IPSec Virtual Private Networks

150.1 History ................................................................. 2026
150.2 Building Blocks of a Standard ........................................ 2026
150.3 Introduction of Function .............................................. 2027
150.4 Understanding the Foundation ................................. 2028
Finding the Gateway
150.5 Modes of Communication ........................................... 2029
Transport Mode • Tunnel Mode
150.6 Protecting and Verifying Data ........................................ 2031
Authentication and Integrity • Confidentiality and Encryption
150.7 Managing Connections ............................................. 2032
150.8 Establishing a VPN .................................................... 2034
150.9 Keeping Track ........................................................ 2035
Communication Policies • Security Association Control
150.10 Providing Multi-Layered Security Flow ...................... 2036
150.11 A Key Point ............................................................ 2038
150.12 Key History ............................................................ 2038
150.13 IPSec IKE ............................................................ 2038
150.14 Phases and Modes ................................................. 2039
150.15 System Trust Establishment ....................................... 2040
150.16 Key Sharing .......................................................... 2040
150.17 One Key ............................................................... 2041
Many Keys
150.18 Key Establishment .................................................... 2043
Manual Keying • Automatic Keying
150.19 Technology Turned Mainstream .................................. 2044
Performance • Interoperability • Scalability
150.20 The Market for VPN ............................................... 2046
Remote Access • Extranet Access • Internal Protection
150.21 Consideration for VPN Implementation ..................... 2048
System Requirements • Security Policy • Application Performance • Training
150.22 Future of IPSec VPNs ............................................... 2049
150.23 Conclusion ............................................................. 2050

James S. Tiller
The Internet has graduated from simple sharing of e-mail to business-critical applications that involve incredible amounts of private information. The need to protect sensitive data over an untrusted medium has led to the creation of virtual private networks (VPNs). A VPN is the combination of tunneling, encryption, authentication, access control, and auditing technologies and services used to transport traffic over the Internet or any network that uses the TCP/IP protocol suite to communicate.

This chapter:
- Introduces the IPSec standard and the RFCs that make up VPN technology
- Introduces the protocols of the IPSec suite and key management
- Provides a technical explanation of the IPSec communication technology
- Discusses implementation considerations and current examples
- Discusses the future of IPSec VPNs and the industry's support for growth of the standard

### 150.1 History

In 1994, the Internet Architecture Board (IAB) issued a report on “Security in the Internet Architecture” (Request For Comment [RFC] 1636). The report stated the general consensus that the Internet needs more and better security due to the inherent security weaknesses in the TCP/IP protocol suite, and it identified key areas for security improvements. The IAB also mandated that the same security functions become an integral part of the next generation of the IP protocol, IPv6. So, from the beginning, this evolving standard will continually be compatible with future generations of IP and network communication technology.

VPN infancy started in 1995 with the AIAG (Automotive Industry Action Group), a nonprofit association of North American vehicle manufacturers and suppliers, and their creation of the ANX (Automotive Network eXchange) project. The project was spawned to fulfill a need for a TCP/IP network comprised of trading partners, certified service providers, and network exchange points. The requirement demanded efficient and secure electronic communications among subscribers, with only a single connection over unsecured channels. As this technology grew, it became recognized as a solution for any organization wishing to provide secure communications with partners, clients, or any remote network. However, the growth and acceptance had been stymied by the lack of standards and product support issues.

In today's market, VPN adoption has grown enormously as an alternative to private networks. Much of this has been due to many performance improvements and the enhancement of the set of standards. VPN connections must be possible between any two or more types of systems. This can be further defined in three groups:

1. Client to gateway
2. Gateway to gateway
3. Client to client

This process of broad communication support is only possible with detailed standards. IPSec (IP Security protocol) is an ever-growing standard to provide encrypted communications over IP. Its acceptance and robustness have fortified IPSec as the VPN technology standard for the foreseeable future. There are several RFCs that define IPSec, and currently there are over 40 Internet Engineering Task Force (IETF) RFC drafts that address various aspects of the standard's flexibility and growth.

### 150.2 Building Blocks of a Standard

The IPSec standard is used to provide privacy and authentication services at the IP layer. Several RFCs are used to describe this protocol suite. The interrelationship and organization of the documents are important to understand to become aware of the development process of the overall standard.
As Exhibit 150.1 shows, there are seven groups of documents that allow for the association of separate aspects of the IPSec protocol suite to be developed independently while a functioning relationship is attained and managed.

The Architecture is the main description document that covers the overall technology concepts and security considerations. It provides the access point for an initial understanding of the IPSec protocol suite.

The ESP (Encapsulating Security Payload) protocol (RFC 2406) and AH (Authentication Header) protocol (RFC 2402) document groups detail the packet formats and the default standards for packet structure that include implementation algorithms.

The Encryption Algorithm documents are a set of documents that detail the use of various encryption techniques utilized for the ESP. Examples of documents include DES (Data Encryption Standard RFC 1829) and Triple DES (draft-simpson-dsex-02) algorithms and their application in the encryption of the data.

The Authentication Algorithms are a group of documents describing the process and technologies used to provide an authentication mechanism for the AH and ESP protocols. Examples would be HMAC-MD5 (RFC 2403) and HMAC-SHA-1 (RFC 2404).

All of these documents specify values that must be consolidated and defined for cohesiveness into the DOI, or Domain of Interpretation (RFC 2407). The DOI document is part of the IANA assigned numbers mechanism and is a constant for many standards. It provides the central repository for values for the other documents to relate to each other. The DOI contains parameters that are required for the other portions of the protocol to ensure that the definitions are consistent.

The final group is Key Management, which details and tracks the standards that define key management schemes. Examples of the documents in this group are the Internet Security Association and Key Management Protocol (ISAKMP) and Public Key Infrastructure (PKI). This chapter unveils each of these protocols and the technology behind each that makes it the standard of choice in VPNs.

**150.3 Introduction of Function**

IPSec is a suite of protocols used to protect information, authenticate communications, control access, and provide non-repudiation. Of this suite there are two protocols that are the driving elements:

1. Authentication Header (AH)
2. Encapsulating Security Payload (ESP)
AH was designed for integrity, authentication, sequence integrity (replay resistance), and non-repudiation—but not for confidentiality for which the ESP was designed. There are various applications where the use of only an AH is required or stipulated. In applications where confidentiality is not required or not sanctioned by government encryption restrictions, an AH can be employed to ensure integrity, which in itself can be a powerful foe to potential attackers. This type of implementation does not protect the information from dissemination but will allow for verification of the integrity of the information and authentication of the originator. AH also provides protection for the IP header preceding it and selected options. The AH includes the following fields:

- IP Version
- Header Length
- Packet Length
- Identification
- Protocol
- Source and Destination Addresses
- Selected Options

The remainder of the IP header is not used in authentication with AH security protocol. ESP authentication does not cover any IP headers that precede it.

The ESP protocol provides encryption as well as some of the services of the AH. These two protocols can be used separately or combined to obtain the level of service required for a particular application or environmental structure. The ESP authenticating properties are limited compared to the AH due to the non-inclusion of the IP header information in the authentication process. However, ESP can be more than sufficient if only the upper layer protocols need to be authenticated. The application of only ESP to provide authentication, integrity, and confidentiality to the upper layers will increase efficiency over the encapsulation of ESP in the AH. Although authentication and confidentiality are both optional operations, one of the security protocols must be implemented. It is possible to establish communications with just authentication and without encryption or null encryption (RFC 2410). An added feature of the ESP is payload padding, which conceals the size of the packet being transmitted and further protects the characteristics of the communication.

