DESAFÍOS Y OPORTUNIDADES PARA LA SEGURIDAD DE WPA/WPA2 EN REDES DOMÉSTICAS: UN ENFOQUE BASADO EN EL IDIOMA NATIVO DE LOS USUARIOS

DANIEL FELIPE MONTENEGRO

REDES DE COMPUTADORES

PROFESOR OCTAVIO JOSE SALCEDO

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE INGENIERÍA

RESUMEN

Esta investigación examina las vulnerabilidades de seguridad de los enrutadores de redes domésticas protegidos por WPA/WPA2. A pesar de su amplia adopción, se ha descubierto que estos estándares tienen debilidades explotables, especialmente en la captura del "4-way handshake". La descifrado offline de la clave precompartida plantea un desafío significativo, a menudo requiriendo ataques de fuerza bruta o diccionario. Sin embargo, los diccionarios existentes se centran principalmente en el inglés y carecen de recursos especializados para contextos de habla hispana.

Para abordar esta limitación, este estudio utiliza el corpus CORPES, una colección de elementos del idioma español categorizados por su frecuencia de uso, para construir un diccionario adaptado. Este diccionario especializado tiene como objetivo mejorar la recuperación de claves precompartidas en entornos de habla hispana. Al proporcionar un enfoque nuevo que se alinea con el idioma nativo de los usuarios, esta investigación contribuye a mejorar la seguridad de WPA/WPA2 en redes domésticas.

Los hallazgos de este estudio arrojan luz sobre los desafíos y oportunidades específicos para mejorar la seguridad de WPA/WPA2 en contextos de habla hispana, ofreciendo conocimientos valiosos para administradores de redes e investigadores que trabajan en el campo de la seguridad de redes inalámbricas.

ABSTRACT

This research examines the security vulnerabilities of WPA/WPA2-protected home network routers. Despite their widespread adoption, these standards have been found to have exploitable weaknesses, particularly in the capture of the 4-way handshake. Offline cracking of the pre-shared key poses a significant challenge, often requiring brute-force or dictionary attacks. However, existing dictionaries primarily focus on English, lacking specialized resources for Spanish-speaking contexts.

To address this limitation, this study utilizes the CORPES corpus, a collection of Spanish language elements categorized by their frequency of use, to construct a tailored dictionary. This specialized dictionary aims to enhance the recovery of pre-shared keys in Spanish-speaking environments. By providing a fresh approach that aligns with the native language of users, this research contributes to improving the security of WPA/WPA2 in home networks.

The findings of this study shed light on the specific challenges and opportunities for enhancing WPA/WPA2 security in Spanish-speaking contexts, offering valuable insights for network administrators and researchers working in the field of wireless network security.

KEYWORDS: WPA, WPA2, 4-way handshake, IEEE 802.11i, CORPES, WiFi security, password cracking, Offline cracking, Brute-force attack, Dictionary attack, Combinator attack

INTRODUCCIÓN

Desde la propuesta inicial del sistema de cifrado WEP (Wired Equivalent Privacy) en 1999, seguido por su corrección de seguridad en 2003 con la implementación de WPA (Wi-Fi Protected Access) basado en el estándar IEEE 802.11i, hasta la ratificación completa del estándar en 2004 con la creación de WPA2 (Wi-Fi Protected Access 2), la seguridad de los enrutadores de redes domésticas ha evolucionado y hoy contamos incluso con la versión más reciente del estándar que es WPA3 (Wi-Fi Protected Access 3) anunciada en el año 2018 y la cual pese a estar siendo difundida, aún no hace presencia en todos los hogares, esto debido a la existencia de dispositivos que no son compatibles con este tipo de conexión e interactúan con las redes locales; esta dificultad hace que WPA2 sea la configuración más predominante de seguridad en la actualidad y la opción a instalar en los hogares por parte de la mayoría de ISPs (Internet Service Providers).

