NetworkCare
The knowledge behind the network

Lucent Technologies
Bell Labs Innovations



IPSec Virtual Private Networks: A Technical Review

By Jim Tiller, Network Security Consultant, CISSP

Introduction

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 (VPN). A secure VPN is the combination of tunneling, encryption, authentication, access control, and auditing technologies and services used to transport traffic over the Internet or any insecure network that uses the TCP/IP protocol suite for communication. Several other standards and protocols can provide tunneling of data, but, not all of them implement encryption and authentication. The term VPN is a widely encompassing acronym that represents many communication standards. For the purposes of this paper, all references are made to secure VPNs and the IPSec's implementation to provide security in the form of encryption and authentication.

In 1994, the Internet Architecture Board (IAB) issued a report on "Security in the Internet Architecture" (Request For Comment [RFC] 1636). The report expressed 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 continue to be compatible with future generations of IP and network communication technology.

VPN started in 1995 with the AIAG (Automotive Industry Action Group), a non-profit 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, its growth and acceptance had been stymied by the lack of standards and by product support issues.

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

- Client to Gateway
- Gateway to Gateway
- Client to Client

This process of broad communication support is only possible through detailed standards. IPSec (IP Security Protocol) is an ever growing standard for providing encrypted communications over IP. Its acceptance and robustness has fortified the 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.

The goals of this paper are to:

- Introduce the IPSec standard and the RFCs that make up the VPN technology.
- Introduce the protocols of the IPSec suite and Key management.
- Provide a technical explanation of the IPSec communication technology.

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, and the interrelationship and organization of the documents are necessary for understanding the development process of the overall standard.

As Figure 1 shows, seven groups of documents allow the separate aspects of the IPSec protocol suite to be developed independently while a functioning relationship is attained and managed.

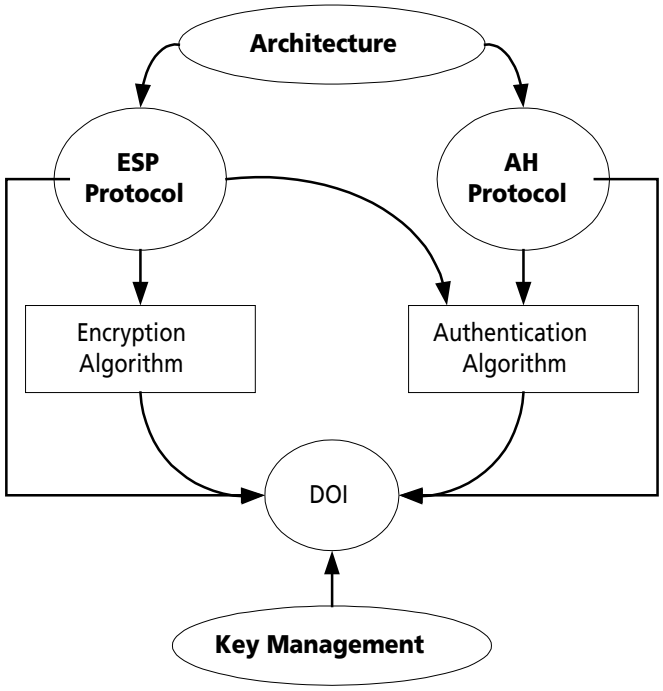

Figure 1

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

The Authentication Algorithms documents describe the processes and technologies used to provide an authentication mechanism for the AH and ESP Protocols. Examples include HMAC-MD5 (RFC 2403) and HMAC-SHA-1 (RFC 2404).

All of these documents specify values which 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, allowing other documents to relate to each other. The DOI contains parameters that are required for 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 ISAKMP and PKI.

This paper will unveil each of these protocols and the underlying technologies which make them the standard of choice in VPNs.

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:

- Authentication Header (AH)
- 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 for 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. 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 is 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 authentication only, 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. A 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 may commence.

Understanding the Foundation

Security Associations are the infrastructure of IPSec. Of all the portions of the 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.

Two SAs are required for authenticated, confidential, bi-directional communications between systems. Each SA can be defined by three components:

- Security Parameter Index (SPI)
- Destination IP Address
- Security Protocol Identifier (AH or ESP)

An SPI is a 32-bit value used to distinguish 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 an 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.