The authenticating process of these protocols is necessary to create a security association (SA), the foundation of an IPSec VPN. An SA is built from the authentication provided by the AH or ESP protocol and becomes the primary function of key management to establish and maintain the SA between systems. Once the SA is achieved, the transport of data can commence.

### 150.4 Understanding the Foundation

Security associations are the infrastructure of IPSec. Of all the portions of IPSec protocol suite, the SA is the focal point for vendor integration and the accomplishment of heterogeneous virtual private networks. SAs are common among all IPSec implementations and must be supported to be IPSec compliant. An SA is nearly synonymous with VPN, but the term "VPN" is used much more loosely. SAs also exist in other security protocols. As described later, much of the key management used with IPSec VPNs is existing technology without specifics defining the underlying security protocol, allowing the key management to support other forms of VPN technology that use SAs.

SAs are simplex in nature in that two SAs are required for authenticated, confidential, bi-directional communications between systems. Each SA can be defined by three components:

1. Security parameter index (SPI)
2. Destination IP address
3. Security protocol identifier (AH or ESP)
An SPI is a 32-bit value used to distinguish among different SAs terminating at the same destination and using the same IPSec protocol. This data allows for the multiplexing of SAs to a single gateway. Interestingly, the destination IP address can be unicast, multicast, or broadcast; however, the standard for managing SAs currently applies to unicast applications or point-to-point SAs. Many vendors will use several SAs to accomplish a point-to-multipoint environment.

The final identification—the security protocol identifier—is the security protocol being utilized for that SA. Note that only one security protocol can be used for communications provided by a single SA. In the event that the communication requires authentication and confidentiality by use of both the AH and ESP security protocols, two or more SAs must be created and added to the traffic stream.

### 150.4.1 Finding the Gateway

Prior to any communication, it is necessary for a map to be constructed and shared among the community of VPN devices. This acts to provide information regarding where to forward data based on the required ultimate destination. A map can contain several pieces of data that exist to provide connection point information for a specific network and to assist the key management process. A map typically will contain a set of IP addresses that define a system, network, or groups of each that are accessible by way of a gateway’s IP address.

An example of a map that specifies how to get to network 10.1.0.0 by a tunnel to 251.111.27.111 and use a shared secret with key management, might look like:

```
begin static -map
target "10.1.0.0/255.255.0.0"
mode "ISAKMP-Shared"
tunnel "251.111.27.111"
end
```

Depending on the vendor implemented, keying information and type may be included in the map. A shared secret or password may be associated with a particular destination. An example is a system that wishes to communicate with a remote network via VPN and needs to know the remote gateway’s IP address and the expected authentication type when communication is initiated. To accomplish this, the map may contain mathematical representations of the shared secret in the map to properly match the secret with the destination gateway. A sample of this is a Diffie–Hellman key, explained in detail later.

### 150.5 Modes of Communication

The type of operation for IPSec connectivity is directly related to the role the system is playing in the VPN or the SA status. There are two modes of operation, as shown in Exhibit 150.2, for IPSec VPNs: transport mode and tunnel mode.

Transport mode is used to protect upper layer protocols and only affects the data in the IP packet. A more dramatic method, tunnel mode, encapsulates the entire IP packet to tunnel the communications in a secured communication.

Transport mode is established when the endpoint is a host, or when communications are terminated at the endpoints. If the gateway in gateway-to-host communications was to use transport mode, it would act as a host system, which can be acceptable for direct protocols to that gateway. Otherwise, tunnel mode is required for gateway services to provide access to internal systems.

#### 150.5.1 Transport Mode

In transport mode, the IP packet contains the security protocol (AH or ESP) located after the original IP header and options and before any upper layer protocols contained in the packet, such as TCP and UDP. When ESP is utilized for the security protocol, the protection, or hash, is only applied to the
upper layer protocols contained in the packet. The IP header information and options are not utilized in the authentication process. Therefore, the originating IP address cannot be verified for integrity against the data. With the use of AH as the security protocol, the protection is extended forward into the IP header to provide integrity of the entire packet by use of portions of the original IP header in the hashing process.

**150.5.2 Tunnel Mode**

Tunnel mode is established for gateway services and is fundamentally an IP tunnel with authentication and encryption. This is the most common mode of operation. Tunnel mode is required for gateway-to-gateway and host-to-gateway communications. Tunnel mode communications have two sets of IP headers—inside and outside.

The outside IP header contains the destination IP address of the VPN gateway. The inside IP header contains the destination IP address of the final system behind the VPN gateway. The security protocol appears after the outer IP header and before the inside IP header. As with transport mode, extended portions of the IP header are utilized with AH that are not included with ESP authentication, ultimately providing integrity only of the inside IP header and payload.

The inside IP header’s TTL (Time To Live) is decreased by one by the encapsulating system to represent the hop count as it passes through the gateway. However, if the gateway is the encapsulating system, as when NAT is implemented for internal hosts, the inside IP header is not modified. In the event the TTL is modified, the checksum must be recreated by IPSec and used to replace the original to reflect the change, maintaining IP packet integrity.

During the creation of the outside IP header, most of the entries and options of the inside header are mapped to the outside. One of these is ToS (Type of Service), which is currently available in IPv4.
150.6 Protecting and Verifying Data

The AH and ESP protocols can provide authentication or integrity for the data, and the ESP can provide encryption support for the data. The security protocol's header contains the necessary information for the accompanying packet. Exhibit 150.3 shows each header's format.

150.6.1 Authentication and Integrity

Security protocols provide authentication and integrity of the packet by use of a message digest of the accompanying data. By definition, the security protocols must use HMAC-MD5 or HMAC-SHA-1 for hashing functions to meet the minimum requirements of the standard. The security protocol uses a hashing algorithm to produce a unique code that represents the original data that was hashed and reduces...
the result into a reasonably sized element called a digest. The original message contained in the packet accompanying the hash can be hashed by the recipient and then compared to the original delivered by the source. By comparing the hashed results, it is possible to determine if the data was modified in transit. If they match, then the message was not modified. If the message hash does not match, then the data has been altered from the time it was hashed. Exhibit 150.4 shows the communication flow and comparison of the hash digest.

**150.6.2 Confidentiality and Encryption**

The two modes of operation affect the implementation of the ESP and the process of encrypting portions of the data being communicated. There is a separate RFC defining each form of encryption and the implementation of encryption for the ESP and the application in the two modes of communication. The standard requires that DES be the default encryption of the ESP. However, many forms of encryption technologies with varying degrees of strength can be applied to the standard. The current list is relatively limited due to the performance issues of high-strength algorithms and the processing required. With the advent of dedicated hardware for encryption processes and the advances in small, strong encryption algorithms such as ECC (Elliptic Curve Cryptosystems), the increase in VPN performance and confidentiality is inevitable.

In transport mode, the data of the original packet is encrypted and becomes the ESP. In tunnel mode, the entire original packet is encrypted and placed into a new IP packet in which the data portion is the ESP containing the original encrypted packet.

**150.7 Managing Connections**

As mentioned earlier, SAs furnish the primary purpose of the IPSec protocol suite and the relationship between gateways and hosts. Several layers of application and standards provide the means for controlling, managing, and tracking SAs.

