports can be used by an attacker to a greater extent than ICMP-messages. This should be carried in mind when configuring an OSI-firewall.

Using Responding-address to “hide” privileged addresses

In the connection establishment phase, all ISO connection oriented protocols use a *responding address* as part of the information in a *Connection Confirm* (CC) packet. Responding address is used to confirm the address given in the Connection Request packet, or to give a new address for the connection. There could be a number of reasons for specifying a new address, e.g. if an error has occurred with the addressed (N)-entity.
From the application to the transport level responding address is:

- Responding-Application-Entity-Title (responding-AE-title)
- Responding-Presentation-Selector (responding-P-selector)
- Responding-Session-Selector (responding-S-selector)
- Responding-Transport-Selector (responding-T-selector)

This can be used by a firewall offering services to the outside. The idea is best illustrated with an example:

Suppose an organization for security reasons has configured their firewall to offer two different FTAM (File Transfer, Access and Management)-services to the outside. One of them, *limited_ftam*, is a subset of the other, *unlimited_ftam*. Limited_ftam is meant to be a service offered to business partners of the organization, while unlimited_ftam is restricted to trusted employees. Limited_ftam is stripped from functionality that is considered as “risky”.
Only limited_ftam is registered (with full address) in a catalogue (X.500), making unlimited_ftam “invisible” from the outside. When someone tries to access the service, the firewall uses a login-procedure, verifying the identity of the initiator. If the initiator is a business partner, the connection is established in the usual manner. If it is a trusted employee, the connection is established, but at the application level the firewall answers with unlimited_ftam as the responding-AE-title. As a consequence the connection is established to a different appli-

cation entity than what was originally specified for the connection. If the initiator is neither a trusted employee or business partner, the request is blocked.

The advantage of using responding address this way, is that it is easy to hide that the user is accessing restricted versions of a service, or services supporting extra security functions. From the outside only restricted (safe) services are visible, and users with extra privileges can transparently access the full service.

CENTRALISING SECURITY FUNCTIONS IN THE FIREWALL

Trusted routers [Hoff][2] are network layer gateways that includes security functions (like encryption, authentication etc.). The idea is to protect information that is exchanged, over possibly several untrusted subnetworks, between two trusted subnetworks. The trusted subnetworks are regarded as safe, and it is only the external communication between the trusted routers that is protected. The benefits of using a trusted router are many:

- Internal machines do not have to think about security concerning external communication. This is centralized in the router.
- All external communication will be subject to the same security policy.
- It is easier to maintain/administer security-protocols, especially key management.

ISO have standardized a framework for security. It is therefore natural to include ISO-security-protocols in an OSI-firewall. The trusted router concept can easily be expanded to include all layers of the OSI-stack, and not just the network layer. A “trusted OSI-firewall” can therefore make use of GULS (Generic Upper Layers Security [ISO/IEC 8182][3]), TLSP (Transport Layer Security Protocol [ISO/IEC 10736][4]), or NLSP (Network Layer Security Protocol [ISO/IEC 11577][5]) when implementing security functions to be used for external communication.
If it is desirable to use protection on the internal network, but the internal security policy is different from the external (for e.g. efficiency reasons), the “trusted OSI-firewall” can map between internal and external security policy. Readers unfamiliar with ISO security protocols and upper layers security model, will find relevant information in [Ford][9].

Figure 5 illustrates the use of TLSP between a firewall, and a trusted end system.
Figure 6 illustrates use of NLSP to achieve trusted router functionality in a screening router. In this figure NLSP is used between the screening router and a trusted external end system, but it could also have been between two trusted screening routers. The same concept will apply to a application layer solution, conforming to the GULS-standard. On what level security functions should be applied must be determined by local policy and demands.

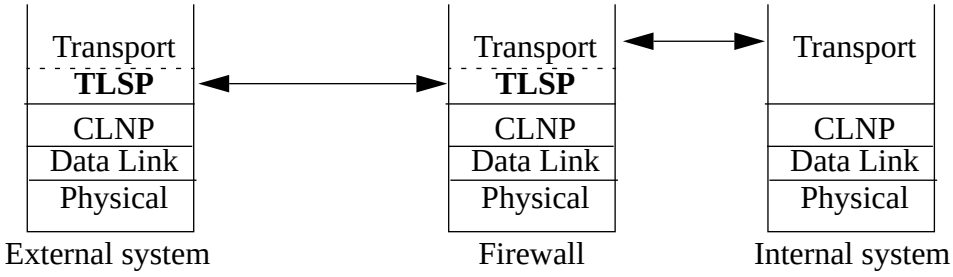

Figure 5: A “trusted firewall” can make use of TLSP to implement security services such as integrity and confidentiality.