La existencia de este estándar durante tanto tiempo como su difusión alrededor del mundo, solo hacen de WPA2 un viejo conocido, y es que hoy en día existen numerosas vulnerabilidades reportadas y ampliamente documentadas; donde se deja ver que la seguridad ofrecida no es tan robusta como se pensaba. Una de las vulnerabilidades más importantes es la que se relaciona con WPS (Wi-Fi Protected Setup), la cual por medio del PIN asociado al router permite en muy poco tiempo encontrar la clave precompartida de la red inalámbrica. Aunque este tipo de ataque es altamente efectivo, es fácilmente evitable desactivando la opción de WPS en el router de una red doméstica, sin embargo, este cambio no hace que la red sea automáticamente a prueba de la vulnerabilidad existente más explotada: la captura del 4-way handshake, un intercambio de mensajes de autenticación entre el router y el dispositivo a conectarse, donde en uno de ellos se encuentra encriptada la clave precompartida.

FORMULACIÓN DEL PROBLEMA

El intercambio de mensajes puede ser capturado por cualquier persona que se encuentre en el radio de la señal y no es necesario contar con un equipo especializado o conocimientos avanzados, ya que solo basta con tener una tarjeta de red que admita el modo monitor y ejecutar scripts distribuidos en sistemas operativos de código abierto, como por ejemplo Kali Linux; una distribución especializada en encontrar y explotar vulnerabilidades que no solo se limitan a las redes de computadores.

Desencriptar la clave precompartida de manera offline es posible una vez se tiene capturado 4-way handshake. Abordar este problema, casi siempre implica dos alternativas similares, pero que tienen grandes diferencias en su rendimiento.

La primera opción, que es a su vez lo primero que se le puede venir a la cabeza a cualquier persona es intentar un ataque de fuerza bruta, el cual consiste en probar todas las posibles permutaciones formadas por los caracteres que pueden conformar la clave, es decir probar una por una todas las claves posibles hasta encontrar la correcta. Al principio, puede parecer una alternativa no tan descabellada y al alcance de la mano, sin embargo, más adelante nos daremos cuenta porque, al menos para descifrar la

clave precompartida de una red protegida con WPA/WPA2, no es algo viable y tampoco materialmente posible.

La segunda alternativa es mucho más acertada, pero eso no implica que sea sencilla y en cambio se agrega un nivel extra de dificultad: ya no haremos la desgastante labor de probar todas las claves existentes, ahora solo probaremos un selecto conjunto de claves candidatas. Este método es mucho más efectivo, pero compromete a quien lo usa a cumplir con condiciones claras y específicas, por ejemplo una de ellas puede parecer tan obvia como problemática, y es que es un requisito que la clave a encontrar debe estar en el diccionario a utilizar. Otra es que la calidad del ataque depende directamente de la calidad del diccionario, ya que en un supuesto muy mal construido, podemos generar un diccionario el cual tenga todas las posibles claves y en resumen de cuentas ya no estaríamos haciendo realmente un ataque de diccionario, sino en cambio uno de fuerza bruta. Para fortuna nuestra, no todos los escenarios son tan desalentadores, en la otra cara de la moneda, también existe un escenario poco probable, pero posible y es que nuestro diccionario puede ser nada más un puñado de claves y la clave a encontrar este entre ellas por lo que solo nos tomara unos cuantos milisegundos la desencriptación en este contexto ideal.

RECURSOS

Corpus del Español del Siglo XXI (CORPES)

La Real Academia Española (RAE) define corpus como un conjunto formado por miles de textos empleados habitualmente para conocer el significado y características de palabras, expresiones y construcciones a partir de los usos reales registrados. Un corpus general (llamado de referencia) tiene como propósito básico el de servir para obtener las características globales que presenta una lengua en un momento determinado de su historia. (1)

El corpus del español del siglo XXI se denomina CORPES XXI, los textos que integran el CORPES se seleccionan de acuerdo con una serie de parámetros y son tratados con un sistema de codificación especialmente diseñado para este corpus y para la recuperación de sus datos desde cualquiera de esos parámetros. Desde diciembre de 2013 es posible realizar consultas al CORPES XXI a través de una aplicación específica que permite recuperar los casos contenidos en el corpus de una palabra, una expresión o una categoría o subcategoría gramatical. (1)