Finding the Gateway

Prior to any communication, a map must be constructed and shared among the community of VPN devices. This provides information about 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.

Below is 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 using a shared secret with key management:

```
target "10.1.0.0/255.255.0.0"
```

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 to properly match the secret with the destination gateway. A sample of this is a Diffie-Hellman key, explained in detail later in this paper.

Modes of Communication

The type of operation for IPSec connectivity is directly related to the role the system plays in the VPN or the SA status. As shown in Figure 2, there are two modes of operation for IPSec VPNs:

- Transport mode
- Tunnel mode

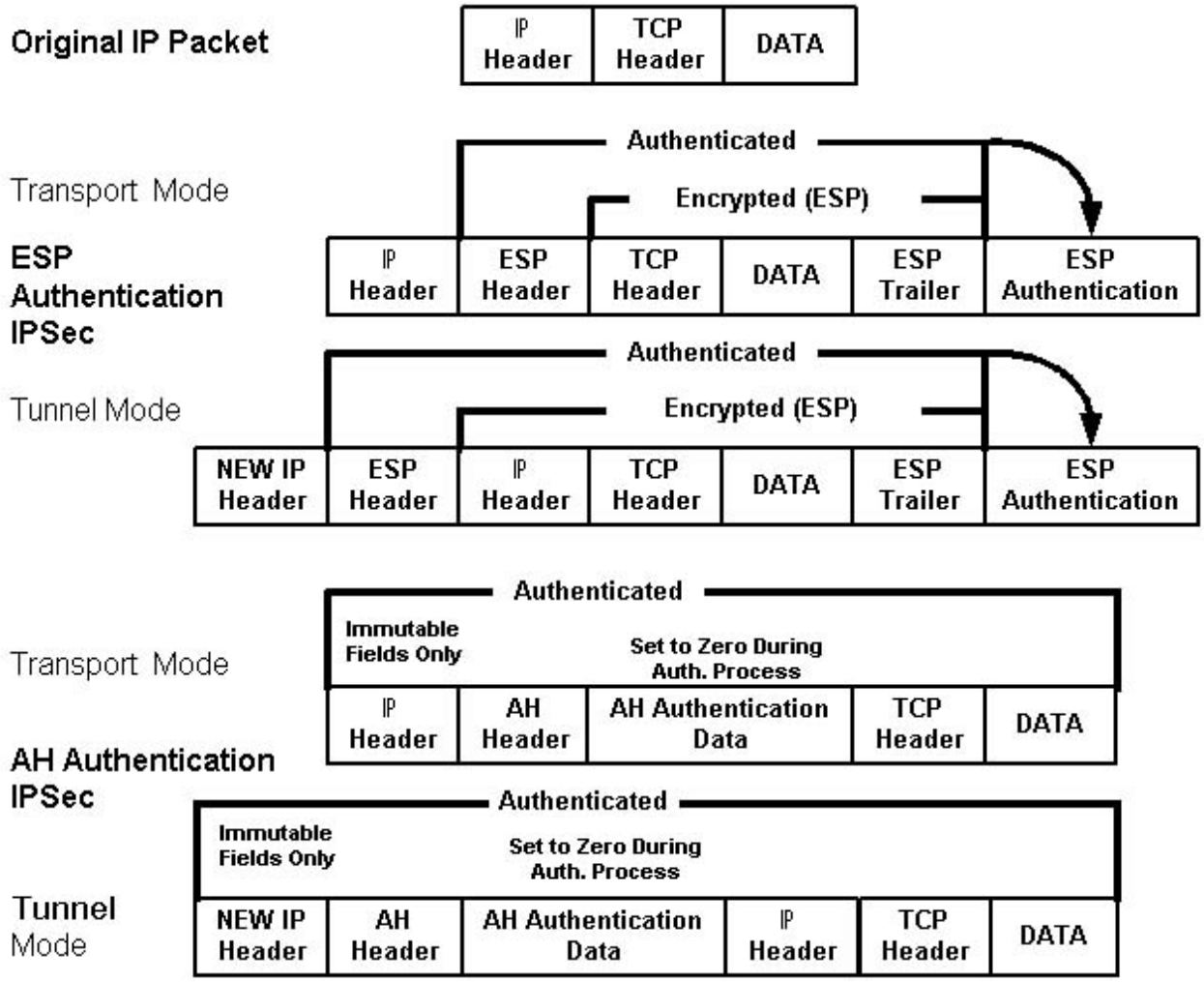

Figure 2

Transport mode is used to protect upper layer protocols and only effects 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 a gateway to host communications were 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.

