Seguridad en servicios de voz sobre IP y mensajería instantánea

PID_00191701

Antoni Martínez Ballesté

Universitat Oberta de Catalunya

Índice

Introducción

Internet, aparte de ser una fuente inmensa de búsqueda y compartición de información, es un entorno que permite la comunicación entre usuarios. Tanto es así, que el correo electrónico ha acabado siendo una forma habitual de comunicación y envío de información entre personas y ha relegado el correo postal a propósitos más concretos o que no tienen implementación posible sobre un sistema telemático. Además, la tecnología IP (y por extensión Internet) se ha usado para implantar aplicaciones de comunicación síncrona o en tiempo real. Desde las primeras aplicaciones de chat o mensajería instantánea, hasta la integración de la videoconferencia en las redes sociales y en los teléfonos móviles, han ido surgiendo aplicaciones de todo tipo que enriquecen las posibilidades de comunicación entre personas y abaratan costes destinados a la remisión de información.

Si bien estos sistemas basados en IP e Internet han supuesto muchas mejoras en la productividad o en la comunicación interpersonal, lo cierto es que también representan un foco de ataques informáticos. Hay que tener presente que, por medio de la mensajería instantánea o voz sobre IP, se transmite información que puede ser considerada privada y sensible: por ejemplo, se pueden cerrar acuerdos comerciales, se puede proporcionar información sobre un proyecto futuro, se puede comprar interactuando con sistemas automáticos de reconocimiento de voz, etc. Así pues, es conveniente comprender cuáles son los problemas de seguridad a los cuales estas tecnologías se exponen, y también tener en cuenta qué medidas y tecnologías se pueden utilizar para mitigar los efectos de ataques potenciales.

Más allá de proteger los sistemas y las redes que sustentan estos sistemas de comunicación, será importante conocer cómo proteger la información que circula por ellos, y también velar por la correcta autenticación de las partes que intervienen en las comunicaciones.

Este módulo describe los problemas de seguridad informática que pueden presentar los sistemas de comunicaciones síncronas, en concreto la voz sobre IP y la mensajería instantánea. También recoge la serie de técnicas que conviene aplicar para proteger los sistemas y la información que circula por ellos.

Para la comprensión de los conceptos de este módulo, es conveniente que tengáis conocimientos fundamentales de las redes de computadores, concretamente, de todo lo que rodea la comunicación con TCP/IP. También es importante haber estudiado con anterioridad las técnicas criptográficas y los productos de seguridad ampliamente usados en la transmisión segura de la información.

Objetivos

Los objetivos que habréis alcanzado una vez finalizado el estudio del módulo son los siguientes:

1. Comprender cómo funcionan los servicios de voz sobre IP y de mensajería instantánea, centrándose en los elementos de la arquitectura y los protocolos que los hacen posibles.

2. Conocer cómo afectan los ataques sobre los equipos y las redes al buen funcionamiento de estos servicios.

3. Evaluar el impacto de los ataques sobre la confidencialidad de la información que pueden experimentar estos sistemas.

4. Evaluar el impacto de los ataques sobre la autenticación de la información y los participantes que pueden experimentar estos sistemas.

5. Aplicar herramientas de seguridad a los protocolos y componentes de la arquitectura de estos servicios para evitar problemas de seguridad o minimizar su impacto en caso de que se produzcan.

1. Servicios de comunicación síncrona

En este apartado introducimos dos entornos, cuyos problemas de seguridad analizaremos: la voz sobre IP y la mensajería instantánea. Cabe decir que otros sistemas de comunicación síncrona, como por ejemplo la videoconferencia, pueden experimentar ataques similares, y se pueden usar las técnicas descritas en este módulo para mitigarlos. Así pues, con la intención de no extender demasiado el módulo con conceptos repetitivos, solo nos centramos en estos dos entornos.

Cronológicamente hablando, la primera aplicación de comunicación síncrona es la mensajería instantánea. El origen se podría remontar a la instrucción *talk* disponible en los sistemas Unix, con la cual un usuario del sistema podía enviar mensajes cortos de texto a otro usuario del mismo sistema de manera inmediata. Este servicio ha ido evolucionando y dando decenas de aplicaciones diferentes.

La mensajería instantánea (IM) consiste en la comunicación entre dos o más usuarios mediante mensajes cortos de texto que se envían en tiempo real.

IM

Las siglas IM, que denominan la mensajería instantánea, vienen del inglés *instant messaging*.

La voz sobre IP (VoIP) permite aprovechar la infraestructura de red de datos para el establecimiento de conversaciones telefónicas, tanto punto a punto como en grupo. Esta tecnología permite el ahorro en recursos TIC, puesto que no hay una red telefónica paralela que se tenga que mantener. En la misma línea, para unas nuevas instalaciones, solo hay que pensar en un único cableado de datos que también servirá para la voz.

La voz sobre IP (VoIP) es una tecnología que permite el establecimiento de llamadas telefónicas usando los datagramas IP como medio de transporte.

VoIP

La abreviatura VoIP, que denomina la voz sobre IP, viene del inglés *voice over the Internet protocol*.

En el fondo, hay diferentes variantes que se pueden englobar dentro de la VoIP. Por un lado, cuando las llamadas se pueden establecer hacia la red telefónica convencional, se denomina telefonía IP. Por otro lado, si la llamada se transmite más allá de la red interna, es decir, se transmite por Internet, hablamos de voz sobre Internet o telefonía sobre Internet. De este modo, el uso de IP para el tránsito de voz dentro de una red cerrada y controlada (por ejemplo,

Cisco y Skype

Como ejemplo, la empresa Cisco desarrolla soluciones de VoIP para grandes corporaciones. Por otro lado, la aplicación Skype se ha convertido en un estándar en la telefonía y la videoconferencia sobre Internet (incluso su tecnología es utilizada por otros proveedores de servicios de Internet).

una LAN) se correspondería puramente con la VoIP. De todos modos, se utiliza VoIP para referirse a todas estas variantes, y además, los protocolos empleados no son diferentes.

La integración de estas tecnologías en la web ha permitido que la mensajería instantánea, la voz sobre Internet y la videoconferencia se puedan gestionar desde un navegador sin la necesidad de instalar software específico. Este sería el caso, por ejemplo, de las herramientas de Google y Facebook para comunicaciones síncronas.

Finalmente, es conocido que algunas herramientas de IM han incorporado tecnologías de voz sobre Internet (como es el caso de Messenger) o, por el contrario, algunas aplicaciones de voz sobre Internet también incorporan IM. También cabe añadir la posibilidad de la videoconferencia en muchos de estos sistemas populares, como por ejemplo el mencionado Skype.

En este módulo estudiamos la seguridad relacionada con la VoIP y la IM y, para hacerlo, empezaremos en primer lugar por ver cómo funcionan estas dos tecnologías.

1.1. Funcionamiento de la voz sobre IP

En este subapartado introduciremos la tecnología sobre la cual se desarrolla la VoIP. De este modo, podremos entender los problemas de seguridad y comprender cómo se pueden evitar o solucionar. En primer lugar, describiremos la arquitectura de un sistema VoIP en una red, en dispositivos y en servidores. Después introduciremos los protocolos que se usan en esta tecnología: los propuestos por el UIT y los de la IETF.

UIT

UIT es la Unión Internacional de Telecomunicaciones, en inglés International Telecomunications Union. Es la responsable de definir muchos de los protocolos usados en telefonía fija y móvil, sistemas de vídeo conferencia, cableados, etc.

1.1.1. Arquitectura

IETF

La Internet Engineering Task Force (IETF) es uno de los organismos más influyentes en cuanto a estándares de Internet. Los estándares se publican en unos documentos llamados RFC (*request for comments).*

Un sistema VoIP funciona sobre una red IP. En esta red se establece la comunicación entre dos o más agentes de usuario por medio del envío de paquetes de señalización, los cuales describen quién es el destinatario, especifican cuál es la localización (dirección IP) del originador de la llamada, envían señales de llamada al agente de usuario del destinatario, etc. Una vez se ha establecido la comunicación, se inicia un envío de paquetes de voz para transmitir la conversación entre los participantes. En general, la señalización se puede establecer sobre un protocolo de transporte orientado a conexiones (es decir, TCP), mientras que el transporte de voz acontece de usuario a usuario por medio de datagramas y UDP.

Lo más habitual hoy en día es que el sistema funcione sobre una red de área local (LAN) o bien con una comunicación entre dos o más usuarios conectados a Internet. Así pues, para poner un ejemplo de LAN, supondremos un entorno de oficinas. En las instalaciones hay un conmutador Fast Ethernet que

centraliza las conexiones de la red de datos formada por varios ordenadores que usan IP conectados a este conmutador. También suponemos que hay un encaminador que conecta la LAN a Internet.

Los elementos que forman parte de la arquitectura VoIP son los teléfonos IP, las centralitas IP y las pasarelas de telefonía. Adicionalmente, puede haber otros servidores complementarios.

Los teléfonos IP son los nodos desde los cuales a menudo se inician o se reciben las comunicaciones (es decir, agentes de usuario). Hay diferentes variantes de estos teléfonos: por un lado, hay teléfonos parecidos a los de la telefonía convencional pero que realmente usan VoIP; de la otra, hay teléfonos de software que están instalados en los ordenadores y, por lo tanto, necesitan un micrófono y una salida de audio para poder hacer efectivas las llamadas. Adicionalmente, se pueden usar adaptadores a VoIP enchufados a la línea de teléfonos convencionales, o bien teléfonos USB que, en el fondo, son dispositivos que interactúan con un software instalado en el ordenador. Cada teléfono IP tiene un número de extensión que lo identifica y que se usa a la hora de especificar el destinatario de la llamada. Algunos sistemas de VoIP permiten especificar el destinatario de la llamada usando su nombre de usuario, como si se tratara de una dirección de correo electrónico.