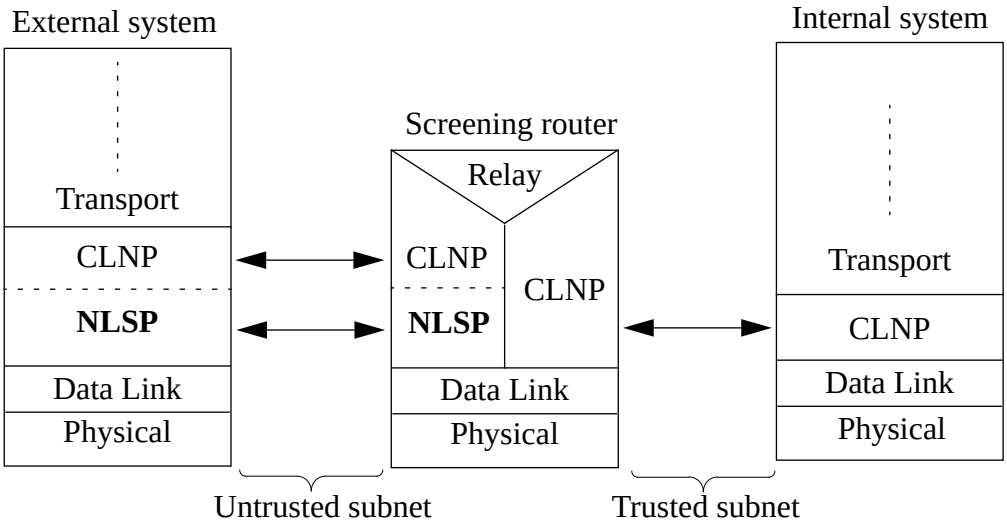

Figure 6: NLSP can be used when including security functions in a firewall. The firewall is responsible for security regarding external communication.

CONCLUSIONS

ISO protocols are optimized by repeating a minimum of information in each PDU. While this improves the overall performance, it limits the conclusions one might reach about the purpose of a particular PDU. As a consequence the functionality of a screening router will be strongly limited compared to the TCP/IP counterpart, because it would require the screening router to keep track of multiple PDUs and log context information, prior to making a filtering decision. The possibility of using segmentation on multiple layers will also complicate the situation for a screening router, one can therefore assume that application level gateways will be even more attractive in an OSI-environment. It is a strong weakness that the transport layer establishment phase does not reveal what application is going to use the connection. We therefore recommend that PSAP addresses are composed in a way that facilitates filtering on service without having to establish the transport connection. This can be done on incoming packets if the T- and P-selector values in a PSAP address are identical. ISO protocols have more accurate error reporting, with the possibility of revealing important information to the outside, and permits reuse of a transport connection, with the possibility of accessing an unauthorized service via an "old" transport connection. These problems are easy to come around, but it is important for a system administrator to be aware of them when configuring a firewall. ISO have standardized a framework for security, it would therefore be natural to include ISO security protocols in a firewall to protect external communication. This concept has been introduced for routers [Hoff][2], but can be expanded to the full OSI-stack, introducing the "Trusted OSI-Firewall".

REFERENCES

1. [Cheswick and Bellovin] William R.Cheswick and Steven M.Bellovin, "Firewalls and Internet Security - Repelling the Wily Hacker", Addison-Wesley Professional Computing Series, 1994.

2. [Hoff] Pål Hoff, "Inter-LAN Security and Trusted Routers",

proceedings of the Internet Society symposium on network and distributed systems security. San Diego, February 1994.

3. [ISO/IEC 8182] Revised text of CD 11586-1, Information Technology - Open Systems Interconnection - Generic Upper Layers Security - Part 1: Overview, Models, and Notation. 9.august 1993.

4. [ISO/IEC 10736] Information Technology - Telecommunication and Information Exchange Between Systems - Transport Layer Security Protocol. 1994.

5. [ISO/IEC 11577] Information Technology - Telecommunication and Information Exchange Between Systems - Network Layer Security Protocol. 1994.

6. [ITU-T X.225] ITU (CCITT) Blue book. Recommendation X.225, Session Protocol Specification, Geneva 1989.

7. [Lazear] Walter D. Lazear, “OSI Packet Filtering Study”, proceedings of the 4th joint European networking conference, Trondheim 10-13 May 1993.

8. [Piscitello and Chapin] David M. Piscitello and A. Lyman Chapin, “Open System Networking - TCP/IP and OSI”, Addison-Wesley Professional Computing Series, 1993.

9. [Ford] Warwick Ford, “Computer Communications Security - Principles, standard protocols and techniques”, Prentice Hall P T R, 1994.