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.

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:

- Outside
- Inside

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 Time To Live (TTL) 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 Type of Service (ToS),which is currently available in IPv4.

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.

Authentication and Integrity

Security protocols provide packet authentication and integrity 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, the message was not modified. If the message hash does not match, the data has been altered from the time it was hashed.

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 advances in small strong encryption algorithms such as ECC (Elliptic Curve Cryptosystems), the increase of 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.

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 of controlling, managing, and tracking SAs.

Various applications may require the unification of services demanding combined SAs to accomplish the required transport. An example is 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, and it can provide a layered effect of communications. SA bundles can be utilized by applications in two formats:

- Fine Granularity
- Coarse Granularity

Fine granularity is the assignment of SAs for each communication process. Data transmitted over a single SA are protected by a single security protocol. The data is protected by an AH or ESP, but not both, since 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 allows the communication two levels of protection by way of more than one SA. Consider the example of a host on the Internet that establishes 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:

- Transport adjacency
- Iterated tunneling

Transport adjacency is 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, since 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 Figure 3: Example A)

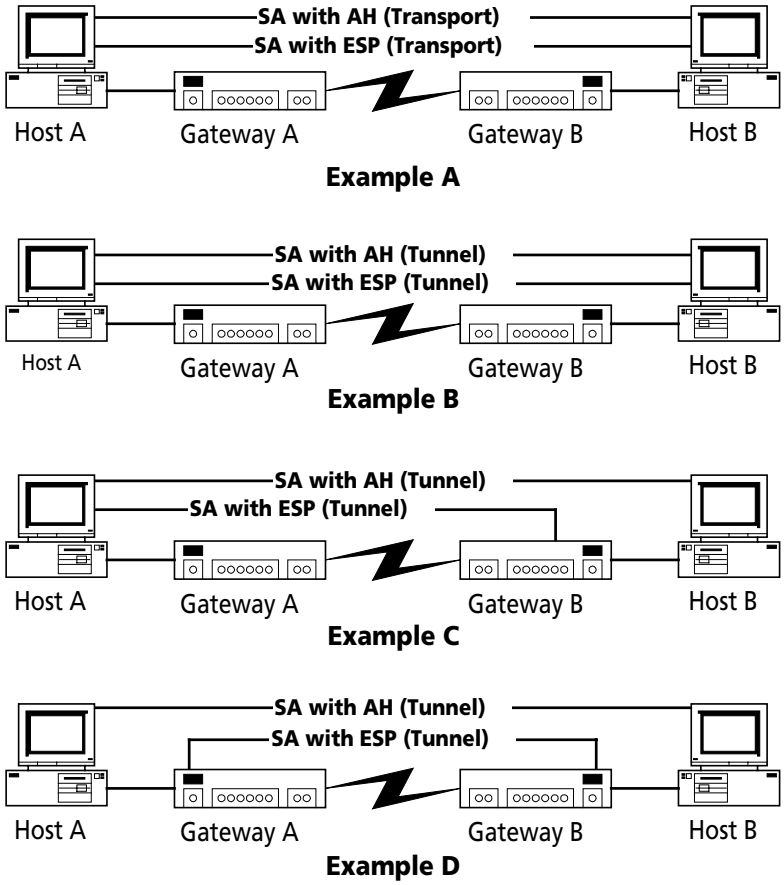

Figure 3

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 since 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 and/or ESP is contained in an ESP providing the tunnel. (See Figure 3: 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 Figure 3: Example C)

In the event that neither SA terminates 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 secured communication based on client to client SAs. This provides for several layers of authentication and data protection. (See Figure 3: Example D)

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 Figure 4, 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.