Teléfonos de software

Los teléfonos de software reciben, en inglés, el nombre de *softphones*.

TCP

Recordad que TCP (*transmission control protocol*) establece una sesión entre los comunicantes, de modo que el protocolo se encarga de que todos los datagramas lleguen correctamente y se entreguen a la aplicación en el mismo orden en que la aplicación origen los ha emitido.

Figura 1. A la izquierda, teléfono IP; a la derecha, un teléfono IP de software

En los teléfonos IP tiene lugar una de las etapas más importantes de la VoIP: la digitalización y compresión o descompresión de la voz. Para llevar a cabo estas tareas, se pueden usar diferentes codificadores, en función de los requisitos de anchura de banda, calidad de voz y retardos máximos. En este sentido, un codificador que mantenga una buena calidad de voz para poco ancho de banda, suele implicar la introducción de un retraso en la comunicación. Como ya hemos apuntado, la voz comprimida se transmite en paquetes IP que van sobre UDP, puesto que el uso de TCP (que garantiza la entrega en orden de todos los paquetes emitidos) introduciría retrasos que podrían llegar a degradar en exceso la calidad de servicio de la llamada.

UDP

Recordad que UDP (*user datagram protocol*) es el envío de datagramas de manera independiente, sin establecer una conexión entre los comunicantes.

La centralita de VoIP, o IP-PBX[1], es un ordenador con un software que hace las funciones de centralita telefónica. Las centralitas son un elemento clave a la hora de establecer y gestionar las llamadas. Por cuestiones de disponibilidad y eficiencia, puede haber varias centralitas actuando de manera coordinada. Otra opción es que la centralita sea un dispositivo externamente similar a un conmutador o encaminador, más que un ordenador donde se ejecuta un software.

(1)Abreviatura en inglés de *IP private branch exchange*.

Voz IP con muchos usuarios

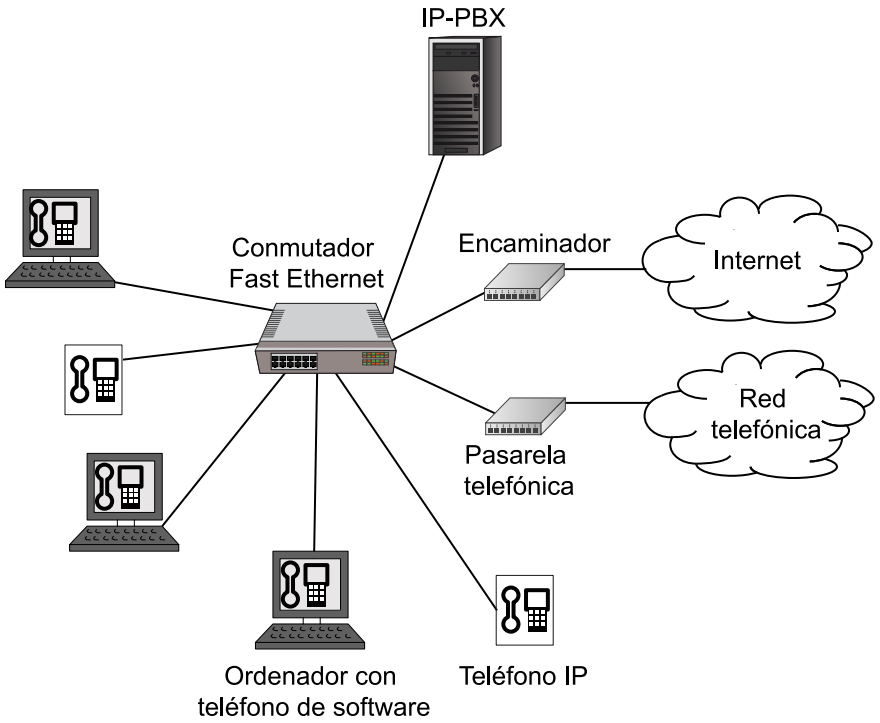
Un caso ilustrativo extremo de varias centralitas actuando de manera coordinada sería el de un sistema de voz sobre Internet con centenares de miles de usuarios, el cual requeriría centenares de servidores distribuidos geográficamente.

Finalmente, la pasarela telefónica o *gateway* es un dispositivo que conecta la red de datos por donde circula el tránsito de VoIP con la red telefónica convencional, ya sea a través de líneas analógicas o a través de líneas RDSI[2].

(2)RDSI significa red digital de servicios integrados.

Además de los elementos anteriores, en un sistema de VoIP suele haber otros servidores. Por ejemplo, servidores con funcionalidad de buzón de voz, servidores de contabilidad y facturación, servidores de directorio y control de agentes de usuario (*gatekeepers*), servidores de configuración, etc. A pesar de no ser obligatorios, suelen ser el foco de algunos ataques. La figura 2 muestra un escenario de ejemplo con todos estos elementos.

Figura 2. Escenario de ejemplo de una red con VoIP



Una vez definidos los elementos que forman una red de VoIP, introduciremos las tendencias principales por lo que respecta a protocolos que permiten el establecimiento y el control de llamadas.

1.1.2. Protocolos UIT

En primer lugar, introducimos los protocolos definidos por la UIT como estándares para la VoIP, aprovechando algunos de los estándares que se usan ampliamente en la telefonía convencional.

La UIT define, dentro del estándar H.323, un conjunto de protocolos para el establecimiento y el control de llamadas a través de la VoIP.

Por ejemplo, dentro del protocolo H.323 se definen los protocolos siguientes en relación con la señalización y el control de la llamada:

- H.225, que define el uso del protocolo Q.931 en el establecimiento de llamadas telefónicas en la RDSI, enviando estos paquetes Q.391 sobre conexiones TCP en lugar de hacerlo sobre el canal de control del RDSI. También incluye las especificaciones para el registro y la admisión de agentes de usuario en las llamadas (RAS, *registration, admission and status*), usando UDP.

- H.245, que define mensajes para concretar termas relacionados con la compresión de la voz, los puertos que hay que emplear, las capacidades de los teléfonos IP, etc., sobre la conexión TCP que ha iniciado la H.225/Q.931.

El protocolo H.323 también especifica que, para el transporte multimedia de la voz, hay que utilizar una capa RTP (protocolo definido por la IETF) para insertar en los paquetes de voz información de secuencia y de tiempo real, la cual es esencial para garantizar una buena comunicación. Como veremos más adelante, alterar esta información es uno de los ataques más frecuentes en la VoIP. Adicionalmente, la UIT define el protocolo H.248 para la conversión entre VoIP y la red telefónica convencional (tendría relación, pues, con las pasarelas telefónicas).

RTP

RTP significa *real-time transport protocol*. Está definido en el RFC 3550.

En el protocolo H.323 los paquetes se codifican en forma binaria con la técnica PER[3], para transmitirlos eficientemente en las redes de comunicaciones.

[3]PER significa *packet encoding rules*.

A pesar de que la UIT se considera el referente en cuanto a estándares para telefonía, lo cierto es que la tendencia la marcan otras alternativas. Por un lado, algunos de los grandes fabricantes de tecnología para VoIP (por ejemplo, Cisco) emplean protocolos propios, de modo que hay que utilizar pasarelas entre arquitecturas H.323 y sus sistemas. Por otro lado, muchas de las aplicaciones de VoIP basadas en teléfonos de software (como por ejemplo Skype, o Google Voice), se han decantado por el uso del protocolo SIP, que introducimos a continuación. Los protocolos H.323 se recogen en la tabla 1.

Tabla 1. Alternativa H.323 para VoIP

Control y establecimiento de llamada		Registro y admisión	Control del estado de la red	Voz comprimida, G.711, G.729,...
H.245	H.225/Q.931	H.225/RAS	RTCP	RTP
TCP			UDP	
IP				

1.1.3. Protocolo IETF

El protocolo SIP, definido por la IETF en las RFC 2543 y 3261, ha acabado por ser el estándar *de facto* en aplicaciones de VoIP y también en IM.

El protocolo SIP se diseñó para poder establecer sesiones multimedia sobre una red IP con el objetivo de que sirviera para un espectro de aplicaciones amplio.

SIP

SIP significa protocolo de iniciación de sesiones (en inglés, *session initiation protocol*). Está definido en las RFC 2543 y 3261.

A diferencia del protocolo H.323, SIP usa mensajes de texto en las peticiones y respuestas (señalización y control de llamadas), de una manera similar a lo que hacen FTP y HTTP. Los identificadores de usuario de los sistemas VoIP usan la arroba, igual que los identificadores (o direcciones) de correo electrónico.

Las funcionalidades de SIP son muy diversas, y para implementarlas se usan diferentes peticiones o respuestas. Las funcionalidades que recoge SIP son las siguientes:

- Ubicación de usuarios, que permite conocer la localización de un agente de usuario o de un usuario dentro de una red de VoIP.
- Disponibilidad de usuarios, que permite saber si el usuario se encuentra disponible.
- Capacidades de usuario, para poder saber qué parámetros conviene usar para la llamada (ancho de banda, codificadores aceptados, etc.).
- Establecimiento de sesión, para establecer la llamada en el caso de VoIP, o la conversación en caso de IM.
- Gestión de la sesión, para modificar los parámetros durante la conexión, añadir usuarios, acabar la llamada, etc.

Para implementar estas funcionalidades se suelen usar protocolos auxiliares, como por ejemplo el RTP, el RTSP[4], el MGCP[5] y el SDP[6]. El primero, tal como hemos visto para el caso del protocolo H.323, añade información en tiempo real a los paquetes de información. El RTSP sirve para controlar el envío de información multimedia (por ejemplo, para detener el envío de vídeo en una

(4)RTSP significa *real-time streaming protocol*. Lo define el RFC 2326.

conferencia de VoIP). El MGCP es el protocolo encargado de conectar un sistema SIP con la red telefónica convencional por medio de la pasarela, tal como lo hace el protocolo H.248. Finalmente, el SDP permite describir la información necesaria para iniciar una sesión (por ejemplo, la localización del destinatario, la dirección de quien inicia la conversación, puertos que se usarán, etc.).