Hashcat

Hashcat es una herramienta de recuperación de contraseñas avanzada y de código abierto. Es conocido por ser el cracker de contraseñas más rápido del mundo y cuenta con una amplia gama de características. Hashcat es compatible con múltiples sistemas operativos, como Linux, Windows y macOS, y es compatible con diferentes tipos de dispositivos, incluidas CPU, GPU y APU. Es una poderosa herramienta ampliamente utilizada por profesionales de la seguridad y auditores para probar la fortaleza de las contraseñas y realizar análisis de seguridad. (2)

Kali linux

Kali Linux (anteriormente conocido como BackTrack Linux) es una distribución de Linux basada en Debian de código abierto destinada a pruebas de penetración avanzadas y auditorías de seguridad. Kali Linux contiene modificaciones específicas de la industria, así como varios cientos de herramientas dirigidas a diversas tareas de

seguridad de la información, como pruebas de penetración, investigación de seguridad, informática forense, ingeniería inversa, gestión de vulnerabilidades y pruebas de equipo rojo. Es una solución multiplataforma, accesible y de libre disposición para profesionales como aficionados a la seguridad de la información. (3)

Raspberry Pi 4
Es una computadora de escritorio de tamaño reducido basada en una placa de desarrollo de bajo costo y alto rendimiento que ofrece capacidades avanzadas para diversos proyectos. Para utilizar una Raspberry Pi 4 solo es necesaria una fuente de alimentación USB-C de 15W, una tarjeta microSD, un teclado, un mouse y un cable HDMI para conectar la placa a un monitor. (4)

Adaptador USB Inalámbrico de Alta Sensibilidad TL-WN722N
El adaptador USB inalámbrico TL-WN722N permite conectar una computadora de escritorio o portátil a una red inalámbrica y acceder a una conexión a Internet de alta velocidad. Tiene una antena externa de alta ganancia desmontable de 4dBi para asegurar una transmisión y recepción de señal más fuerte, fortaleciendo notablemente la potencia de la señal del adaptador USB. Es compatible con Windows 11 / 10 / 8.1 / 8/7 / XP, Mac OS X, Linux. (5)

IdeaPad 3-14ALC6 Laptop - Type 82KT
Procesador AMD Ryzen™ 3 5300U Processor(Ryzen™ 3 5300U), Memoria 2x 4 GB DDR4-3200, Sistema operativo Windows 11 Home Single Language 64 (SP:Spanish), Unidad de disco duro 512 GB SSD PCIe

INVESTIGACIONES RELACIONADAS

A continuación, se destacan todas las investigaciones que fundamentan la base teórica en la que se basó este proyecto, estos avances permitieron ser la orientación de la investigación expuesta en este documento:

- N. Pimple, T. Salunke, U. Pawar and J. Sangoi, "Wireless Security — An Approach Towards Secured Wi-Fi Connectivity," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 872-876, doi: 10.1109/ICACCS48705.2020.9074350.
- S. Vinjosh Reddy, K. Sai Ramani, K. Rijutha, S. Mohammad Ali and C. Pradeep Reddy, "Wireless hacking - a WiFi hack by cracking WEP," 2010 2nd International Conference on Education Technology and Computer, Shanghai, China, 2010, pp. V1-189-V1-193, doi: 10.1109/ICETC.2010.5529269.
- H. Peng, "WIFI network information security analysis research," 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, China, 2012, pp. 2243-2245, doi: 10.1109/CECNet.2012.6201786.
- Arash Habibi Lashkari, Mir Mohammad Seyed Danesh and B. Samadi, "A survey on wireless security protocols (WEP, WPA and WPA2/802.11i)," 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, China, 2009, pp. 48-52, doi: 10.1109/ICCSIT.2009.5234856.
- J. Liu, X. Ye, J. Zhang and J. Li, "Security Verification of 802.11i 4-

Way Handshake Protocol," 2008 IEEE International Conference on Communications, Beijing, China, 2008, pp. 1642-1647, doi: 10.1109/ICC.2008.317.