Various applications may require the unification of services, demanding combined SAs to accomplish the required transport. An example would be an application that requires authentication and confidentiality by utilizing AH and ESP and requires that further groups of SAs provide hierarchical communication. This process is called an SA Bundle, which can provide a layered effect of communications. SA bundles can be utilized by applications in two formats: fine granularity and coarse granularity.

Fine granularity is the assignment of SAs for each communication process. Data transmitted over a single SA is protected by a single security protocol. The data is protected by an AH or ESP, but not both because SAs can have only one security protocol.

Coarse granularity is the combination of services from several applications or systems into a group or portion of an SA bundle. This affords the communication two levels of protection by way of more than one SA.

Exhibit 150.5 conveys the complexity of SAs, and the options available become apparent considering that SAs in a SA bundle can terminate at different locations.

Consider the example of a host on the Internet that established a tunnel-mode SA with a gateway and a transport-mode SA to the final destination internal host behind the gateway. This implementation affords the protection of communications over an untrusted medium and further protection once on the internal network for point-to-point secured communications. It also requires an SA bundle that terminates at different destinations.

There are two implementations of SA Bundles:

1. Transport adjacency
2. Iterated tunneling
Transport adjacency involves applying more than one security protocol to the same IP datagram without implementing tunnel mode for communications. Using both AH and ESP provides a single level of protection and no nesting of communications because the endpoint of the communication is the final destination. This application of transport adjacency is applied when transport mode is implemented for communication between two hosts, each behind a gateway. (See Exhibit 150.5: Example A.)

In contrast, iterated tunneling is the application of multiple layers of security protocols within a tunnel-mode SA(s). This allows for multiple layers of nesting because each SA can originate or terminate at different points in the communication stream. There are three occurrences of iterated tunneling:

- Endpoints of each SA are identical
- One of the endpoints of the SAs is identical
- Neither endpoint of the SAs is identical

Identical endpoints can refer to tunnel-mode communications between two hosts behind a set of gateways where SAs terminate at the hosts and AH (or ESP) is contained in an ESP providing the tunnel. (See Exhibit 150.5: Example B.)

With only one of the endpoints being identical, an SA can be established between the host and gateway and between the host and an internal host behind the gateway. This was used earlier as an example of one of the applications of SA Bundling. (See Exhibit 150.5: Example C.)

In the event of neither SA terminating at the same point, an SA can be established between two gateways and between two hosts behind the gateways. This application provides multi-layered nesting
and communication protection. An example of this application is a VPN between two gateways that provide tunnel mode operations for their corresponding networks to communicate. Hosts on each network are provided secure communication based on client-to-client SAs. This provides for several layers of authentication and data protection. (See Exhibit 150.5: Example D.)

150.8 Establishing a VPN

Now that the components of a VPN have been defined, it is necessary to discuss the form that they create when combined. To be IPSec compliant, four implementation types are required of the VPN. Each type is merely a combination of options and protocols with varying SA control. The four detailed here are only the required formats, and vendors are encouraged to build on the four basic models.

The VPNs shown in Exhibit 150.6 can use either security protocol. The mode of operation is defined by the role of the endpoint—except in client-to-client communications, which can be transport or tunnel mode.

In Example A, two hosts can establish secure peer communications over the Internet. Example B illustrates a typical gateway-to-gateway VPN with the VPN terminating at the gateways to provide connectivity for internal hosts. Example C combines Examples A and B to allow secure communications from host to host in an existing gateway-to-gateway VPN. Example D details the situation when a remote host connects to an ISP, receives an IP address, and then establishes a VPN with the destination network’s gateway. A tunnel is established to the gateway, and then a tunnel- or transport-mode communication is established to the internal system. In this example, it is necessary for the remote system to apply the transport header prior to the tunnel header. Also, it will be necessary for the gateway to allow IPSec connectivity and key management protocols from the Internet to the internal system.

EXHIBIT 150.6 VPN types.
150.9 Keeping Track

Security associations and the variances of their applications can become complicated; levels of security, security protocol implementation, nesting, and SA Bundling all conspire to inhibit interoperability and to decrease management capabilities. To ensure compatibility, fundamental objectives are defined to enable coherent management and control of SAs. There are two primary groups of information, or databases, that are required to be maintained by any system participating in an IPSec VPN Security Policy Database (SPD) and Security Association Database (SAD).

The SPD is concerned with the status, service, or character provided by the SA and the relationships provided. The SAD is used to maintain the parameters of each active association. There are a minimum of two of each database—one for tracking inbound and another for outbound communications.

150.9.1 Communication Policies

The SPD is a security association management construct designed to enforce a policy in the IPSec environment. Consequently, an essential element of SA processing is an underlying security policy that specifies what services are to be offered to IP datagrams and in what fashion they are implemented. SPD is consulted for all IP and IPSec communications, inbound and outbound, and therefore is associated with an interface. An interface that provides IPSec, and ultimately is associated with an SPD, is called a “black” interface. An interface where IPSec is not being performed is called a “red” interface and no data is encrypted for this network by that gateway. The number of SPDs and SADs are directly related to the number of black and red interfaces being supported by the gateway. The SPD must control traffic that is IPSec based and traffic that is not IPSec related. There are three modes of this operation:

1. Forward and do not apply IPSec
2. Discard packet
3. Forward and apply IPSec

In the policy, or database, it is possible to configure traffic that is only IPSec to be forwarded, hence providing a basic firewall function by allowing only IPSec protocol packets into the black interface. A combination will allow multi-tunneling, a term that applies to gateways and hosts. It allows the system to discriminate and forward traffic based on destination, which ultimately determines if the data is encrypted or not. An example is to allow basic browsing from a host on the Internet while providing a secured connection to a remote gateway on the same connection. A remote user may dial an ISP and establish a VPN with the home office to get their mail. While receiving the mail, the user is free to access services on the Internet using the local ISP connection to the Internet.

If IPSec is to be applied to the packet, the SPD policy entry will specify a SA or SA bundle to be employed. Within the specification are the IPSec protocols, mode of operation, encryption algorithms, and any nesting requirements.

A selector is used to apply traffic to a policy. A security policy may determine several SAs be applied for an application in a defined order, and the parameters of this bundled operation must be detailed in the SPD. An example policy entry may specify that all matching traffic be protected by an ESP using DES, nested inside an AH using SHA-1. Each selector is employed to associate the policy to SAD entries.

The policies in the SPD are maintained in an ordered list. Each policy is associated with one or more selectors. Selectors define the IP traffic that characterizes the policy. Selectors have several parameters that define the communication to policy association, including:

- Destination IP address
- Source IP address
- Name
- Data sensitivity
- Transport protocol
- Source and destination TCP ports

Destination address may be unicast, multicast, broadcast, a range of addresses, or a wildcard address. Broadcast, range, and wildcard addresses are used to support more than one destination using the same SA. The destination address defined in the selector is not the destination that is used to define an SA in the SAD (SPI, destination IP address, and IPSec protocol). The destination from the SA identifier is used as the packet arrives to identify the packet in the SAD. The destination address within the selector is obtained from the encapsulating IP header. Once the packet has been processed by the SA and un-encapsulated, its selector is identified by the IP address and associated to the proper policy in the inbound SPD. This issue does not exist in transport mode because only one IP header exists. The source IP address can be any of the types allowed by the destination IP address field.