(5)MGCP significa *media gateway control protocol*. Lo define el RFC 3435.

(6)SDP significa *session description protocol*. Lo define el RFC 4566.

En un escenario basado en SIP, la centralita (u otros servidores complementarios) pueden hacer las funcionalidades de registro de usuarios, intermediarios de llamada, etc., lo que proporciona un amplio abanico de posibilidades a la hora de implantar soluciones centralizadas, distribuidas, con posibilidades de balanceo de carga, etc.

Los mensajes más habituales en una sesión SIP son el REGISTER y el INVITE:

- Cuando un usuario se valida en el sistema, manda un mensaje REGISTER al servidor de registros (en inglés, *registrar server*).

- Cuando un usuario quiere iniciar una conversación, manda un mensaje INVITE a uno de los servidores intermediarios de llamada.

A continuación, mostramos un ejemplo de mensaje INVITE, donde podemos apreciar que se usan identificadores similares a los del correo electrónico (tipo URI[7]), tanto para los usuarios como para los identificadores de llamada.

(7)URI significa *uniform resource identifier*.

```
INVITE sip:toni@voip.uoc.edu SIP/2.0
Via: SIP/2.0/UDP toni.intranet.uoc.edu;branch=z9hG4bK776asdhds
Max-Forwards: 70
To: Jordi <sip:jordi@voip.uoc.edu>
From: Toni <sip:toni@voip.uoc.edu >;tag=1928301774
Call-ID: a84b4c76e66710@toni.intranet.uoc.edu
CSeq: 314159 INVITE
Contact: <sip:toni@toni.intranet.uoc.edu >
Content-Type: application/sdp
Content-Length: 142
...
```

Los protocolos relacionados con el sistema SIP se recogen en la tabla 2.

Tabla 2. Alternativa SIP para VoIP

Control de flujos multimedia	**Control y establecimiento de llamada (SDP)**	**Control del estado de la red**	**Voz comprimida, G.711, G.729,...**
RTSP	SIP	RTCP	RTP
TCP	UDP o TCP	UDP	
IP			

1.2. Funcionamiento de la mensajería instantánea

Las aplicaciones de IM permiten conversaciones entre usuarios usando mensajes cortos de texto que se envían en tiempo real. En general, las conversaciones se pueden establecer potencialmente entre un grupo de contactos, al cual se pueden ir añadiendo usuarios: un usuario se registra en el servicio IM y después va añadiendo (o invitando) a otros usuarios del servicio, que pasarán a formar parte de su lista de contactos.

Desde los inicios de la popularización de Internet (a mediados de los noventa) han ido surgiendo decenas de aplicaciones de IM. Cada una de las aplicaciones que ha ido apareciendo ha ido definiendo los protocolos y las implementaciones de modo que, en general, la interoperabilidad entre diferentes aplicaciones era prácticamente inviable. Aun así, algunas aplicaciones se han popularizado mucho en los últimos años y han llegado a ser herramientas cotidianas para muchos usuarios.

Como hemos comentado, estas herramientas incorporan voz sobre Internet o posibilidad de transmitir vídeo, tanto desde aplicaciones específicas como desde el propio navegador web. Estos sistemas también permiten enviar ficheros (documentos, ejecutables, fotografías, etc.) o bien enlaces web.

1.2.1. Arquitectura

La arquitectura básica que definimos está compuesta por agentes de usuario (software para establecer conversaciones y enviar mensajes) y por diferentes servidores. Para el caso de los servidores, en nuestra arquitectura definimos tres tipos:

- **Servidor de despacho**. Este es el servidor con el cual contacta un agente de usuario cuando se conecta al sistema.

- **Servidor de notificaciones**. Este servidor contiene información sobre los usuarios que están activos, es decir, en disposición de efectuar o recibir llamadas.

- **Servidor de centralita**. Este servidor centraliza las comunicaciones entre dos o más participantes, recogiendo los mensajes que generan y enviándolos hacia los otros participantes.

Se entiende que si el sistema tiene un número elevado de usuarios, habrá varios servidores que actuarán de manera coordinada, como en el supuesto que hemos comentado en la VoIP.

1.2.2. Funcionamiento de una sesión

Por lo que se refiere a los protocolos, abundan los protocolos propietario, aunque también es habitual encontrar sistemas basados en SIP. Esto implica que la IM puede sufrir problemas similares a los que sufre la VoIP, claramente en el supuesto de que ambas implementaciones utilicen SIP. Consiguientemente, muchas de las soluciones a los problemas de seguridad serán comunes a VoIP y a IM.

Ejemplo

En nuestro escenario, suponemos un sistema basado en mensajes de texto tipo SIP para el establecimiento de llamadas y otras gestiones. Estas operaciones requerirán el envío de paquetes de señalización, donde se describirá cuál es el contacto destinatario, cuál es su IP, etc.

Una sesión se da mediante una serie de pasos, que se ilustran en la figura 3. Supongamos que el usuario 1 quiere iniciar una conversación con uno de sus contactos, el usuario 7:

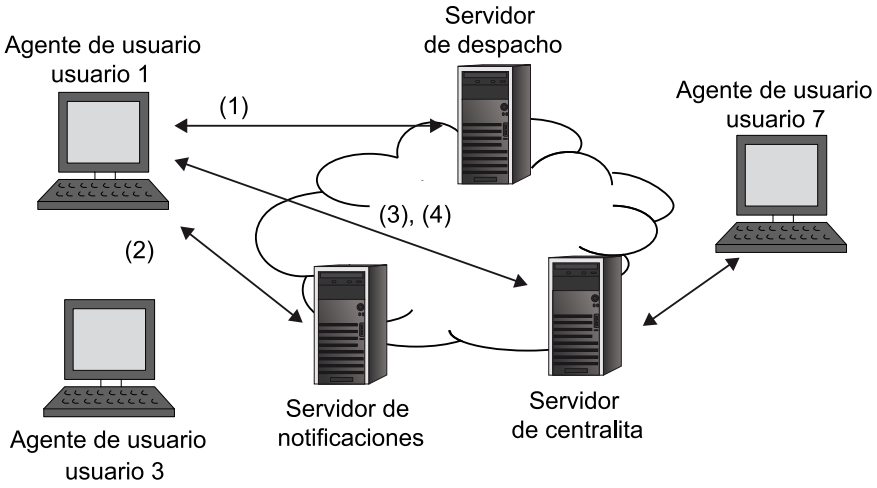
1) En primer lugar, el usuario 1 se valida contra el servidor de despacho usando un mensaje REGISTER. Si la validación es correcta, el servidor de despacho envía al agente de usuario del usuario 1 una clave y una dirección IP donde está el servidor de notificaciones que tiene que usar.

2) El agente de usuario se conecta al servidor de notificaciones correspondiente y, a cambio de la clave que le ha proporcionado el servidor de despacho y su identificador, obtiene la lista de contactos {usuario 2, ..., usuario 20} y cuáles de estos contactos están activos. También se le asigna la IP del servidor de centralita. El servidor de notificaciones tiene que avisar a los agentes de usuario de los contactos que están activos que el usuario 1 se acaba de incorporar.

3) Cuando el usuario 1 quiere establecer una conexión con el usuario 7, el agente de usuario del primero manda un mensaje INVITE hacia el servidor de centralita. El mensaje contiene el identificador del destinatario (usuario 7), con lo cual el servidor de centralita contactará con el agente de usuario del usuario 7 para establecer la conversación.

4) Se generan mensajes y se transmiten en paquetes UDP.

Figura 3. Ejemplo de establecimiento de sesión en IM



En el supuesto que presentamos aquí, todos los mensajes de texto que se envían los participantes pasan por el servidor de centralita. Aun así, algunos sistemas de mensajería instantánea apuestan porque las comunicaciones entre usuarios se hagan directamente entre sus agentes de usuario, es decir, sin centralizarse en un servidor.

En general, de manera similar a lo que sucede en VoIP, las conexiones y el control de sesiones se harán mediante una conexión TCP, mientras que el contenido (texto) se podrá enviar por medio de UDP. De hecho, hay tantos escenarios como implementaciones posibles, y más si tenemos en cuenta que las aplicaciones de mensajería instantánea también ofrecen funcionalidades de envío de archivos, por ejemplo. Esto último puede comportar el establecimiento de conexiones TCP para el envío de datos.

Una vez hemos visto la arquitectura y los protocolos en los que se basan dos de los servicios de comunicaciones síncronas más utilizados, es el momento de estudiar los problemas de seguridad que presentan.

2. Denegación y degradación del servicio

Los sistemas presentados funcionan con protocolos de transmisión de datos que son ejecutados por ordenadores. Algunos de los sistemas efectúan las funcionalidades de servidor, con todo lo que esto implica. Por estos motivos, los servicios de VoIP y de IM son susceptibles de recibir ataques similares a los que puede recibir cualquier otro sistema de transmisión de información por medio de redes de computadores, como por ejemplo, el servicio web, el correo electrónico, etc.

Para resolver los problemas de seguridad, o intentar evitarlos, se utilizan técnicas que ya se usan en los sistemas de transmisión de información. Así pues, encontraremos ataques relacionados con los equipos (que afectan a la disponibilidad) y ataques relacionados con la seguridad de la información (que afectan a la confidencialidad y la integridad de la información). Dentro de los ataques relacionados con los equipos, hay ataques específicos contra los servidores y ataques específicos contra los agentes de usuario (es decir, teléfonos VoIP y software de IM). Finalmente, daremos un vistazo a las recomendaciones de fabricantes y organismos para dar seguridad a los equipos

2.1. Ataques contra los servidores

Los atacantes pueden colapsar los servidores que toman parte en un servicio de VoIP o de IM con el objetivo de denegar el servicio a los usuarios autorizados. El efecto de estos ataques va desde una pérdida en la calidad de servicio (comunicaciones que tardan más de lo habitual en establecerse, retrasos y cortes en las conversaciones, etc.) hasta la imposibilidad total de usar el servicio.