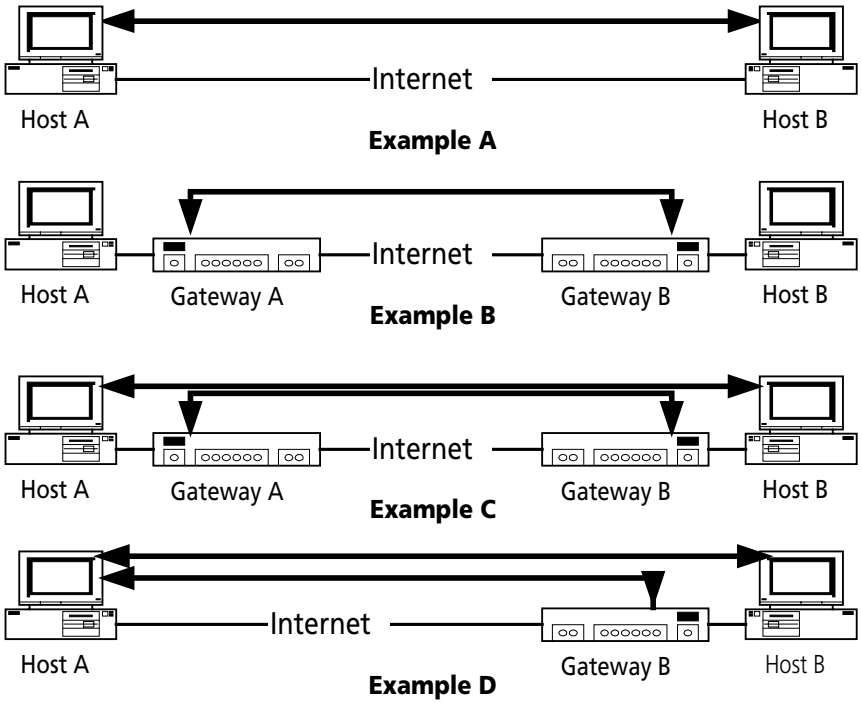

Figure 4

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, 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.

Keeping Track

Security associations and the varieties 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 must be maintained by any system participating in a IPSec VPN:

- Security Policy Database (SPD)
- 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 is a minimum of two of each database - one for tracking inbound and another for outbound communications.

Communication Policies

The SPD is a security association management constructed 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 offered to IP datagrams and in what fashion they are implemented. SPD is consulted for all IP and IPSec communications, inbound and outbound, and is therefore 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 both traffic that is IPSec-based and traffic that is not IPSec related. There are three modes of this operation:

- Forward and do not apply IPSec
- Discard packet
- 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 an SA or SA bundle to be employed. Within the specification is 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 that 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 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:

- Sequence number counter for outbound communications
- Sequence number overflow counter that sets an option flag to prevent further communications utilizing the specific SA
- A 32 bit anti-replay window that is used to identify the packet for that point in time traversing the SA and which provides the means to identify that packet for future reference
- Lifetime of the SA that is determined by a byte count or time frame, or a combination of the two
- The algorithm used in the AH
- The algorithm used in the authenticating ESP
- The algorithm used in the encryption of the ESP
- IPSec mode of operation, Transport or Tunnel mode
- 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.

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.

Key History

Key management has a far from obvious definition and interchanged acronyms only add to the 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. These formats provide a consistent framework, which 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:

- ISAKMP
- Oakley
- 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 term IKE and ISAKMP are used interchangeably with 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, enterprise-wide network security architecture that may be utilized by IPSec.

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.

Phases and Modes

Phase one takes place when the two ISAKMP peers establish a secure, authenticated communication channel. 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
- 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
- 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-phased approach is that the second phase can provide additional ISAs, which eliminates the re-authorization 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 acknowledgements. 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 because it provides keying material for the initiation of SAs for the security protocol. This is the point where the 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.

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:

- Shared secret
- Certificate
- Public/Private key

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.

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.

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
- Asymmetrical

Symmetrical key encryption occurs when the same key is used for the encryption of information into human unintelligible data, or cipher text, and for the decryption of that cipher text 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 are mathematically related and generated by a complicated formula. The concept of asymmetry 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.