There are two sets of names that can be included in the Name field: User ID and System Name.

User ID can be a user string associated with a fully qualified domain name (FQDN), as with person@company.com. Another accepted form of user identification is X.500 distinguished name. An example of this type of name could be: C=US,O=Company,OU=Finance,CN=Person. System Name can be a FQDN, box.company.com, or an X.500 distinguished name.

Data sensitivity defines the level of security applied to that packet. This is required for all systems implemented in an environment that uses data labels for information security flow.

Transport protocol and ports are obtained from the header. These values may not be available because of the ESP header or not mapped due to options being utilized in the originating IP header.

### 150.9.2 Security Association Control

The SPD is policy driven and is concerned with system relationships. However, the SAD is responsible for each SA in the communications defined by the SPD. Each SA has an entry in the SAD. The SA entries in the SAD are indexed by the three SA properties: destination IP address, IPSec protocol, and SPI. The SAD database contains nine parameters for processing IPSec protocols and the associated SA:

1. Sequence number counter for outbound communications
2. Sequence number overflow counter that sets an option flag to prevent further communications utilizing the specific SA
3. A 32-bit anti-replay window that is used to identify the packet for that point in time traversing the SA and provides the means to identify that packet for future reference
4. Lifetime of the SA that is determined by a byte count or timeframe, or a combination of the two
5. The algorithm used in the AH
6. The algorithm used in the authenticating the ESP
7. The algorithm used in the encryption of the ESP
8. IPSec mode of operation: transport or tunnel mode
9. Path MTU (PMTU) (this is data that is required for ICMP data over an SA)

Each of these parameters is referenced in the SPD for assignment to policies and applications. The SAD is responsible for the lifetime of the SA, which is defined in the security policy. There are two lifetime settings for each SA: soft lifetime and hard lifetime.

Soft lifetime determines a point when to initiate the process to create a replacement SA. This is typical for rekeying procedures. Hard lifetime is the point where the SA expires. If a replacement SA has not been established, the communications will discontinue.

### 150.10 Providing Multi-Layered Security Flow

There are many systems that institute multi-layered security (MLS), or data labeling, to provide granularity of security based on the data and the systems it may traverse while on the network. This
model of operation can be referred to as Mandatory Access Control (MAC). An example of this security model is the Bell–LaPadula model, designed to protect against the unauthorized transmission of sensitive information. Because the data itself is tagged for review while in transit, several layers of security can be applied. Other forms of security models such as Discretionary Access Control (DAC) that may employ access control lists or filters are not sufficient to support multi-layer security. The AH and ESP can be combined to provide the necessary security policy that may be required for MLS systems working in a MAC environment.

This is accomplished using the authenticating properties of the AH security protocol to bind security mappings in the original IP header to the payload. Using the AH in this manner allows the authentication of the data against the header. Currently, IPv4 does not validate the payload with the header. The sensitivity of the data is assumed only by default of the header.

To accomplish this process each SA, or SA Bundle, must be discernable from other levels of secured information being transmitted. An example is: “SENSITIVE” labeled data will be mapped to a SA or a SA Bundle, while “CLASSIFIED” labeled data will be mapped to others. The SAD and SPD contain a parameter called Sensitivity Information that can be accessed by various implementations to ensure that the data being transferred is afforded the proper encryption level and forwarded to the associated SAs.

There are two forms of processing when MAC is implemented:

1. Inbound operation
2. Outbound operation

When a packet is received and passed to the IPSec functions, the MLS must verify the sensitivity information level prior to passing the datagram to upper layer protocols or forwarding. The sensitivity information level is then bound to the associated SA and stored in the SPD to properly apply policies for that level of secured data.

Outbound requirements of the MLS are to ensure that the selection of a SA, or SA Bundle, is appropriate for the sensitivity of the data, as defined in the policy. The data for this operation is contained in the SAD and SPD, which is modified by defined policies and the previous inbound operations.

Implementations of this process are vendor driven. Defining the level of encryption, type of authentication, key management scheme, and other security-related parameters associated with a data label are available for vendors to implement. The mechanism for defining policies that can be applied is accessible and vendors are beginning to become aware of these options as comfort and maturity of the IPSec standard are realized.
150.11 A Key Point

Key management is an important aspect of IPSec or any encrypted communication that uses keys to provide information confidentiality and integrity. Key management and the protocols utilized are implemented to set up, maintain, and control secure relationships and ultimately the VPN between systems. During key management, there are several layers of system insurance prior to the establishment of an SA, and there are several mechanisms used to accommodate these processes.

150.12 Key History

Key management is far from obvious definition, and lackadaisical conversation with interchanged acronyms only adds to the perceived misunderstandings. The following is an outline of the different protocols that are used to get keys and data from one system to another.

The Internet Security Association and Key Management Protocol (ISAKMP) (RFC 2408) defines the procedures for authenticating a communicating peer and key generation techniques. All of these are necessary to establish and maintain an SA in an Internet environment. ISAKMP defines payloads for exchanging key and authentication data. As shown Exhibit 150.7, these formats provide a consistent framework that is independent of the encryption algorithm, authentication mechanism being implemented, and security protocol, such as IPSec.

The Internet Key Exchange (IKE) protocol (RFC 2409) is a hybrid containing three primary, existing protocols that are combined to provide an IPSec-specific key management platform. The three protocols are:

1. ISAKMP
2. Oakley
3. SKEME (Secure Key Exchange Mechanism)

Different portions of each of these protocols work in conjunction to securely provide keying information specifically for the IETF IPSec DOI. The terms IKE and ISAKMP are used interchangeably by various vendors, and many use ISAKMP to describe the keying function. While this is correct, ISAKMP addresses the procedures and not the technical operations as they pertain to IPSec. IKE is the term that best represents the IPSec implementation of key management.

Public Key Infrastructure (PKI) is a suite of protocols that provide several areas of secure communication based on trust and digital certificates. PKI integrates digital certificates, public key cryptography, and certificate authorities into a total, enterprisewide network security architecture that can be utilized by IPSec.

150.13 IPSec IKE

As described earlier, IKE is a combination of several existing key management protocols that are combined to provide a specific key management system. IKE is considerably complicated, and several variations are available in the establishment of trust and providing keying material.

Oakley and ISAKMP protocols, which are included in IKE, each define separate methods of establishing an authenticated key exchange between systems. Oakley defines modes of operation to build a secure relationship path, and ISAKMP defines phases to accomplish much the same process in a hierarchical format. The relationship between these two is represented by IKE with different exchanges as modes, which operate in one of two phases. Implementing multiple phases may add overhead in processing, resulting in performance degradation, but several advantages can be realized. Some of these are:

- First phase creation assisted by second phase
- First phase key material used in second phase
- First phase trust used for second phase
The first phase session can be disbursed among several second phase operations to provide the construction of new ISAKMP security associations (ISA for purposes of clarity in this document) without the renegotiation process between the peers. This allows for the first phase of subsequent ISAs to be preempted via communications in the second phase.

Another benefit is that the first phase process can provide security services for the second phase in the form of encryption keying material. However, if the first phase does not meet the requirements of the second phase, no data can be exchanged or provided from the first to the second phase.