Uno de los métodos habituales de conseguir una denegación de servicio es por medio de la inundación de peticiones hacia los servidores. En general, estos ataques se suelen originar desde miles de máquinas que inician el ataque a petición del atacante.

Ahora bien, como en todos los servidores, una de las maneras de conseguir que deje de prestar el servicio es por medio de la intrusión en el sistema. El atacante consigue entrar en el sistema, ya sea explotando vulnerabilidades o bien usando ingeniería social (por ejemplo, llamando al administrador y pidiéndole las credenciales de autenticación). A continuación, el atacante toma el control de la máquina y los procesos para los que presta servicio.

También es posible bloquear el software que reside en los servidores por medio del envío de peticiones mal formadas, que colgarán el servidor. Este método forma parte del aprovechamiento de vulnerabilidades del software.

Para mitigar estos posibles problemas, es evidente que hay que aplicar todas las políticas de protección de sistemas que aplicaríamos a cualquiera otro servidor. En este sentido, apuntamos las acciones siguientes:

- **Actualización del sistema operativo y del software en ejecución**. Los sistemas operativos presentan vulnerabilidades y, a medida que se van descubriendo, los fabricantes van proporcionando "parches" para corregirlos. Lo mismo suele suceder con los diferentes programas que puede haber instalados en el servidor. Por lo tanto, habrá que tomar conciencia de la importancia de tener los sistemas actualizados.

- **Protección de los servidores con cortafuegos**. Un cortafuego es un software que gestiona qué tránsito puede salir hacia la red o bien entrar hacia el equipo u otra red. Los cortafuegos que filtran paquetes, por medio de la inspección de los puertos o direcciones de red, son más que suficientes para controlar los accesos al servidor, o bien en la red donde se ubique el servidor. Como en un sistema moderadamente grande de VoIP habrá varios servidores relacionados con el servicio VoIP para dar cobertura a una organización, conviene que estos estén ubicados dentro de una zona desmilitarizada[8]. Aun así, el uso de cortafuegos para proteger servidores de VoIP puede comportar algunos problemas que hay que tener en cuenta, y que detallamos más adelante.

(8)El concepto de zona desmilitarizada (DMZ, del inglés *demilitarized zone*) se estudia en el módulo "Sistema de cortafuegos".

- **Instalación de un sistema detector de intrusiones**. Los sistemas detectores de intrusiones (IDS[9]) analizan la actividad en una máquina o red para poder detectar comportamientos anómalos por parte de usuarios y procesos, con el objetivo de notificar que hay una intrusión al sistema. Un IDS es el complemento idóneo del cortafuegos para garantizar al máximo que, si hay alguna intrusión que ha hecho saltar el sistema de cortafuegos, podrá ser detectado.

(9)IDS viene del inglés *intrusion detection system*. Este concepto se ve en el módulo "Sistemas de detección de intrusos en red".

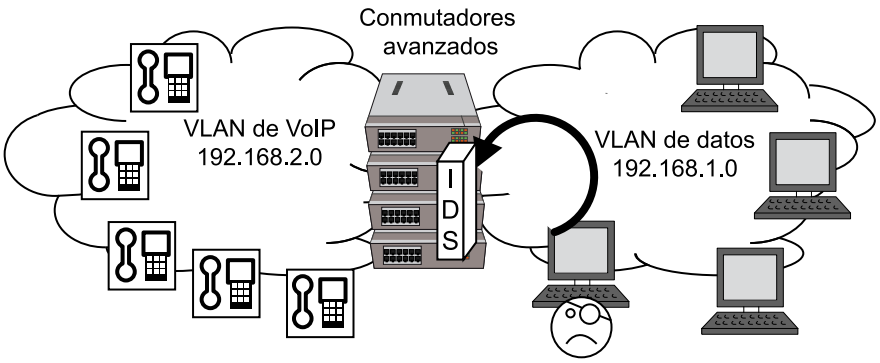
- **Autenticación fuerte de usuarios administradores**. La mayoría de software y dispositivos actuales permiten la gestión desde interfaces web donde el usuario administrador debe validarse. Fortalecer esta validación por medio de certificados digitales ayuda a evitar que atacantes entren en el sistema y rompan los sistemas de autenticación por contraseña.

Para el caso de los servidores de VoIP dentro de una organización, los fabricantes recomiendan, si es posible, usar tecnología VLAN para poder definir un segmento de red exclusivo para la VoIP. De este modo, el tránsito de la VoIP iría por unos puertos en concreto y un atacante conectado a la VLAN de datos donde se conectan los equipos tendría más complicado acceder al segmento reservado a la voz. Adicionalmente, se recomienda que en el segmento de VLAN dedicado a voz resida un IDS, incluso, si es posible, implementado dentro mismo del conmutador para detectar posibles intentos de intrusión. De hecho, los conmutadores avanzados con tecnología VLAN suelen tener este tipo de características, entre las cuales destacamos la detección de intrusiones o la protección hacia algunos ataques de denegación de servicio. La figura 4 ilustra el uso de VLAN y un IDS en un sistema de VoIP.

VLAN

La tecnología VLAN permite organizar máquinas conectadas a un conmutador como si estuvieran conectadas en diferentes redes, independientemente de su conexión física a la red.

Figura 4. Uso de VLAN en un sistema de VoIP



A pesar de que las propuestas anteriores maximizan el fracaso de los ataques más habituales contra los servidores, la variedad de sistemas y aplicaciones es tan amplia que siempre hay algún nuevo ataque que podría tener éxito.

2.2. Ataques contra los teléfonos IP

Los teléfonos IP también pueden recibir ataques de denegación (o degradación) de servicio, en el sentido de que pueden ser inundados con peticiones SIP, o bien con paquetes de voz que interrumpan constantemente el buen flujo de una conversación telefónica. Al fin y al cabo, se trata de nodos accesibles dentro de una red IP. De estos ataques específicos encontramos dos variantes:

- **Envío de paquetes de fin de llamada**. El atacante envía a los agentes de usuario que estén en una llamada un paquete de fin de llamada que sea conforme al protocolo usado en el sistema.

- **Envío de paquetes con información RTP falsa**. El protocolo RTP proporciona control del tiempo real en un flujo multimedia por medio de la utilización de números de secuencia (para que el receptor pueda determinar el orden de envío correcto hacia el descompresor), y el marcaje de paquetes con información del instante de tiempo (se conoce como *timestamping*).

Claramente, si un atacante envía paquetes de voz con información de RTP falsa, el descompresor producirá un audio lleno de cortes, errores y ruido.

En este caso, si se producen los ataques puede ser por dos motivos. Uno de ellos, que el atacante haya entrado en el servicio de VoIP y lo esté atacando "desde dentro". Se entiende que el atacante puede entrar si las propuestas del apartado anterior para evitar ataques a los servidores no han funcionado. El otro motivo podría ser que un usuario tuviera el equipo infectado por un código malicioso y este fuera el verdadero responsable de ir enviando ataques para degradar la calidad del servicio de otros usuarios del servicio de VoIP.

Los fabricantes principales de tecnología para VoIP recogen en las documentaciones una serie de recomendaciones para dotar de seguridad el servicio de los teléfonos IP. A continuación las reproducimos:

- **Recomendaciones sobre el hardware**. Algunos fabricantes de tecnología VoIP se decantan por emplear exclusivamente teléfonos IP en lugar de teléfonos de software. Los primeros son menos susceptibles de recibir ataques e infecciones por parte de software malicioso. Además, es más probable que un usuario inexperto se descargue cualquier teléfono de software que, en realidad, aproveche para piratear el sistema de VoIP, ya sea para enviar conversaciones a un tercero, o para atacar los servidores de VoIP desde dentro de la red.

- **Recomendaciones sobre el direccionamiento IP**. La asignación de direcciones IP a los teléfonos tendría que ser manual, para minimizar el riesgo de sufrir ataques. Ahora bien, si la asignación solo puede ser automática por la dimensión del sistema de VoIP, las IP solo se tendrían que asignar a direcciones físicas[10] conocidas. Como las direcciones físicas se pueden falsear, los fabricantes recomiendan que el usuario ponga al teléfono una contraseña para que el aparato obtenga una dirección IP. Se recomienda usar exclusivamente direcciones privadas[11] para los teléfonos IP. Una dirección privada corresponde a una serie de rangos reservados con este objetivo, y no se pueden asignar a interfaces de red conectadas a la "parte pública" de las redes. Finalmente, los servidores con información de configuración, directorio, etc. tan solo tendrían que dar información a los teléfonos IP que pertenezcan a una lista controlada por el administrador del servicio de VoIP.

(10)Las direcciones físicas se conocen como direcciones MAC (*medium access control*).

(11)Un ejemplo de dirección privada es 192.168.0.7.

2.3. Otros ataques contra voz sobre IP

Los sistemas de VoIP pueden ser el foco de ataques específicos relacionados con las características propias del servicio. A continuación presentamos algunos:

- **Manipulación de la configuración de los agentes de usuario**. En algunas implementaciones, los teléfonos IP utilizan servicios de transferencia de ficheros contra un servidor de configuración para determinar parámetros como permisos de usuario, obtener entradas de directorios telefónicos, o incluso permitir la actualización de software. Si un atacante tiene acceso a estos servidores de configuración, o bien es capaz de suplantar el servidor original, los teléfonos IP se configurarían erróneamente para que el atacante pudiera hacer efectivos varios ataques.