RESULTADOS ESPERADOS

Entender la naturaleza de la clave precompartida de una red protegida con WPA/WPA2 es fundamental para comprender la magnitud del problema que estamos enfrentando, desde el estándar IEEE 802.11i se establece que la clave debe tener una longitud mínima de 8 caracteres y una longitud máxima de 63 caracteres.

Los caracteres permitidos son los ASCII imprimibles (del 0x20 al 0x7E), lo que nos indica que para cada posición tenemos 95 posibilidades. Haciendo un cálculo simple podemos afirmar que el número de claves diferentes para un número *n* arbitrario entre 8 y 63, es igual a: *95^8 + 95^9 + 95^10 + … + 95^n*. Al realizar este cálculo para la cota superior de nuestro problema que es 63, la aproximación del resultado sería un número cercano a 4e+124 claves.

Estas son en definitiva demasiadas claves, pero si nos aventuramos en la colosal labor de descifrarlas tendríamos que hacer de esta cifra algo un poco más real y que esté expresado en términos de nuestras capacidades actuales. En la computadora expuesta en la sección de recursos se realizó un benchmark en Hashcat usando la herramienta integrada para este propósito “-b” y eligiendo el Hash-Mode “-m 22000” (WPA-PBKDF2-PMKID+EAPOL); el resultado aproximado de esta prueba de rendimiento fue de 50000 H/s.

Si tenemos en cuenta que un año tiene un estimado de 31536000 segundos, es posible afirmar que esta computadora se tardaría alrededor de 2.5e+112 años para descifrar todas las posibles claves existentes en este cifrado. Al no contar materialmente con el tiempo suficiente, se traslada el problema del proyecto a la construcción de un diccionario que permita en un tiempo óptimo o al menos razonable (comparado con un ataque de fuerza bruta) encontrar la clave precompartida de una red protegida con WPA/WPA2.

En la red abundan propuestas de diccionarios especializados para este tipo de ataques, incluso algunos pueden ser encontrados fácilmente precompartidos en el directorio /usr/share/wordlists de la versión más reciente de Kali Linux. El más popular de todos estos diccionarios es de lejos uno llamado rockyou el cual tiene casi 15 millones de claves y usado de la mano con un sólido conjunto de reglas en Hashcat puede ser altamente efectivo. Sin embargo, todos estos diccionarios tienen algo en común que puede llegar a ser una falencia y es que plantean claves predominantemente en inglés o que están pensadas como si hubieran sido creadas por una persona la cual su lengua nativa es el inglés.

Este proyecto usa como soporte de su investigación al CORPES creado por la RAE, el cual es un listado de elementos usados en el idioma español y categorizados por su frecuencia de uso por cada millón de palabras. Usar este recurso nos permite crear una aproximación de un diccionario especializado en contextos donde la persona que creó la clave tiene como lengua nativa el idioma español y así ofrecer una propuesta fresca que no compita directamente con los valiosos recursos que ya pone la comunidad en manos de todos nosotros.

OBJETIVOS ESPECÍFICOS

- Desarrollar un diccionario especializado basado en el corpus CORPES (Corpus del Español del Siglo XXI) que permita descifrar contraseñas en idioma español mediante el análisis semántico.
- Evaluar el rendimiento del diccionario especializado mediante la comparación con un ataque hipotético de fuerza bruta dirigido a la misma clave precompartida, con el fin de determinar su efectividad y eficiencia en la recuperación de contraseñas.
- Realizar pruebas de campo utilizando Kali Linux para analizar la vulnerabilidad de una red en un entorno de pruebas, en el que se tenga control sobre el router, con el objetivo de identificar y evaluar las posibles debilidades de seguridad y recomendaciones para su mejora.

