R. Ocaña, and I. Agudo, "Comunicaciones VoIP cifradas usando Intel SGX", *V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC)*, 2019.
NICS Lab. Publications: https://www.nics.uma.es/publications

Comunicaciones VoIP cifradas usando Intel SGX

Raúl Ocaña Isaac Agudo
Network, Information and Computer Security (NICS) Lab
Universidad de Málaga, 29071
{roa, isaac}@lcc.uma.es

***Resumen*—Cada día es más frecuente encontrar servicios en internet gestionados desde plataformas online y con la expansión de la tecnología IoT, los *smartphones*, las *smartTV* y otros tantos dispositivos: la autenticación, la distribución y al fin y al cabo, la comunicación entre extremos puede verse seriamente comprometida si dicha plataforma es atacada. La inclusión de nuevas medidas de seguridad en este tipo de ecosistemas requiere de un cambios sustancial de la arquitectura subyacente en muchos casos, por lo que su avance es lento. En este trabajo se trata de forma concreta el desarrollo de una alternativa *OpenSource* a uno de estos servicios, la telefonía IP (VoIP), que esta expandiéndose cada día más, empezando por redes locales y privadas y llegando a grandes centralitas de conmutación de tele operadoras, consiguiendo así una transmisión de voz segura extremo a extremo transparente para los servidores VoIP, que no requiera modificar la infraestructura subyacente.**

***Index Terms*—Intel SGX, VoIP, Seguridad Extremo a Extremo, Cifrado, Comunicaciones**

Tipo de contribución: *Investigación en desarrollo.*

I. Introducción

Uno de los protocolos predominantes para el desarrollo de las comunicaciones de Voz sobre IP (VoIP) es el protocolo RTP (Real-time Transport Protocol). Mediante el uso de este protocolo se puede producir una comunicación fluida y síncrona, si bien no siempre se hace uso de medidas de seguridad que lo acompañen. El uso de las alternativas seguras, SRTP [2] o ZRTP [3], no está tan extendido como sería deseable; esto puede poner en riesgo la seguridad tanto de la información tratada durante estas llamadas, la de sus interlocutores y su privacidad.

Lo sistemas que dan soporte las comunicaciones VoIP no dejan de ser servicios en línea, y por tanto es amplia la lista de posibles amenazas heredadas: desde ataques lanzados a redes subyacentes, a los protocolos de transporte, a los dispositivos de transmisión VoIP y sus aplicaciones, servidores, puertas de enlace, o incluso a su protocolo de configuración DHCP, llegando incluso al sistema operativo [1]. Esto unido a que en algunos casos la confianza en los proveedores de estos servicios no es plena obliga a trabajar asumiendo una configuración de sistemas *Honestos-pero-Curiosos* [10], es decir que proporcionan el servicio deseado pero pueden estar interesados en los datos de los usarios, en este caso, sus llamadas.

La necesidad por tanto de proveer a estos sistemas de una seguridad adecuada es una tarea compleja, que implica muchos factores independientes. Es por ello que se presenta la idea de encapsular la seguridad de estos protocolos, de manera paralela a la ejecución de las propias aplicaciones y del sistema operativo, en pequeños espacios de memoria, aislados y sellados contra el acceso tanto externo como interno a la propia máquina. Estos espacios de memoria son los Enclaves [4], una tecnología relativamente reciente de Intel llamada SGX (*Secure Guard Extensions*) que permite el almacenamiento y ejecución de código de forma aislada al sistema operativo utilizando al procesador como agente. De esta manera se podría generar confianza entre ambos extremos ignorando cualquier elemento intermedio de la cadena de comunicación, el cual será transparente a nuestro protocolo de seguridad subyacente.

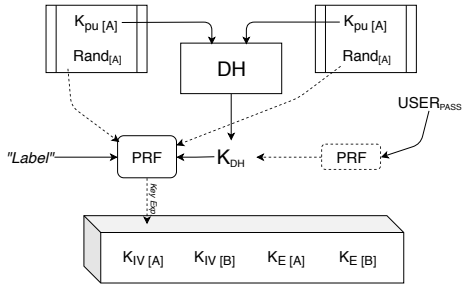

Figura 1. Derivación de claves entre procesos.