With the first phase providing peer identification, the second phase may provide the creation of the security protocol SAs without the concern for authentication of the peer. If the first phase were not available, each new SA would need to authenticate the peer system. This function of the first phase is an important feature for IPSec communications. Once peers are authenticated by means of certificates or shared secret, all communications of the second phase and internal to the IPSec SAs are authorized for transport. The remaining authentication is for access control. By this point, the trusted communication has been established at a higher level.

### 150.14 Phases and Modes

Phase one takes place when the two ISAKMP peers establish a secure, authenticated channel with which to communicate. Each system is verified and authenticated against its peer to allow for future communications. Phase two exists to provide keying information and material to assist in the establishment of SAs for an IPSec communication.

Within phase one, there are two modes of operation defined in IKE: main mode and aggressive mode. Each of these accomplishes a phase one secure exchange, and these two modes only exist in phase one. Within phase two, there are two modes: Quick Mode and New Group Mode.

Quick Mode is used to establish SAs on behalf of the underlying security protocol. New Group Mode is designated as a phase two mode only because it must exist in phase two; however, the service provided by New Group Mode is to benefit phase one operations. As described earlier, one of the advantages of a two-phase approach is that the second phase can be used to provide additional ISAs, which eliminates the reauthorization of the peers.

Phase one is initiated using ISAKMP-defined cookies. The initiator cookie (I-cookie) and responder cookie (R-cookie) are used to establish an ISA, which provides end-to-end authenticated communications. That is, ISAKMP communications are bi-directional and, once established, either peer may initiate a Quick Mode to establish SA communications for the security protocol. The order of the cookies is crucial for future second phase operations. A single ISA can be used for many second phase operations, and each second phase operation can be used for several SAs or SA Bundles. Main Mode and Aggressive Mode each use Diffie–Hellman keying material to provide authentication services.

While Main Mode must be implemented, Aggressive Mode is not required. Main Mode provides several messages to authenticate. The first two messages determine a communication policy; the next two messages exchange Diffie–Hellman public data; and the last two messages authenticate the Diffie–Hellman Exchange. Aggressive Mode is an option available to vendors and developers that provides much more information with fewer messages and acknowledgments. The first two messages in Aggressive Mode determine a communication policy and exchange Diffie–Hellman public data. In addition, a second message authenticates the responder, thus completing the negotiation.

Phase two is much simpler in nature in that it provides keying material for the initiation of SAs for the security protocol. This is the point where key management is utilized to maintain the SAs for IPSec communications. The second phase has one mode designed to support IPSec: Quick Mode. Quick Mode verifies and establishes the keying process for the creation of SAs. Not related directly to IPSec SAs is the New Group Mode of operation; New Group provides services for phase one for the creation of additional ISAs.
150.15  System Trust Establishment

The first step in establishing communications is verification of the remote system. There are three primary forms of authenticating a remote system:

1. Shared secret
2. Certificate
3. Public/private key

Of these methods, shared secret is currently used widely due to the relatively slow integration of Certificate Authority (CA) systems and the ease of implementation. However, shared secret is not scalable and can become unmanageable very quickly due to the fact that there can be a separate secret for each communication. Public and private key use is employed in combination with Diffie–Hellman to authenticate and provide keying material. During the system authentication process, hashing algorithms are utilized to protect the authenticating shared secret as it is forwarded over untrusted networks. This process of using hashing to authenticate is nearly identical to the authentication process of an AH security protocol. However, the message—in this case a password—is not sent with the digest. The map previously shared or configured with participating systems will contain the necessary data to be compared to the hash.

An example of this process is a system, called system A, that requires a VPN to a remote system, called system B. By means of a preconfigured map, system A knows to send its hashed shared secret to system B to access a network supported by system B. System B will hash the expected shared secret and compare it to the hash received from system A. If the two hashes match, an authenticated trust relationship is established.

Certificates are a different process of trust establishment. Each device is issued a certificate from a CA. When a remote system requests communication establishment, it will present its certificate. The recipient will query the CA to validate the certificate. The trust is established between the two systems by means of an ultimate trust relationship with the CA and the authenticating system. Seeing that certificates can be made public and are centrally controlled, there is no need to attempt to hash or encrypt the certificate.

150.16  Key Sharing

Once the two systems are confident of each other’s identity, the process of sharing or swapping keys must take place to provide encryption for future communications. The mechanisms that can be utilized to provide keying are related to the type of encryption to be utilized for the ESP. There are two basic forms of keys: symmetrical and asymmetrical.

Symmetrical key encryption occurs when the same key is used for the encryption of information into human unintelligible data (or ciphertext) and the decryption of that ciphertext into the original information format. If the key used in symmetrical encryption is not carefully shared with the participating individuals, an attacker can obtain the key, decrypt the data, view or alter the information, encrypt the data with the stolen key, and forward it to the final destination. This process is defined as a man-in-the-middle attack and, if properly executed, can affect data confidentiality and integrity, rendering the valid participants in the communication oblivious to the exposure and the possible modification of the information.

Asymmetrical keys consist of a key-pair that is mathematically related and generated by a complicated formula. The concept of asymmetrical comes from the fact that the encryption is one way with either of the key-pair, and data that is encrypted with one key can only be decrypted with the other key of the pair. Asymmetrical key encryption is incredibly popular and can be used to enhance the process of symmetrical key sharing. Also, with the use of two keys, digital signatures have evolved and the concept of trust has matured to certificates, which contribute to a more secure relationship.
150.17 One Key

Symmetrical keys are an example of DES encryption, where the same keying information is used to encrypt and decrypt the data. However, to establish communications with a remote system, the key must be made available to the recipient for decryption purposes. In early cases, this may have been a phone call, e-mail, fax, or some form of nonrelated communication medium. However, none of these options are secure or can communicate strong encryption keys that require a sophisticated key that is nearly impossible to convey in a password or phrase.

In 1976, two mathematicians, Bailey W. Diffie at Berkeley and Martin E. Hellman at Stanford, defined the Diffie–Hellman agreement protocol (also known as exponential key agreement) and published it in a paper entitled “New Directions in Cryptography.” The protocol allows two autonomous systems to exchange a secret key over an untrusted network without any prior secrets. Diffie and Hellman postulated that the generation of a key could be accomplished by fundamental relationships between prime numbers. Some years later, Ron Rivest, Adi Shamir, and Leonard Adelman, who developed the RSA Public and Private key cryptosystem based on large prime numbers, further developed the Diffie–Hellman formula (i.e., the nuts and bolts of the protocol). This allowed communication of a symmetrical key without transmitting the actual key, but rather a mathematical portion or fingerprint.

An example of this process is system A and system B require keying material for the DES encryption for the ESP to establish an SA. Each system acquires the Diffie–Hellman parameters, a large prime number \( p \) and a base number \( g \), which must be smaller than \( p - 1 \). The generator, \( g \), is a number that represents every number between 1 and \( p \) to the power of \( k \). Therefore, the relationship is \( g^k \equiv n \mod p \).

Both of these numbers must be hardcoded or retrieved from a remote system. Each system then generates a number \( X \), which must be less than \( p - 2 \). The number \( X \) is typically created by a random string of characters entered by a user or a passphrase that can be combined with date and time to create a unique number. The hardcoded numbers will not be exceeded because most, if not all, applications employ a limit on the input.