METODOLOGÍA

El proyecto se fundamenta en la premisa de que las claves son creadas por humanos y para humanos. Partiendo de esta idea, es razonable considerar que las contraseñas pueden tener una relación semántica o mnemotécnica, es decir, una conexión con el significado o la memoria. Desde esta perspectiva, se presupone la estrecha vinculación existente entre el lenguaje y la creación de contraseñas por parte de las personas.

Cuando los seres humanos inventan contraseñas, es altamente probable que utilicen como componentes palabras que forman parte de su lenguaje o que guarden relación con su contexto personal. Por ejemplo, pueden incluir su propio nombre o palabras que evocan recuerdos significativos para ellos. A diferencia de las máquinas, las personas tienden a preferir no utilizar conjuntos de caracteres aleatorios.

El objetivo central de este proyecto consiste en encontrar esas contraseñas que son un conjunto de elementos mezclados de manera arbitraria y que, como resultado, generan una clave diseñada para ser recordada por seres humanos. En otras palabras, se busca identificar y comprender los patrones y bloques semánticos que las personas emplean para crear contraseñas y utilizar esa información para desarrollar estrategias de generación de claves más efectivas.

Al enfocarnos en la búsqueda de contraseñas que tengan un componente semántico o mnemotécnico, podemos mejorar la experiencia de los usuarios al recordar sus claves y, al mismo tiempo, garantizar un nivel adecuado de seguridad. Al comprender cómo las personas relacionan su lenguaje y sus experiencias con la creación de contraseñas, podemos desarrollar herramientas y técnicas que se ajusten mejor a las necesidades humanas, brindando una mayor comodidad y facilidad de uso en entornos digitales seguros.

DESARROLLO E IMPLEMENTACIÓN

En este estudio, se llevó a cabo la configuración de una Raspberry Pi 4 con el sistema operativo Kali Linux, junto con la instalación de un adaptador WiFi de alta sensibilidad TL-WN722N. El propósito principal fue capturar los datos de una red de pruebas creada en un enrutador utilizando la herramienta hcxdumptool de Hashcat.

Después de la instalación de Kali Linux, se procedió a configurar el adaptador WiFi TL-WN722N en la Raspberry Pi 4. Este

adaptador inalámbrico permite poner la tarjeta de red en modo monitor, lo cual posibilita la captura de señales de redes WiFi cercanas. En nuestro caso, nos enfocamos únicamente en capturar las señales del router utilizado en nuestro laboratorio de pruebas. Se verificó que el adaptador estuviera reconocido y configurado correctamente en la Raspberry Pi 4.

Una vez finalizada la configuración del adaptador WiFi, se utilizó la herramienta hcxdumptool de Hashcat para capturar los datos de la red de pruebas. Hcxdumptool es una herramienta especializada en la captura de paquetes, la cual permite obtener el intercambio de mensajes de autenticación en 4 pasos en redes protegidas con WPA/WPA2. Esta funcionalidad resultó fundamental para capturar los datos necesarios para llevar a cabo los análisis posteriores.

En primer lugar, se cargó el corpus CORPES, que contiene palabras en español y su frecuencia de aparición. A continuación, se aplicó una función de transliteración para convertir las palabras a una forma más sencilla y estandarizada. Esto facilita la generación de permutaciones y la creación de diccionarios.

Se realizaron análisis estadísticos, como la generación de un histograma de las longitudes de las palabras presentes en el corpus. Se calcularon la media y la desviación estándar de las longitudes para establecer un rango óptimo de longitud de palabras para el diccionario.

Luego, se eliminaron del corpus las palabras cuyas longitudes estaban fuera del rango establecido. También se agruparon las transliteraciones de las palabras y se calculó la suma de sus frecuencias, lo que permitió obtener un ranking de las transliteraciones más comunes.

Se exploraron diferentes enfoques para determinar la cantidad de palabras y la profundidad de permutaciones a considerar en la generación del diccionario. Se utilizaron probabilidades y ponderaciones basadas en la frecuencia de las palabras y en la teoría de la información para determinar la relevancia de las combinaciones de palabras.