Este trabajo parte del desarrollo de una aplicación de Chat cifrado basado en un intercambio DH, donde todas las operaciones criptográficas se implementan dentro una enclave Intel SGX [8] e intenta ampliar su ámbito de aplicación a las llamadas VoIP. Todo el código del proyecto se puede encontrar en el repositorio oficial [9].

II. Implementación

Para gestionar el desarrollo de esta investigación se ha usado como motor de llamadas VoIP la librería PJSIP[1], de dominio público y que ofrece diferentes aplicaciones básicas para pruebas y realización de llamadas. En términos de desarrollo se plantean las siguientes etapas:

II-A. Emparejamiento y generación de claves

La generación de claves es un proceso crítico en todo el diseño; para esta tarea se ha planteado un proceso en cascada que está inspirado en el diseño de TLS (Transport Layer Security) [7] para la derivación y expansión de claves mediante el uso de la función pseudoaleatoria PRF (Pseudo Random Function).

Tal y como podemos ver en la Figura 1, para la derivación de las claves se hace uso del protocolo Diffie-Helman (DH), que se ejecutará bien en un canal paralelo o en el contexto del protocolo SIP (Session Initiation Protocol) usando mensajes instantáneos cuando el servidor los soporte. En ambos casos el servidor será el mismo, variando sólo el canal a través del cual se negocie dicha clave inicial. También se define un mecanismo simplificado en el caso en el que no sea posible ejecutar el protocolo DH, en cuyo caso se genera

[1]Repositorio oficial de PJSIP: https://github.com/pjsip/pjproject.

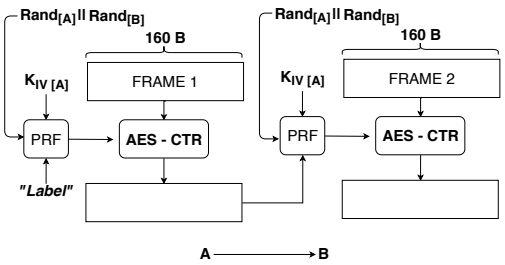

Figura 2. Cifrado de paquetes.

la clave maestra del sistema (MK) a través de la función PRF y usando como entrada una contraseña compartida entre los interlocutores. A partir de esa clave maestra y a través de la función PRF se genera un nuevo bloque de claves (*key expansion*) que alberga las claves de cifrado de ambos clientes, así como la clave de derivación de los IV usados en el cifrado de las tramas. Para dicha expansión de claves se utiliza una *etiqueta* definida y dos números aleatorios aportados por los clientes, esto hace que las claves sean diferentes en cada ejecución del protocolo aunque se utilizara la misma contraseña.

II-B. Proceso de cifrado

El cifrado se ha implementado dentro del mecanismo de codificación de voz, integrándolo en el códec G.722 [5]. A la hora de realizar la codificación se aplica un proceso extra de cifrado, usando el modelo de cifrado secuencial encadenado descrito a continuación (ver Figura 2). Al inicio, se parte de un IV definido y concreto, que ambos clientes son capaces de generar con facilidad tanto en transmisión, como en recepción; este se crea a partir de la concatenación de los números aleatorios proveídos por el cliente que inicia la comunicación, a partir de ahora A, y el que la recibe, en adelante B. En posteriores tramas el cifrado utilizará como *etiqueta* para la generación del IV (también mediante PRF) la última trama enviada. Las tramas son de 160 Bytes y para cifrarlas utilizamos el algoritmo AES en modo CTR [6] usando en cada trama el IV correspondiente. Al no tener que enviar el IV, evitamos tener que alterar el tamaño de las tramas de voz. Este modelo de encadenamiento de las tramas se asimila al modo CBC. Si bien en la figura se muestra el cifrado de las tramas de A a B el cifrado en el otro sentido se realiza de forma análoga pero con las claves correspondientes.

II-C. La aplicación y sus contextos

Como se ha mencionado al principio de este documento este desarrollo usa Intel SGX como sello de garantía. En ambos clientes, los enclaves son los encargados de generar, procesar y almacenar la clave compartida y sus derivadas, así como inicializar los métodos de cifrado y descifrado, que posteriormente serán usados por el códec G.722.