![Diffie–Hellman exchange protocol](EXHIBIT 150.8)

EXHIBIT 150.8 Diffie–Hellman exchange protocol.
As shown in Exhibit 150.8, a new key is generated with these numbers, $g^x \mod p$. The result $Y$, or fingerprint, is then shared between the systems over the untrusted network. The formula is then exercised again using the shared data from the other system and the Diffie–Hellman parameters. The results will be mathematically equivalent and can be used to generate a symmetrical key. If each system executes this process successfully, they will have matching symmetrical keys without transmitting the key itself. The Diffie–Hellman protocol was finally patented in 1980 (U.S. Patent 4200770) and is such a strong protocol that there are currently 128 other patents that reference Diffie–Hellman.

To complicate matters, Diffie–Hellman is vulnerable to man-in-the-middle attacks because the peers are not authenticated using Diffie–Hellman. The process is built on the trust established prior to keying material creation. To provide added authentication properties within the Diffie–Hellman procedure, the Station-to-Station (STS) protocol was created. Diffie, Oorschot, and Wiener completed STS in 1992 by allowing the two parties to authenticate themselves to each other by the use of digital signatures created by a public and private key relationship.

An example of this process, as shown in Exhibit 150.9, transpires when each system is provided a public and private key-pair. System A will encrypt the $Y$ value (in this case $Y_A$) with the private key. When system B receives the signature, it can only be decrypted with the system A public key. The only plausible result is that system A encrypted the $Y_A$ value authenticating system A. The STS protocol allows for the use of certificates.
to further authorize the public key of system A to ensure that the man-in-the-middle has not compromised the key-pair integrity.

### 150.17.1 Many Keys

Asymmetrical keys, such as PGP (Pretty Good Privacy) and RSA, can be used to share the keying information. Asymmetrical keys were specifically designed to have one of the keys in a pair published. A sender of data can obtain the public key of the preferred recipient to encrypt data that can only be decrypted by the holder of the corresponding private key. The application of asymmetrical keys in the sharing of information does not require the protection of the public key in transit over an untrusted network.

### 150.18 Key Establishment

The IPSec standard mandates that key management must support two forms of key establishment: manual and automatic.

The other IPSec protocols (AH and ESP) are not typically affected by the type of key management. However, there may be issues with implementing anti-replay options, and the level of authentication can be related to the key management process supported. Indeed, key management can also be related to the ultimate security of the communication. If the key is compromised, the communication can be in danger of attack. To thwart the eventuality of such an attack, there are re-keying mechanisms that attempt to ensure that if a key is compromised its validity is limited either by time, amount of data encrypted, or a combination of both.

#### 150.18.1 Manual Keying

Manual key management requires that an administrator provide the keying material and necessary security association information for communications. Manual techniques are practical for small environments with limited numbers of gateways and hosts. Manual key management does not scale to include many sites in a meshed or partially meshed environment. An example is a company with five sites throughout North America. This organization wants to use the Internet for communications, and each office site must be able to communicate directly with any other office site. If each VPN relationship had a unique key, the number of keys can be calculated by the formula \( n(n-1)/2 \), where \( n \) is the number of sites. In this example, the number of keys is 10. Apply this formula to 25 sites (i.e., five times the number of sites in the previous example) and the number of keys skyrocket to 300, not 50. In reality, the management is more difficult than it may appear by the examples. Each device must be configured, and the keys must be shared with all corresponding systems. The use of manual keying conspires to reduce the flexibility and options of IPSec. Anti-replay, on-demand re-keying, and session-specific key management are not available in manual key creation.

#### 150.18.2 Automatic Keying

Automatic key management responds to the limited manual process and provides for widespread, automated deployment of keys. The goal of IPSec is to build off existing Internet standards to accommodate a fluid approach to interoperability. As described earlier, the IPSec default automated key management is IKE, a hybrid based in ISAKMP. However, based on the structure of the standard, any automatic key management can be employed. Automated key management, when instituted, may create several keys for a single SA. There are various reasons for this, including:

- Encryption algorithm requires more than one key
- Authentication algorithm requires more than one key
Encryption and authentication are used for a single SA

- Re-keying

The encryption and authentication algorithms’ use of multiple keys, or if both algorithms are used, then multiple keys will need to be generated for the SA. An example of this would be if Triple-DES is used to encrypt the data. There are several types of applications of Triple-DES (DES-EEE3, DES-EDE3, and DES-EEE2) and each uses more than one key (DES-EEE2 uses two keys, one of which is used twice).

The process of re-keying is to protect future data transmissions in the event a key is compromised. This process requires the rebuilding of an existing SA. The concept of re-keying during data transmission provides a relatively unpredictable communication flow. Being unpredictable is considered a valuable security method against an attacker.

Automatic key management can provide two primary methods of key provisioning:

1. Multiple string
2. Single string

Multiple strings are passed to the corresponding system in the SA for each key and for each type. For example, the use of Triple-DES for the ESP will require more than one key to be generated for a single type of algorithm, in this case, the encryption algorithm. The recipient will receive a string of data representing a single key; once the transfer has been acknowledged, the next string representing another key will be transmitted.

In contrast, the single string method sends all the required keys in a single string. As one might imagine, this requires a stringent set of rules for management. Great attention is necessary to ensure that the systems involved properly map the corresponding bits to the same key strings for the SA being established. To ensure that IPsec-compliant systems properly map the bit to keys, the string is read from the left, highest bit order first for the encryption key(s) and the remaining string is used for the authentication. The number of bits used is determined by the encryption algorithm and the number of keys required for the encryption being utilized for that SA.

150.19 Technology Turned Mainstream

VPNs are making a huge impact on the way communications are viewed. They are also providing ample fodder for administrators and managers to have seemingly endless discussions about various applications. On one side are the possible money savings, and the other are implementation issues. There are several areas of serious concern, including:

- Performance
- Interoperability
- Scalability
- Flexibility

150.19.1 Performance

Performance of data flow is typically the most common concern, and IPsec is very processor intensive. The performance costs of IPsec are the encryption being performed, integrity checking, packet handling based on policies, and forwarding, all of which become apparent in the form of latency and reduced throughput. IPsec VPNs over the Internet increase the latency in the communication that conspires with the processing costs to discourage VPN as a solution for transport-sensitive applications. Process time for authentication, key management, and integrity verification will produce delay issues with SA
establishment, authentication, and IPSec SA maintenance. Each of these results in poor initialization response and, ultimately, disgruntled users.

The application of existing hardware encryption technology to IPSec vendor products has allowed these solutions to be considered more closely by prospective clients wishing to seize the monetary savings associated with the technology. The creation of a key and its subsequent use in the encryption process can be offloaded onto a dedicated processor that is designed specifically for these operations. Until the application of hardware encryption for IPSec, all data was managed through software computation that was also responsible for many other operations that may be running on the gateway.

Hardware encryption has released IPSec VPN technology into the realm of viable communication solutions. Unfortunately, the client operating system participating in a VPN is still responsible for the IPSec process. Publicly available mobile systems that provide hardware-based encryption for IPSec communications are becoming available, but are some time away from being standard issue for remote users.

150.19.2 Interoperability

Interoperability is a current issue that will soon become antiquated as vendors recognize the need to become fully IPSec compliant—or consumers will not implement their product based simply on its incompatibility. Shared secret and ISAKMP key management protocol are typically allowing multi-vendor interoperability. As Certificate Authorities and the technology that supports them become fully adopted technology, they will only add to the cross-platform integration. However, complex and large VPNs will not be manageable using different vendor products in the near future. Given the complexity, recentness of the IPSec standard, and the various interpretations of that standard, the time to complete interoperability seems great.