- **Manipulación de los registros de contabilidad**. Si un atacante puede tener acceso a los servidores de contabilidad y facturación, podrá modificar los registros correspondientes a su usuario para abaratar considerablemente el coste que pueda pagar por el servicio. Con la misma facilidad, podría alterar las cuentas de otros usuarios e incrementarles la factura añadiendo servicios que realmente no se han prestado.

- **Manipulación del equipo de usuario**. En determinados casos, cuando los usuarios usan un teléfono basado en software, un programa malicioso podría entrar en el ordenador del usuario con finalidades de espía. Así como existe software espía que recoge información sobre contraseñas y otros datos que pueda teclear el usuario para enviarlas a un atacante, se puede dar el caso de un software que envíe las conversaciones de VoIP hacia un atacante.

- **Envío de llamadas publicitarias no solicitadas**. Del mismo modo que quienes envían correos publicitarios masivos usan servidores SMTP víctima como plataforma, un atacante podría entrar en un servidor de un servicio de VoIP y usar sus recursos para efectuar llamadas publicitarias. Estas llamadas degradarían la calidad del servicio (por la disminución de recursos disponibles), a la vez que causarían una molestia evidente a los usuarios del sistema y receptores eventuales de la publicidad masiva. Quienes envían correos masivos publicitarios a menudo utilizan ordenadores personales infectados con software malicioso con el objetivo de que envíe también mensajes de este tipo. Así pues, se podría pensar en atacar teléfonos IP para efectuar llamadas publicitarias en nombre de un usuario de un servicio de VoIP.

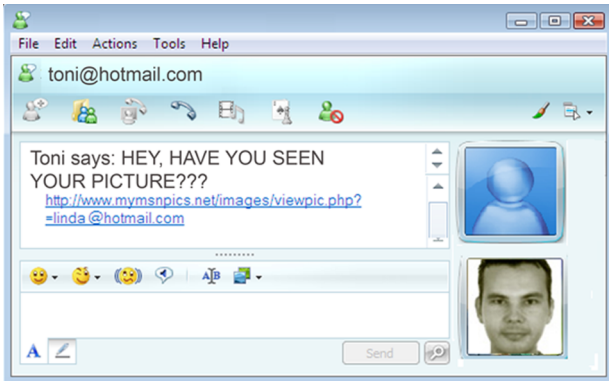
Los dos últimos ataques también tienen su equivalente en la IM. La solución o prevención de los problemas anteriores de seguridad dependerá de cada uno de los escenarios concretos. Aun así, en gran medida, la protección de redes, servidores y agentes de usuario en un servicio tendrían que minimizar las posibilidades de que los ataques anteriores tengan éxito.

2.4. Ataques contra el software de mensajería instantánea

El software de IM también puede ser víctima de los ataques contra los teléfonos IP. En este sentido, los atacantes pueden enviar mensajes de fin de llamada o bien mensajes de texto falsos o que contengan software malicioso. Hemos comentado que una de las utilidades añadidas a los servicios de IM es la transferencia de ficheros. Adicionalmente, también se pueden enviar enlaces web para poder compartir información con los participantes de una conversación. Pues bien, aprovechando este hecho se puede transferir software malicioso mediante servicios de IM.

Más concretamente, para hacer más efectivo este ataque conviene que previamente se haya tenido éxito en un ataque de suplantación de identidad. Si un contacto abre una conversación y nos pide que descarguemos un fichero, en principio no deberíamos pensar que se trata de un software malicioso. Ahora bien, si ha habido una suplantación de identidad y nuestro contacto utiliza un idioma que no es el habitual, habrá motivos para desconfiar. En la figura 5 se muestra un envío que, previsiblemente, ha efectuado un atacante. Lo detectamos porque la lengua que utiliza el contacto es el inglés, que no es la habitual. Haciendo clic en el enlace mostrado, se descargaría algún tipo de software malicioso.

Figura 5. Ejemplo de un envío por parte de un atacante



Además de la distribución de software malicioso por medio de la suplantación de identidad, algunos atacantes utilizan las deficiencias de programación de las aplicaciones de usuario. En estos casos, la distribución del software malicioso a gran escala puede ser cuestión de minutos: los contactos que se tienen con el servicio de IM forman de hecho una telaraña de conexiones donde miles de usuarios están indirectamente conectados. El programa malicioso se esparce hacia todos los contactos de los usuarios que estén conectados.

Para minimizar el impacto de estos ataques, la única solución es desconectar de la red los usuarios con más contactos, o como solución más drástica, cerrar temporalmente los servidores.

Ahora bien, si lo que se quiere es prevenir que "robots" envíen software malicioso, se propone que el agente de usuario necesite una resolución de captcha para poder enviar ficheros y enlaces.

Captchas

Un captcha es un test que se usa para comprobar que el usuario es humano. La palabra viene de las siglas en inglés *completely automated public Turing test to tell computers and humans apart*. En general, se trata de teclear las palabras que aparecen en una imagen deformada.

Para concluir este apartado, vamos a mencionar que las aplicaciones de IM presentan cada vez menos problemas de seguridad. Hoy en día son pocas las aplicaciones que han llegado a acaparar grandes cuotas de mercado. Gracias al hecho de que son aplicaciones ampliamente conocidas y utilizadas, constantemente son un banco de pruebas de miles de usuarios *hacker*. Así pues, si encuentran problemas de seguridad, se avisará a la comunidad de usuarios o a los fabricantes del software para remediarlo.

3. Problemas de seguridad en la comunicación

Una vez hemos estudiado los aspectos de seguridad relacionados con servidores y agentes de usuario, analizaremos los problemas relacionados con la seguridad de la comunicación. En concreto, veremos la confidencialidad en las comunicaciones, la integridad de lo que se transmite y la autenticación de los participantes.

3.1. Confidencialidad de la comunicación

Los servicios de VoIP e IM a menudo sirven para organizar reuniones a distancia entre diferentes participantes. En estas conversaciones se pueden tratar temas confidenciales.

Del mismo modo que los sistemas de telefonía tradicional pueden ser "pinchados" para enviar la conversación a un atacante (en este caso espía), las conversaciones que circulan por sistemas de VoIP e IM pueden ser interceptadas.

Evidentemente, las LAN actuales, basadas en conmutadores (*switches*) y no en concentradores (*hubs*), intentan evitar por sí solas que usuarios no destinatarios escuchen conversaciones. Aun así, hay que recordar la efectividad de ataques basados en el protocolo ARP en la interceptación de paquetes de datos: del mismo modo que es posible que todos los datagramas IP sean enviados hacia el espía y después reenviados hacia el destinatario, es perfectamente viable que un atacante reciba paquetes que contengan la voz o los mensajes de texto.

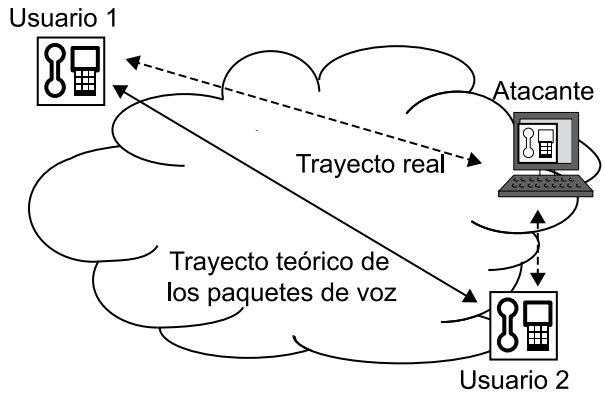
ARP

ARP significa *address resolution protocol*, y es el protocolo encargado de traducir direcciones IP a direcciones físicas de dispositivo de red.

En el caso de la voz, tan solo habrá que descomprimir la información y escuchar la conversación (recordad que los codificadores de voz suelen ser estándar y por lo tanto, su interpretación y descompresión es conocida). Este ataque, para el caso de la VoIP, se muestra en la figura 6.

Para garantizar la confidencialidad, habrá que aplicar algún sistema de cifrado a la comunicación. En el caso de la VoIP, hay que tener en mente que el sistema empleado no ha de introducir un retraso significativo que, añadido al retraso mismo de la compresión de voz y la transmisión de paquetes, degrade la calidad de la conversación.

Figura 6. Escucha de una conversación de VoIP



Los servicios de IM también presentan problemas de seguridad relacionados con la confidencialidad y la integridad de la información.

En la VoIP, los ataques relacionados con la seguridad de la información tienen unas características específicas relacionadas con el hecho de que la información que se transmite es voz digitalizada. En cambio, la transmisión de texto hace menos complejos la captura y el procesamiento de paquetes y, en consecuencia, el éxito de los ataques.

Más concretamente, es relativamente sencillo para un atacante capturar los paquetes de manera automática, procesarlos para seguir la conversación y filtrar los paquetes que puedan contener información sensible. Por ejemplo, es más sencillo filtrar todos los paquetes que contengan el texto *PIN* o contraseña, que aplicar un sistema de reconocimiento de voz para seleccionar aquellos cortes de una conversación de VoIP que pueden resultar interesantes para un atacante.

Por último, aunque no por ello menos importante, hay que tener en cuenta que todos los paquetes que se transmiten en entornos SIP, como por ejemplo los paquetes de descripción de llamada del protocolo SDP, son paquetes en claro. Espiando estos paquetes se puede obtener información diversa con la cual un atacante podría fácilmente hacerse pasar por el destinatario inicial y verdadero de la llamada.

3.2. Integridad de la comunicación

La información que se transmite por un sistema de VoIP es, en esencia, voz humana digitalizada y codificada. La modificación de paquetes no tendría sentido, puesto que la información que se transmite no es textual, sino un audio.