Tal y como se puede ver en la Figura 3, dentro de la aplicación encontramos dos grandes contextos: el primero, que es el que usa SIP o un canal auxiliar previamente acordado para el emparejamiento y la derivación de claves compartidas; y un segundo contexto en el que se lleva a cabo la comunicación RTP y que va a ser el que prácticamente predomine durante el uso de dicha aplicación. En ambos contextos se utiliza un servidor: tanto para el acuerdo de

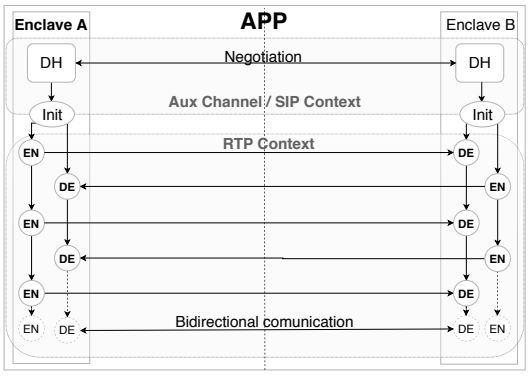

Figura 3. Modelo contextual de la aplicación

claves, como para el desarrollo de la comunicación; que hace de intermediario entre las partes. Gracias al cifrado extremo a extremo implementado dentro de los enclaves SGX, ninguno de los servidores tendrá acceso a la información en claro.

III. Conclusiones y Trabajo Futuro

Este trabajo muestra como se puede integrar un cifrado extremo a extremo de forma casi transparente al servidor VoIP, entrelazando el proceso de cifrado con el de codificación de la voz. Esto permite una comunicación segura extremo a extremo, donde el servidor no tiene acceso en ningún momento a la llamada. Esta solución a nivel de códec evita que el proceso de cifrado tenga lugar en el contexto de la aplicación y que se realicen altos cambios en la librería VoIP, aún así se valora la futura integración de protocolos como SRTP o SRTCP dentro del contexto del enclave, aunque esto implicaría mayores modificaciones en las librerías actuales.

Como trabajo futuro, se plantea la implementación de un cliente plenamente funcional donde el acuerdo de claves DH se realice de forma transparente a través del protocolo SIP, sin la necesidad de utilizar un canal fuera de banda. El encapsulamiento del intercambio DH dentro de SIP mejoraría la escalabilidad y la usabilidad del sistema.

Otras líneas futuras son el estudio de la integración con otros códecs de audio así como la integración con otros clientes o el soporte para autenticación extremo a extremo.

Referencias

[1] SICKER, Douglas C.; LOOKABAUGH, Tom. "VoIP security: Not an afterthought". *Queue*, vol. 2, no 6, p. 56. 2004.

[2] McGrew, D. and E. Rescorla, "Datagram Transport Layer Security (DTLS) Extension to Establish Keys for the Secure Real-time Transport Protocol (SRTP)", RFC 5764, DOI 10.17487/RFC5764, May 2010.

[3] P. Zimmermann, ZRTP: Media Path Key Agreement for Unicast Secure RTP, RFC 6189 (proposed standard), April 2011.

[4] COSTAN, Victor; DEVADAS, Srinivas. İntel SGX Explained". *IACR Cryptology ePrint Archive*, vol. 2016, no 086, p. 1-118. 2016.

[5] MERMELSTEIN, Paul. "G. 722: a new CCITT coding standard for digital transmission of wideband audio signals". *IEEE Communications Magazine*, vol. 26, no 1, p. 8-15. 1988.

[6] Dworkin, Morris. Recommendation for block cipher modes of operation. methods and techniques. No. NIST-SP-800-38A. National Inst of Standards and Technology Gaithersburg MD Computer security Div, 2001.

[7] DIERKS, T. y RESCORLA, E. "The Transport Layer Security (TLS) Protocol Version 1.2". *IETF. Request for Comments, 5246*. 2008.

[8] Raúl Ocaña. .Aplicación de chat segura basada en Intel SGX". *TFG, E.T.S.I. Telecomunicaciones, Universidad de Málaga*. 2018.

[9] Repositorio oficial del proyecto: https://github.com/nicslabdev/VoIP_seguro_IntelSGX

[10] Paverd AJ, Martin A, Brown I. "Modelling and automatically analysing privacy properties for honest-but-curious adversaries". *Tech. Rep.*. 2014.