150.19.3 Scalability

Scalability is obtained by the addition of equipment and bandwidth. Some vendors have created products focused on remote access for roaming users, while others have concentrated on network-to-network connectivity without much attention to remote users. The current ability to scale the solution will be directly related to the service required. The standard supporting the technology allows for great flexibility in the addition of services. It will be more common to find limitations in equipment configurations than in the standard as it pertains to growth capabilities. Scalability ushers in a wave of varying issues, including:

- Authentication
- Management
- Performance

Authentication can be provided by a number of processes, although the primary focus has been on RADIUS (Remote Access Dial-In User Security), Certificates, and forms of two-factor authentication. Each of these can be applied to several supporting databases. RADIUS is supported by nearly every common authenticating system, from Microsoft Windows NT to NetWare’s NDS. Authentication, when implemented properly, should not become a scalability issue for many implementations, because the goal is to integrate the process with existing or planned enterprise authenticating services.

A more interesting aspect of IPSec vendor implementations and the scalability issues that might arise is management. As detailed earlier, certain implementations do not scale, due to the shear physics of shared secrets and manual key management. In the event of the addition of equipment or increased bandwidth to support remote applications, the management will need to take multiplicity into consideration. Currently, VPN management of remote users and networks leaves a great deal to be desired. As vendors and organizations become more acquainted with what can be accomplished, sophisticated management capabilities will become increasingly available.
Performance is an obvious issue when considering the increase of an implementation. Typically, performance is the driving reason, followed by support for increased numbers. Both of these issues are volatile and interrelated with the hardware technology driving the implementation. Performance capabilities can be controlled by the limitation of supported SAs on a particular system—a direct limitation in scalability. A type of requested encryption might not be available on the encryption processor currently available. Forcing the calculation of encryption onto the operating system ultimately limits the performance. A limitation may resonate in the form of added equipment to accomplish the link between the IPSec equipment and the authenticating database. When users authenticate, the granularity of control over the capabilities of that user may be directly related to the form of authentication. The desired form of authentication may have limitations in various environments due to restrictions in various types of authenticating databases. Upgrade issues, service pack variations, user limitations, and protocol requirements also combine to limit growth of the solution.

150.20 The Market for VPN

Several distinct qualities of VPN are driving the investigation by many organizations to implement VPN as a business interchange technology. VPNs attempt to resolve a variety of current technological limitations that represent themselves as costs in equipment and support or solutions where none had existed prior. Three areas that can be improved by VPNs are:

1. Remote user access and remote office connectivity
2. Extranet partner connectivity
3. Internal departmental security

150.20.1 Remote Access

Providing remote user access via a dial-up connection can become a costly service for any organization to provide. Organizations must consider costs for:

- Telephone lines
- Terminating equipment
- Long-distance
- Calling card
- 800/877 number support

Telephone connections must be increased to support the number of proposed simultaneous users that will be dialing in for connectivity to the network. Another cost that is rolled up into the telephone line charge is the possible need for equipment to allow the addition of telephone lines to an existing system. Terminating equipment, such as modem pools, can become expenses that are immediate savings once the VPN is utilized. Long-distance charges, calling cards that are supplied to roaming users, and toll-free lines require initial capital and continuous financial support. In reality, an organization employing conventional remote access services is nothing more than a service provider for its employees. Taking this into consideration, many organizations tend to overlook the use of the Internet connection by remote users. As the number of simultaneous users access the network, the more bandwidth is utilized for the existing Internet service.

The cost savings are realized by redirecting funds, originally to support telephone communications, in an Internet service provider (ISP) and its ability to support a greater area of access points and technology. This allows an organization to eliminate support for all direct connectivity and focus on a single connection and technology for all data exchange—ultimately saving money. With the company access
point becoming as single point of entry, access controls, authenticating mechanisms, security policies, and system redundancy become focused and common among all types of access regardless of the originator's communication technology.

The advent of high-speed Internet connectivity by means of cable modems and ADSL (Asynchronous Digital Subscriber Line) is an example of how a VPN becomes an enabler to facilitate the need for high-speed, individual remote access where none existed before. Existing remote access technologies are generally limited to 128K ISDN (Integrated Services Digital Network) or, more typically, 56K modem access. Given the inherent properties of the Internet and IPSec functioning at the network layer, the communication technology utilized to access the Internet only needs to be supported at the immediate connection point to establish an IP session with the ISP. Using the Internet as a backbone for encrypted communications allows for equal IP functionality with increased performance and security over conventional remote access technology.

Currently, cable modem and ADSL services are expanding from the home-user market into the business industry for remote office support. A typical remote office will have a small Frame Relay connection to the home office. Any Internet traffic from the remote office is usually forwarded to the home office's Internet connection, where access controls can be centrally managed and Internet connection costs are eliminated at the remote office. However, as the number of remote offices and the distances increase, so does the financial investment. Each Frame Relay connection, PVC (permanent virtual circuit), has costs associated with it. Committed Information Rate (CIR), port speed (e.g., 128K), and sometimes a connection fee add to the overall investment. A PVC is required for any connection; so, as remote offices demand direct communication to their peers, a PVC will need to be added to support this decentralized communication. Currently within the United States, the cost of Frame Relay is very low and typically outweighs the cost of an ISP and Internet connectivity. As the distance increases and moves beyond the United States, the costs can increase exponentially and will typically call for more than one telecommunications vendor. With VPN technology, a local connection to the Internet can be established. Adding connectivity to peers is accomplished by configuration modifications; this allows the customer to control communications without the inclusion of the carrier in the transformation.

The current stability of remote, tier three, and lower ISPs is an unknown variable. The arguable service associated with multiple and international ISP connectivity has become the Achilles' heel for VPN acceptance for business-critical and time-critical services. As the reach of tier one and tier two ISPs increases, they will be able to provide contiguous connectivity over the Internet to remote locations using an arsenal of available technologies.

**150.20.2 Extranet Access**

The single, most advantageous characteristic of VPNs is to provide protected and controlled communication with partnering organizations. Years ago, prior to VPN becoming a catchword, corporations were beginning to feel the need for dedicated Internet access. Dedicated access is becoming increasingly utilized for business purposes, whereas before it was viewed as a service for employees and research requirements.

The Internet provides the ultimate bridge between networks that was relatively nonexistent before VPN technology. Preceding VPNs, a corporation needing to access a partner's site was typically provided a Frame Relay connection to a common Frame Relay cloud where all the partners claimed access. Other options were ISDN and dial-on-demand routing. As this requirement grows, several limitations begin to surface. Security issues, partner support, controlling access, disallowing unwanted interchange between partners, and connectivity support for partners without supported access technologies all conspire to expose the huge advantages of VPNs over the Internet. Utilizing VPNs, an organization can maintain a high granularity of control over the connectivity per partner or per user on a partner network.
150.20.3 Internal Protection

As firewalls became more predominant as protection against the Internet, they were increasingly being utilized for internal segmentation of departmental entities. The need for protecting vital departments within an organization originally spawned this concept of using firewalls internally. As the number of departments increase, the management, complexity, and cost of the firewalls increase as well. Also, any attacker with access to the protected network can easily obtain sensitive information due to the fact that the firewall applies only perimeter security.