Aun así, bajo el escenario de ataques basados en ARP, se podría plantear que el atacante no solamente escucha la conversación, sino que la modifica. Un ataque interesante sería cambiar la frecuencia y la modulación de la voz, pa-

ra "feminizar" una voz masculina o al contrario, o bien insertar silencios en el flujo de voz. Este ataque tendría incidencia, en el fondo, en la calidad de la llamada. Parece más interesante que el atacante pueda sustituir un *sí* por un *no* con varias finalidades, como por ejemplo, cambiar la intención de toda una conversación. Aun así, este sería un escenario poco probable dada su complejidad.

También es muy sencillo cambiar el contenido de los paquetes de texto en IM. Un atacante puede cambiar, o incluso insertar, texto puesto que en principio la aplicación no está tan ligada a medidas máximas y especificidades de los codificadores de compresión como lo está la VoIP. Solo hay que pensar en la complejidad técnica que representaría sustituir una palabra *sí* por un *no*, tal como hemos sugerido antes para la VoIP.

Hará falta, pues, asegurar que los teléfonos IP solo procesen paquetes de voz que no se han modificado. Para garantizar la integridad se pueden usar códigos de autenticación de mensaje en cada paquete, teniendo en cuenta para la VoIP los mismos requisitos en cuanto a retrasos que se han señalado en el caso de garantizar la confidencialidad.

3.3. Autenticación de los participantes

Los participantes que toman parte en una conversación de IM o VoIP tienen que estar autenticados convenientemente, es decir, hay que comprobar su identidad. Un fallo en la autenticación de los comunicantes podría resultar en una suplantación de identidad. Por ejemplo, en un escenario de VoIP el atacante podría hacer uso de palabras y expresiones capturadas durante una conversación del usuario suplantado para generar nuevas conversaciones. Este tipo de ataque se podría usar con un servicio automático de contratación de servicios por teléfono.

Interesa más, sin embargo, describir un ataque basado en el envenenamiento de los sistemas de localización IP de usuarios: la redirección de llamadas o conversaciones. Supongamos que un atacante envenena el directorio que almacena la relación entre usuarios y la dirección correspondiente en un sistema VoIP. Entonces será posible que un usuario víctima llame a un teléfono determinado y, en realidad, acabe llamando al atacante. Por ejemplo, un usuario llama al teléfono de la entidad bancaria y el sistema de VoIP lo remite al teléfono de un atacante. Este escenario parece más sencillo de aplicar en llamadas telefónicas que en los casos de "pesca" (*phishing*), que emulan webs que suplantan la identidad (en general, de entidades bancarias).

Un punto clave en la autenticación de los comunicantes es el momento en el que los usuarios se validan en el sistema. Si el sistema solo emplea un sistema de nombre de usuario y contraseña, podrá haber problemas derivados de la ingeniería social. Además, los sistemas de identificación autenticados por contraseña pueden adolecer de los problemas inherentes al uso de contraseñas.

Por ejemplo, el uso de ataques de diccionario o ataques de fuerza bruta para conseguir encontrar contraseñas consideradas débiles. Aun así, a pesar de tener una contraseña robusta, los llamados ataques de ingeniería social pueden lograr con éxito el objetivo de conseguir una contraseña, por robusta que sea.

Ataques de ingeniería social

No se trata solo de llamar a un usuario y pedirle la contraseña de IM o VoIP, haciendo creer que somos el proveedor de servicio, sino que se puede ir mucho más allá. Por ejemplo, en algunas páginas web se podría encontrar algún anuncio o pequeño juego en línea (*online*) que nos hiciera creer que para acceder al mismo deberemos hacerlo por medio del servicio de IM. La web en cuestión presentaría un cuadro de diálogo con una interfaz similar a la del agente de usuario de IM, donde la víctima pondría el nombre de usuario y la contraseña. Enseguida esta información iría hacia los servidores del atacante, que podría usar las credenciales "robadas" para entrar en el servicio de IM con el objetivo de atacar.

Claramente, el uso de autenticación de los participantes por medio de certificados digitales tendría que garantizar que se podrá comprobar la autenticidad de los participantes en la comunicación y, por lo tanto, tendría que hacer más difícil la suplantación de usuarios. Aun así, también es importante la cultura de la precaución a la hora de navegar por Internet y facilitar datos confidenciales.

4. Herramientas para comunicaciones seguras

Una vez hemos visto cuáles son los problemas de seguridad relacionados con las comunicaciones, analizaremos las técnicas de seguridad que se pueden aplicar. Con este objetivo, veremos protocolos concretos que se pueden usar.

Primeramente, tratamos cómo proteger la señalización y la gestión de la llamada por medio de la protección con técnicas y protocolos criptográficos de los paquetes que contienen la información de establecimiento de llamada, gestión de los participantes, etc. En segundo lugar, describiremos los sistemas estándar que permiten garantizar la seguridad en el transporte de la voz.

Para asegurar sistemas basados en SIP, la IETF propuso el uso de protocolos ya existentes. Por lo que respecta a las recomendaciones UIT para VoIP, son similares, pero no se entra en tanto detalle. Este organismo recoge las recomendaciones de seguridad para la H.323 dentro del protocolo H.235.

4.1. Seguridad en la señalización

Para la señalización (establecimiento y gestión de llamadas, básicamente), conviene garantizar la autenticación en el origen de los paquetes, y también que la información no se modifica (es decir, la integridad de los paquetes). Adicionalmente, y como SIP utiliza mensajes en formato de texto y en consecuencia no demasiado difíciles de espiar, también se asegurará la confidencialidad de la información que se transmita durante la señalización.

4.1.1. Sistemas SIP

La IETF propone utilizar, para los sistemas de VoIP, IM y videoconferencia, el formado S/MIME para la autenticación de la información correspondiente a SDP. Aunque se trata de un formato inicialmente pensado para el correo electrónico seguro, lo cierto es que permite la remisión de firmas digitales, cifrados y la gestión de los certificados necesarios.

S/MIME

S/MIME significa *secure multipurpose Internet mail extensions*, y se encuentran especificados en las RFC 3850 y 3851.

Ahora bien, hay que tener presente que la información correspondiente, por ejemplo, al destinatario, no puede estar cifrada: los diferentes servidores por donde se irá encaminando la llamada o las posibles pasarelas por las que circule, tienen que disponer de esta información. Los mensajes S/MIME contienen una firma digital de esta información, de modo que el destinatario puede comprobar, cuando menos, que el mensaje SDP no se ha modificado durante el trayecto. La figura 7 muestra un ejemplo de uso de S/MIME para autenticar la información SDP que contiene un paquete SIP.

Figura 7. Estructura de un paquete SIP que usa S/MIME para autenticar la información SDP

```
INVITE sip:jordi@uoc.edu SIP/2.0
Via: SIP/2.0/udp ...
From: <toni@urv.cat> ...
...
...
Content-Type: multipart/signed;boundary=e4ef8847482d240d0
Accept: application/sdp, multipart/mixed
Content-Length: 3381
--e4ef8847482d240d0
Content-Type: application/pkcs7-mime
smime-type=envelopeddata; name=smime.p7m
Content-Disposition: attachment;handling=required;filename=smime.p7m
Content-Transfer-Encoding: binary
*** envelopedData object containing encrypted SDP body ***
* v=0
* o=- 0 0 IN IP6 2001:db8::27:2
* m=audio 49170 RTP/AVP 112 113
...
**********************************************************
--e4ef8847482d240d0
Content-Type: application/pkcs7-signature;name=smime.p7s
Content-Disposition: attachment;handling=required;filename=smime.p7s
Content-Transfer-Encoding: binary
```

Mensaje SIP

Disposición en claro SDP

Firma SDP

Además de asegurar cierta información por medio del uso de S/MIME, para los sistemas SIP se recomienda usar TLS[12] para enviar paquetes de señalización de manera segura. Este protocolo permite la autenticación de los participantes en una comunicación, a la vez que asegura la integridad y la confidencialidad de los datos que circulan. Aun así, es necesaria la participación de una infraestructura de clave pública[13]. Cualquier protocolo de VoIP que utilice TCP como canal de transporte en el establecimiento y el control de la llamada puede usar de manera sencilla el TLS para asegurar la comunicación.

(12)TLS significa *transport layer security*. Está basado en SSL (*secure socket layers*) y se encuentra definido en el RFC 2246.

(13)En inglés, *public key infraestructure*, PKI.

Ved también

TLS y SSL se describen en el módulo "Seguridad en redes WLAN" de esta asignatura.

TLS proporciona una capa de transporte seguro usando un conjunto amplio de algoritmos y parámetros posibles, los cuales se definen por medio de un protocolo de negociación entre los participantes[14]. Justo es decir que TLS proporciona seguridad basándose, en el fondo, en la confianza que se tenga hacia quien genera el certificado de los comunicantes. Si un comunicante quiere establecer una llamada VoIP con otro, conviene que el primero confíe en el emisor del certificado del segundo participante.

(14)También llamado *handshake* o apretón de manos.

4.1.2. Recomendaciones de la UIT

El estándar H.235 especifica un conjunto de mecanismos de seguridad de aplicación en un entorno de VoIP basado en protocolos del H.323. Estas recomendaciones son en cierto modo similares a las que se especifican para los servicios basados en SIP. En cuanto a los protocolos relacionados con el establecimiento y la gestión de las llamadas, la UIT recomienda lo siguiente:

- Los mensajes del protocolo H.225/RAS pueden transportar información sobre qué protocolos y parámetros de seguridad se usarán en el resto de protocolos implicados.

- El protocolo de control de llamada H.245 se puede proteger por medio del uso de TLS. Así pues, los participantes en una comunicación se pueden autenticar en la ejecución propia del TLS. Dentro de este protocolo de control también se puede especificar con qué algoritmos y parámetros se asegurarán los paquetes de voz.

- Los mensajes del protocolo H.225/Q.931 se pueden proteger usando TLS.

El protocolo H.235 define una serie de perfiles de seguridad, cada uno de los cuales regula o bien diferentes aspectos de la protección de VoIP (por ejemplo, autenticación en el protocolo RAS usando claves compartidas, intercambio de claves para el cifrado de la voz, etc.) o bien diferentes niveles de protección.