Se generaron diccionarios personalizados que equilibraban la frecuencia de aparición de las palabras y la diversidad de las permutaciones. Se evaluaron diferentes configuraciones para obtener un equilibrio óptimo entre el tamaño del diccionario y la efectividad del ataque.

CONCLUSIONES

Los resultados mostraron que el proyecto fue efectivo en el descifrado de claves precompartidas de redes protegidas con WPA/WPA2. En comparación con un enfoque de fuerza bruta, que intentaría todas las combinaciones posibles de caracteres, el uso de un diccionario personalizado redujo significativamente el tiempo necesario para encontrar claves de una longitud determinada.

Sin embargo, se observó que el éxito del proyecto depende en gran medida de la existencia de un patrón semántico en las claves precompartidas. Si las claves no siguen un patrón semántico o están compuestas por palabras con baja frecuencia en el corpus, la efectividad del ataque se reduce.

Además, la creación del diccionario requiere la configuración adecuada de parámetros como la frecuencia de las palabras consideradas y la cantidad de permutaciones generadas. Estos

parámetros son determinantes para el éxito del ataque y dependen tanto del conocimiento del usuario de la herramienta como de la comprensión del objetivo al que se dirige el ataque.

En conclusión, este proyecto demostró que el uso de diccionarios personalizados basados en un corpus lingüístico puede ser altamente efectivo para descifrar claves precompartidas en redes protegidas con WPA/WPA2. Sin embargo, la efectividad del ataque está condicionada por la presencia de un patrón semántico en las claves y por la elección adecuada de los parámetros de configuración del diccionario. Es importante destacar que este enfoque no garantiza el éxito en todos los casos, ya que existen situaciones en las que las claves no siguen un patrón semántico claro o están compuestas por palabras poco frecuentes.

Para mejorar la efectividad de este tipo de ataques, se recomienda utilizar fuentes de datos más amplias y actualizadas para construir el diccionario personalizado. Además, es fundamental que los usuarios comprendan los parámetros de configuración y ajusten adecuadamente la frecuencia de las palabras consideradas y la cantidad de permutaciones generadas.

Es importante tener en cuenta que este proyecto se realizó con fines educativos y de investigación, con el objetivo de analizar la vulnerabilidad de las redes protegidas con WPA/WPA2. El uso indebido de esta técnica para acceder ilegalmente a redes sin autorización es una violación de la privacidad y está sujeto a sanciones legales.

En resumen, el uso de diccionarios personalizados basados en un corpus lingüístico puede ser una herramienta valiosa para descifrar claves precompartidas en redes protegidas con WPA/WPA2. Sin embargo, su efectividad depende de varios factores, como la existencia de un patrón semántico en las claves y la correcta configuración de los parámetros del diccionario.

REFERENCIAS

1. REAL ACADEMIA ESPAÑOLA: Banco de datos (CORPES XXI) [en línea]. *Corpus del Español del Siglo XXI (CORPES).* <http://www.rae.es> [03 DE JUNIO DE 2023].
2. HASHCAT. Advanced password recovery. [En línea]. <https://hashcat.net/hashcat/> [03 DE JUNIO DE 2023].
3. KALI: Kali Linux Features.[En línea]. <https://www.kali.org/docs/introduction/what-is-kali-linux/> [03 DE JUNIO DE 2023].
4. RASPBERRY PI: Raspberry Pi 4. [En línea]. <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/> [03 DE JUNIO DE 2023].
5. TP-LINK. Adaptador USB Inalámbrico de Alta Sensibilidad a 150 Mbps. [En línea]. <https://www.tp-link.com/co/home-networking/adapter/tl-wn722n/> [03 DE JUNIO DE 2023].
6. Hoorvitch, I. (2021). CRACKING WIFI AT SCALE WITH ONE SIMPLE TRICK. Blog obtenido en CYBERARK: The identity security Company. [En línea]. <https://www.tp-link.com/co/home-networking/adapter/tl-wn722n/> [03 DE JUNIO DE 2023].