VLANs (virtual local area networks) with access control lists became a minimized replacement for conventional firewalls. However, the same security issue remained, in that the perimeter security was controlled and left the internal network open for attack.

As IPSec became accepted as a viable secure communication technology and applied in MAC environments, it also became the replacement for other protection technologies. Combined with strategically placed firewalls, VPN over internal networks allows secure connectivity between hosts. IPSec encryption, authentication, and access control provide protection for data between departments and within a department.

150.21 Consideration for VPN Implementation

The benefits of VPN technology can be realized in varying degrees, depending on the application and the requirements it has been applied to. Considering the incredible growth in technology, the advantages will only increase. Nevertheless, the understandable concerns with performance, reliability, scalability, and implementation issues must be investigated.

150.21.1 System Requirements

The first step is determining the foreseeable amount of traffic and its patterns to ascertain the adjacent system requirements or augmentations. In the event that existing equipment is providing all or a portion of the service the VPN is replacing, the costs can be compared to discover initial savings in the framework of money, performance, or functionality.

150.21.2 Security Policy

It will be necessary to determine if the VPN technology and how it is planned to be implemented meet the current security policy. In case the security policy does not address the area of remote access, or in the event a policy or remote access does not exist, a policy must address the security requirements of the organization and its relationship with the service provided by VPN technology.

150.21.3 Application Performance

As previously discussed, performance is the primary reason VPN technology is not the solution for many organizations. It will be necessary to determine the speed at which an application can execute the essential processes. This is related to the type of data within the VPN. Live traffic or user sessions are incredibly sensitive to any latency in the communication. Pilot tests and load simulation should be considered strongly prior to large-scale VPN deployment or replacement of existing services and equipment.

Data replication or transient activity that is not associated with human or application time sensitivity is a candidate for VPN connectivity. The application's resistance to latency must be measured to determine the minimum requirements for the VPN. This is not to convey that VPNs are only good for replication traffic and cannot support user applications. It is necessary to determine the application needs and verify the requirements to properly gauge the performance provisioning of
the VPN. The performance “window” will allow the proper selection of equipment to meet the needs of the proposed solution; otherwise, the equipment and application may present poor results compared to the expected or planned results. Or, more importantly, the acquired equipment is under-worked or does not scale in the direction needed for a particular organization’s growth path. Each of these results in poor investment realization and makes it much more difficult to persuade management to use VPN again.

150.21.4 Training

User and administrator training is an important part of the implementation process. It is necessary to evaluate a vendor’s product from the point of the users, as well as evaluating the other attributes of the product. In the event that user experience is poor, it will reach management and ultimately weigh heavily on the administrators and security practitioners. It is necessary to understand the user intervention that is required in the every-day process of application use. Comprehending the user knowledge requirements will allow for the creation of a training curriculum that best represents what the users are required to accomplish to operate the VPN as per the security policy.

150.22 Future of IPSec VPNs

Like it or not, VPN is here to stay. IP version 6 (IPv6) has the IPSec entrenched in its very foundation; and as the Internet grows, IPv6 will become more prevalent. The current technological direction of typical networks will become the next goals for IPSec; specifically, Quality of Service (QoS). ATM was practically invented to accommodate the vast array of communication technologies at high speeds; but to do it efficiently, it must control who gets in and out of the network.

Ethernet Type of Service (ToS) (802.1p) allows for three bits of data in the frame to be used to add ToS information and then be mapped into ATM cells. IP version 4, as currently applied, has support for a ToS field in the IP Header similar to Ethernet 802.1p; it provides three bits for extended information. Currently, techniques are being applied to map QoS information from one medium to another. This is very exciting for service organizations that will be able sell end-to-end QoS. As the IPSec standard grows and current TCP/IP applications and networks begin to support the existing IP ToS field, IPSec will quickly conform to the requirements.

The IETF and other participants, in the form of RFCs, are continually addressing the issues that currently exist with IPSec. Packet sizes are typically increased due to the added headers and sometimes trailer information associated with IPSec. The result is an increased possibility of packet fragmentation. IPSec addresses fragmentation and packet loss; the overhead of these processes constitutes the largest concern.

IPSec can only be applied to the TCP/IP protocol. Therefore, multi-protocol networks and environments that employ IPX/SPX, NetBEUI, and others will not take direct advantage of the IPSec VPN. To allow non-TCP/IP protocols to communicate over an IPSec VPN, an IP gateway must be implemented to encapsulate the original protocol into an IP packet and then be forwarded to the IPSec gateway. IP gateways have been in use for some time and are proven technology. For several organizations that cannot eliminate non-TCP/IP protocols and wish to implement IPSec as the VPN of choice, a protocol gateway is imminent.

As is obvious, performance is crucial to IPSec VPN capabilities and cost. As encryption algorithms become increasingly sophisticated and hardware support for those algorithms becomes readily available, this current limitation will be surpassed.

Another perceived limitation of IPSec is the export and import restrictions of encryption. There are countries that the United States places restrictions on to hinder the ability of those countries to encrypt possibly harmful information into the United States. In 1996, the International Traffic in Arms Regulation (ITAR) governing the export of cryptography was reconditioned. Responsibility for
cryptography exports was transferred to the Department of Commerce from the Department of State. However, the Department of Justice is now part of the export review process. In addition, the National Security Agency (NSA) remains the final arbiter of whether to grant encryption products export licenses.

The NSA staff is assigned to the Commerce Department and many other federal agencies that deal with encryption policy and standards. This includes the State Department, Justice Department, National Institute for Standards and Technology (NIST), and the Federal Communications Commission. As one can imagine, the laws governing the export of encryption are complicated and are under constant revision. Several countries are completely denied access to encrypted communications to the United States; other countries have limitations due to government relationships and political posture. The current list of (as of this writing) embargoed countries include:

- Syria
- Iran
- Iraq
- North Korea
- Libya
- Cuba
- Sudan
- Serbia

As one reads the list of countries, it is easy to see why the United States is reluctant to allow encrypted communications with these countries. Past wars, conflict of interests, and terrorism are the primary ingredients to become exiled by the United States.

Similar rosters exist for other countries that have the United States listed as “unfriendly,” due to their perception of communication with the United States.

As one can certainly see, the concept of encryption export and import laws is vague, complex, and constantly in litigation. In the event a VPN is required for international communication, it will be necessary to obtain the latest information available to properly implement the communication as per the current laws.

150.23 Conclusion

VPN technology, based on IPSec, will become more prevalent in our every-day existence. The technology is in its infancy; the standards and support for them are growing every day. Security engineers will see an interesting change in how security is implemented and maintained on a daily basis. It will generate new types of policies and firewall solutions—router support for VPN will skyrocket.

This technology will finally confront encryption export and import laws, forcing the hand of many countries. Currently, there are several issues with export and import restrictions that affect how organizations deploy VPN technology. As VPNs become more prevalent in international communications, governments will be forced to expedite the process. With organizations sharing information, services, and product, the global economy will force computer security to become the primary focus for many companies.

For VPNs, latency is the center for concern and, once hardware solutions and algorithms collaborate to enhance overall system performance, the technology will become truly accepted. Once this point is reached, every packet on every network will be encrypted. Browsers, e-mail clients, and the like will have VPN software embedded, and only authenticated communications will be allowed. Clear Internet traffic will be material for campfire stories. It is a good time to be in security.