Uno de los niveles más usados es el protocolo H.235.1 (*baseline security profile*), que garantiza autenticación e integridad, pero no confidencialidad. En concreto, protege de los ataques de tipo "hombre en medio" (*man-in-the-middle*) y de "secuestro de sesión" (*session hijacking*), entre otros.

4.2. Seguridad en el envío de la conversación

Por lo que respecta a los datos correspondientes a la voz, hemos visto que hay que asegurar su integridad (que no se haya modificado la voz o los mensajes de texto). También hay que evitar ataques de envío de paquetes fraudulentos de voz (que implicarían cortes y ruido en las conversaciones). También hay que garantizar el origen de los paquetes de voz y texto.

4.2.1. Sistemas SIP

Los sistemas SIP, como veremos también en el estándar UIT, proponen el uso de SRTP para dar seguridad a los paquetes que contienen voz. SRTP cifra los datos que transporta un paquete RTP, autentica el paquete RTP y además protege de ataques de *replay*:

SRTP

SRTP significa *secure real time transport protocol* y está definido en el RFC 3711. Este protocolo también se puede aplicar a los paquetes de vídeo, en el caso de la videoconferencia.

- La confidencialidad se logra por medio de cifrado AES.

- La autenticación también utiliza el cifrado AES.

- Un ataque de *replay* consiste en el hecho de que un atacante intercepte paquetes válidos y los envíe de manera repetida, o bien con retraso, para malograr la composición del mensaje en la aplicación destinataria. Los ataques de *replay* se evitan con el uso de números de secuencia.

Ahora bien, el problema más importante es cómo se comparte la clave que se usará en AES. Uno de los sistemas empleados en este caso es el protocolo MIKEY, un protocolo de gestión de claves propuesto con la intención de ser utilizado en aplicaciones multimedia de tiempo real. Este protocolo presenta tres maneras de compartir la llamada TEK (*traffic encryption key*):

MIKEY

MIKEY significa *multimedia Internet keying*. Está definido en el RFC 3830.

- Clave precompartida, que es un método eficiente porque usa la criptografía simétrica. A pesar de esto, hay que intercambiar una clave única TGK (*TEK generation key*) para cada participante posible en la comunicación y esto puede implicar problemas de escalabilidad.

- Clave pública, donde la TEK se transmite por medio de criptografía de clave pública, como su nombre indica. Implica el uso de una PKI.

- Diffie-Hellman, donde se emplea este sistema de intercambio. A pesar de que es el método que más recursos consume, tanto de esfuerzo computacional como de ancho de banda, se considera el más seguro puesto que proporciona la llamada *perfect forward secrecy*.

Adicionalmente, el mensaje SDP contiene un atributo llamado *crypto* que especifica los protocolos usados y las claves. Hay que recordar que, como SIP envía los mensajes en claro, se supone que el envío del SDP se hace empleando una comunicación segura, como por ejemplo la propuesta con S/MIME + TLS. Un ejemplo del atributo sería el siguiente, donde en inline se especifica la clave:

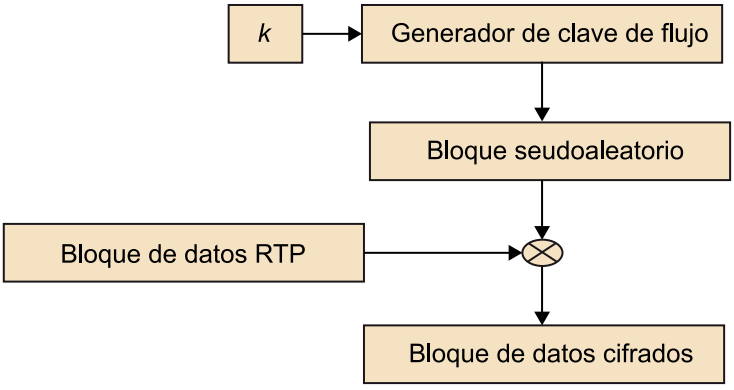
```
a=crypto:1 AES_CM_128_HMAC_SHA1_32
inline:NzB4d1BINUAvLEw6UzF3WSJ+PSdFcGdUJShpX1Zj|2^20
```

En el ejemplo anterior, se suele utilizar el protocolo AES en modo contador (*counter mode*, CM) para obtener confidencialidad. Este protocolo se usa para generar un flujo de bits de cifrado a partir de la clave de cifrado. Estos se aplican en bloques de 128 bits (el número de bits que se especifica a la entrada *crypto*) al campo de datos de los paquetes RTP, haciendo una XOR bit a bit. Este proceso se muestra en la figura 8.

Enlace de interés

En este subapartado, por razones didácticas y de simplicidad, hemos hecho una descripción breve del proceso SRTP. Como lectura recomendada, podéis leer *Securing Internet Telephony Media with SRTP and SDP*.

Figura 8. Cifrado de un paquete RTP por medio de SRTP



Para la autenticación de paquetes, SRTP crea un código de autenticación de mensaje empleando la función *hash* especificada en la entrada *crypto*, en el caso del ejemplo SHA-1, teniendo como entrada la cabecera y los datos del paquete RTP que se tiene que autenticar, y la clave.

Hay que tener presente que estas operaciones criptográficas no tienen que retardar el envío de los paquetes de voz, o cuando menos, garantizar un retraso máximo aceptable para mantener una conversación sin dificultades. En VoIP, teniendo presente que la mayor parte de los datos que genera el servicio se corresponde con la voz, y la necesidad de tener el retraso añadido mínimo posible, conviene prestar especial atención a la eficiencia de las técnicas y a los protocolos de seguridad.

Para la IM, observemos que añadir autenticación implica en determinados casos una pérdida de eficiencia. En las conversaciones de IM a menudo se envían como respuesta mensajes muy cortos (*ok*, *adiós*, etc.). En cambio, en VoIP, la digitalización y la compresión de la voz implica una sucesión de paquetes de un tamaño suficientemente grande como para que salga a cuenta añadir información de autenticación. Aun así, en IM los requisitos de tiempo real son más bajos, puesto que el servicio se puede permitir unos retrasos máximos superiores a los estipulados para VoIP.

4.2.2. Recomendaciones de la UIT para voz sobre IP

Dentro del H.235 hay tres perfiles de seguridad que tienen relación con la protección del envío de audio:

- H.235.6. Perfil de cifrado de voz con gestión de claves por medio del protocolo H.235/H.245.

- H.235.7. Uso del sistema MIKEY para la gestión de claves para SRTP.

- H.235.8. Intercambio de claves para SRTP usando canales seguros de señalización.

De este modo, usando el protocolo H.235.1 más H.235.6 y H.235.8 para intercambiar claves, se podrían garantizar las propiedades de autenticación e integridad (que proporciona el protocolo H.235.1) y confidencialidad (proporcionada por SRTP).

4.2.3. Uso de IPsec

Hay que prever que en organizaciones con sedes distribuidas geográficamente, se efectuarán llamadas que utilizarán Internet o bien la red de tránsito de la operadora de telecomunicaciones. En estas redes, los paquetes de VoIP se tienen que proteger adecuadamente, sobre todo si la información circula por Internet o bien la operadora no garantiza el uso de canales seguros del tipo de red privada virtual.

Ved también

IPsec es un sistema que proporciona confidencialidad y autenticación en redes IP. Se puede obtener más información de este sistema en el módulo "Sistemas de cortafuegos" de esta asignatura.

Así pues, los fabricantes recomiendan que, en estos escenarios, los encaminadores que unen la red interna con la red externa puedan utilizar el protocolo IPsec para asegurar la información de VoIP que circula entre sedes.

4.3. Implicaciones en los sistemas cortafuego

Hemos comentado que el uso de cortafuegos que controlen la circulación de paquetes entre las diferentes redes y los servidores es esencial para prevenir ataques de denegación y degradación del servicio. Además de controlar con diferentes reglas el tránsito de paquetes, hará falta prever la detección de ataques contra estos cortafuegos. En el caso concreto de VoIP, algunos fabricantes recomiendan que los sistemas de voz (tanto servidores como agentes de usuario) se pongan dentro de una zona segura, creada a partir de cortafuegos. Ahora bien, hay que tener algunas consideraciones en el uso de los cortafuegos y VoIP o IM, que a continuación recogemos.

Los cortafuegos suelen llevar a cabo funcionalidades de traducción de direcciones y puertos en los paquetes que entran y salen, con el objetivo de que con una misma dirección "pública" se facilite el acceso a un conjunto de dispositivos de red "internos", cada uno con una dirección de rango privado. Por medio de la monitorización de los puertos y de las direcciones indicados en los paquetes que atraviesan el cortafuegos, se puede determinar si un paquete corresponde a una sesión ya iniciada.

IPtables

En la herramienta IPtables se permiten conexiones nuevas a partir de conexiones ya permitidas por medio de *ESTABLISHED* y RELATED .

Esto permite que un servidor de VoIP o IM interno pueda establecer varias conexiones con el exterior, puesto que para la señalización de la llamada se utilizan protocolos que usan TCP y, claramente, los cortafuegos no tienen ningún tipo de problemas en controlar los paquetes correspondientes a una sesión TCP.

Ahora bien, sí que puede haber problemas para intentar controlar varios flujos de paquetes UDP, que es el caso de los datos, la voz y el vídeo, si procede. Así como los puertos TCP que se emplean para la señalización de llamadas están muy definidos y son conocidos, no pasa lo mismo con los puertos correspondientes a UDP y el tránsito de llamadas. Una solución posible sería que el cortafuegos tuviera abierto, por defecto, un rango de puertos UDP, pero esta solución claramente abre un agujero de seguridad y hace que el cortafuego pueda

perder eficacia. De hecho, ni el número de llamadas que se pueden establecer puede ser previsible: siempre habría puertos para abrir o siempre quedarían puertos abiertos.