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 can be a phone call, e-mail, fax, or some form of non-related 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 from Berkeley and Martin E. Hellman from Stanford, California, defined the Diffie-Hellman agreement protocol (also known as exponential key agreement) and published it in a paper titled, "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 Adleman, who developed the RSA Public and Private key cryptosystem based on large prime numbers, developed the Diffie-Hellman formula (the nuts and bolts of the protocol). This allows communication of a symmetrical key without transmitting the actual key, but rather a mathematical portion or fingerprint.

An example of this process is when 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 $g^k = n \bmod p$.

Both of these numbers must be hard-coded or retrieved from a remote system. Each system then generates a number called "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 pass phrase that can be combined with date and time to create a unique number. The hard-coded numbers will not be exceeded since most, if not all, applications employ a limit on the input.

A new key is generated with these numbers, $g^X \bmod p$. The resulting '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 (US4200770) and is such a strong protocol there are currently 128 other patents that reference Diffie-Hellman.

Many Keys

Asymmetrical keys, such as PGP (Pretty Good Privacy) and RSA, can be used to share the keying information. Asymmetrical keys were designed specifically 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.

Key Establishment

IPSec standard mandates that key management must support two forms of key establishment:

- Manual
- 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, the key management can also be related to the ultimate security of the communication. If the key is compromised the communication can be exposed to 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.

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 number 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 5 sites throughout North America. This organization wants to use the Internet for communications, and each office must be able to communicate directly with any other office. If each VPN relationship had a unique key, the number of keys can be calculated by the formula $n(n\text{-}1)/2$ where n is the number of sites. In this example the number of keys is 10. Apply this formula to 25 sites, 5 times the number of sites in the previous example, and the number of keys skyrockets to 300, not 50. In reality, management is more difficult than it may appear in the examples. Each device must be configured and the keys must be shared

with all corresponding systems. The use of manual keying reduces the flexibility and number of options of IPSec. Anti-replay, on-demand re-keying, and session specific key management are not available in manual key creation.

Automatic Keying

Automatic key management addresses the limited manual process and provides for widespread, automated deployment of keys. The goal of the IPSec is to build on exiting 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 may be employed. Automated key management, when instituted, may create several keys for a single SA. There are various reasons for this:

- 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 multiple keys, or if both algorithms are used, 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), each using more than one key. (DES-EEE2 uses two keys, one of which is used twice.)

The process of re-keying protects 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 may provide two primary methods of key provisioning:

- Multiple string
- Single string

Multiple strings are passed to the corresponding system in the SA for each key and for each type. For instance, the use of triple DES for the ESP will require more than one key to be generated for a single type algorithm, in this case the encryption algorithm. The recipient will receive a string of data representing a single key, and 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 may 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, with the highest bit order listed first for the encryption key(s) and the remaining string used for 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.

Future of IPSec VPNs

Like it or not, IPSec VPNs are 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 and specifically, Quality of Service (QoS), will become the next goals for IPSec. 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 mapped into ATM cells. IP version 4, 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.

Conclusion

VPN technology, based on IPSec, will become more prevalent in our everyday existence. The technology is in its infancy and the standards and support for it 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. The advent of hardware VPN solutions will become common place.

Currently, support for a wide range of operating systems is still lacking. The true growth of IPSec VPNs will be realized as the standard is absorbed into operating systems and implemented directly into the TCP/IP protocol stack. Even though the hardware solutions are the current drivers of VPN technology, their expense, combined with the complexity of the smaller access community, will increase the standard's integration into the operating system level. Hardware solutions provide significant improvements for enterprise-wide solutions but become cumbersome for use in such applications as remote access and extranets, where budgets tend to be more scrutinized.

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 effect how organizations deploy VPN technology. As VPNs become more prevalent in international communications, governments will be forced to expedite the processes.

With organizations sharing information, services, and products, global economy will force computer security to become the primary focus for many companies.

For VPN, latency is for a central 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, operating systems, and the like will have VPN capabilities embedded and only authenticated communications will be allowed.

For further information, see NetworkCare's website at http://www.lucent-networkcare.com or call 1-888-767-2988 in the U.S. or 1-727-217-2303 outside the U.S.

Copyright © 2000, Lucent Technologies NetworkCare

This is an unpublished work protected under the copyright laws. All rights reserved.
WP.EN.NS.IPSEC.0100