Lo ideal sería que el cortafuegos mismo, al detectar que una sesión SIP se ha hecho efectiva, abriera un determinado rango de puertos UDP para que el canal de voz también se pudiera materializar. El cortafuegos podría analizar inmediatamente cuál es el puerto que se usa de manera efectiva para el tránsito de voz, datos o vídeo, y cerrar el resto. Aun así, este análisis tendría que implicar usar cortafuegos con gran capacidad de computación si no queremos que su aplicación degrade la eficiencia del cortafuego y añada un cuello de botella considerable.

Los cortafuegos con inspección de aplicación tienen una clara ventaja respecto de los que solo se limitan a analizar brevemente las capas de red y transporte de los paquetes que reciben. Este tipo de cortafuegos analizan los paquetes (por ejemplo, el campo del mensaje SIP donde se especifica qué puertos UDP se usarán) y lo tienen menos complicado para saber a qué llamada o conversación pertenece cada paquete UDP y, por lo tanto, qué se puede abrir y hacia qué máquina se tiene que enviar.

A pesar de esto, hemos visto que en las conexiones para sistemas de seguridad que usan protocolos de cifrado y autenticación, si la información está cifrada, los cortafuegos de aplicación lo tienen difícil para gestionar conexiones (a no ser que se los permita descifrar la información, cosa que implicaría gestión de claves, etc.).

De todos estos aspectos se desprende que el uso de cortafuegos en los sistemas de VoIP e IM seguros tiene implicaciones en la gestión de conexiones y, de rebote, en la eficiencia misma del cortafuegos. Es por eso que muchas aplicaciones emplean sistemas de túnel, con los cuales todo el tránsito de voz UDP pasa a la red por medio de un puerto determinado y usando conexiones TCP.

Resumen

En este módulo hemos estudiado los problemas de seguridad que presentan las aplicaciones de voz sobre IP y mensajería instantánea. Muchos de los problemas de seguridad que presentan son, en cierta forma, comunes y, por lo tanto, las soluciones a aplicar suelen ser similares.

Hemos empezado viendo cómo funcionan los servicios de voz sobre IP y mensajería instantánea. Hemos visto claramente que funcionan mediante dos tipos de comunicación: comunicación de señalización, en general sobre un canal TCP, y comunicación de la información (voz digitalizada o bien mensajes cortos de texto). Esta última comunicación se suele llevar a cabo en forma de un flujo de paquetes UDP. El protocolo SIP utiliza mensajes de texto también para la señalización (es decir, establecer y gestionar llamadas). El protocolo H.323 de la UIT usa paquetes binarios. Aun así, los estándares no especifican nunca que la información se tiene que cifrar.

Como tendencias de protección en voz sobre IP, hemos visto que se tiene que señalizar sobre un transporte seguro TLS, mientras que la voz puede hacer uso del SRTP, un añadido al protocolo RTP que permite confidencialidad de los paquetes y autenticación. Esta tendencia se usa tanto en sistemas SIP como en sistemas de la UIT. Para el envío de paquetes sobre Internet se recomienda, si es posible, el uso de IPsec.

Otro aspecto fundamental de los servicios estudiados es que se sustentan en una serie de servidores con diferentes funcionalidades. Estos servidores suelen ser el blanco de ataques de usuarios maliciosos, que también pueden tener como objetivo atacar los terminales de voz sobre IP o el software de usuarios de mensajería instantánea. Así pues, hay que proteger estos servidores teniendo en cuenta, sin embargo, que el uso de cortafuegos no es tan simple como se podría pensar, sobre todo si se utiliza comunicación cifrada.

En cuanto a mensajería instantánea, después de indicar que es menos complejo efectuar ataques contra la confidencialidad en mensajes de texto que en paquetes de voz, o bien es más sencillo modificar mensajes de texto que conversaciones de voz, hemos enfocado el problema principal de seguridad de la IM: la distribución de software malicioso.

Actividades

1. Averiguad cuáles son los mecanismos de seguridad que se usan en las herramientas más populares de VoIP y mensajería instantánea. Con este objetivo, usad un detector (*sniffer*) para capturar los detalles de sesiones en estos programas. Deberéis analizar si hay una protección de transporte, por ejemplo, utilizando SSL o TLS, o bien si se utilizan técnicas de autenticación de establecimiento de sesión. Finalmente, decid si se establecen diferentes comunicaciones y medidas de seguridad para la señalización y para el transporte de mensajes y voz.

2. La mensajería instantánea es una herramienta que se ha integrado a la web, sea desde sus inicios con las páginas web que permiten chats, o sea desde los portales de redes sociales. Analizad si el establecimiento de conversaciones entre miembros en redes sociales, por ejemplo, sucede de manera segura según lo que se ha visto en el módulo.

3. La Voice over IP Security Alliance (VOIPSA) es una alianza de fabricantes y desarrolladores de productos para VoIP, que vela por la promoción de las técnicas de seguridad para VoIP. Entrad en la web y echad un vistazo a algunos de los artículos más recientes en materia de seguridad en VoIP.

4. Dentro de la misma web, acceded al apartado sobre herramientas de seguridad en VoIP. Intentad probar algunos de los detectores (*sniffers*) y analizadores de SIP o, incluso de voz.

5. Consultad en Internet y buscad posibles ataques orientados a los teléfonos IP que se hayan descrito.

Glosario

agente de usuario *m* Elemento de la arquitectura VoIP o de mensajería instantánea por medio del cual el usuario interactúa con el servicio. En el caso de VoIP se trata de los teléfonos IP.

contacto *m y f* Usuario de un servicio de mensajería instantánea con el cual un usuario puede establecer una conversación.

H.323 *m* Conjunto de protocolos propuesto de la UIT sobre el establecimiento de llamadas de voz sobre redes IP. Utiliza protocolos de señalización de la telefonía tradicional, como por ejemplo, el Q.931.

IETF *f* Fuerza operacional de la ingeniería para Internet. Organización que se encarga de la estandarización de protocolos que intervienen en el funcionamiento de Internet, como por ejemplo, el TCP, el HTTP, el SIP, etc.
en Internet engineering task force

IM *f* Véase **mensajería instantánea**.

International Telecommunications Union *f* Véase **UIT**.

Internet engineering task force *f* Véase **IETF**.

IP-PBX *m* Servidor de la arquitectura VoIP encargado de gestionar el establecimiento de llamadas.

malware *m* Véase **software malicioso**.

media gateway control protocol *m* Véase **MGCP**.

mensajería *f* Tecnología que permite mantener conversaciones usando mensajes cortos de texto que se transmiten en tiempo real.
sigla IM

MGCP *m* Protocolo de control de la pasarela de medios que se encarga de la conversión de medio entre telefonía tradicional y VoIP, y la adaptación de señales de establecimiento de llamada.
en media gateway control protocol

MIKEY *m* Distribución de claves por Internet en multimedia. Protocolo sencillo que permite compartir y generar claves de sesión para el cifrado de contenido en sesiones multimedia.
en multimedia Internet keying

multimedia Internet keying *m* Véase **MIKEY**.

pasarela telefónica *f* Servidor de la arquitectura VoIP encargado de conectar la red de VoIP a la red telefónica convencional.

real-time transport protocol *m* Véase **RTP**.

RTP *m* Protocolo de transporte en tiempo real que añade a los datagramas transportados en UDP información sobre el instante de tiempo y con qué número de secuencia se han generado.
en real-time transport protocol

SDP *m* Protocolo que, dentro de un entorno SIP, se encarga de describir parámetros diversos sobre la sesión multimedia que se tiene que iniciar.
en session description protocol

session description protocol *m* Véase **SDP**.

session initiation protocol *m* Véase **SIP**.

SIP *m* Protocolo de inicio de sesiones que utiliza peticiones en mensajes de texto para establecer conexiones multimedia en entornos IP.
en session initiation protocolo

softphone *m* Véase **teléfono IP**.

software malicioso *m* Software diverso cuyo objetivo es causar alguna molestia, leve (por ejemplo, aparición de publicidad masiva) o grave (pérdida de datos del sistema), a los usuarios.
en malware

teléfono *m* Dispositivo que hace las funciones de digitalización y reproducción de voz y que permite el establecimiento de llamadas en un entorno de VoIP. Software que hace las funciones de teléfono IP.
en softphone

UIT *m* Unión Internacional de Telecomunicaciones. Organismo que se encarga de la estandarización de protocolos y sistemas que intervienen en las telecomunicaciones.
en International Telecommunications Union

voz sobre IP *f* Tecnología que permite establecer conversaciones telefónicas usando datagramas IP como medio de transporte.
sigla VoIP

VoIP *f* Véase **voz sobre IP**.

Bibliografía

Baugher, M. y otros. *Securing Internet Telephony Media with SRTP and SDP* (en línea). Cisco.

Butcher, D. y otros (2007). "Security Challenge and Defense in VoIP Infrastructures". *IEEE Transacions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* (vol. 37, núm. 6, pág. 1152-1162).

Endler, D. y otros (2006). *Hacking Exposed VoIP: Voice Over IP Security Secrets & Solutions*. McGraw-Hill Professional Publishing.

Gollmann, D. (2005). *Computer Security* (2.ª ed.). Wiley.

Herrera Joancomartí, J. (2006). *Aspectos avanzados de seguridad en redes*. Editorial UOC.

Mannan, M. y otros (2005). *On Instant Messaging Worms, Analysis and Countermeasures, Proceedings of the 2005 ACM workshop on Rapid Malcode*.

Sisalem, D. y otros (2009). *SIP Security*. Wiley.

Symantec. *Securing Instant Messaging*.

Zhua, Y. y otros (2011). "Traffic analysis attacks on Skype VoIP calls". *Computer Communications* (vol. 34, núm. 10, pág. 1202-1212).