Máster en Ingeniería Informática

Sistemas de Ciberseguridad 2024-2025
Grupo 1

Workshop 1

"Conning the Crypto Conman: End-to-End Analysis of Cryptocurrency-based Technical Support Scams"

Luis Daniel Casais Mezquida – 100429021

Profesor

Sergio Pastrana Portillo

1. Resumen del artículo

El artículo evalúa una estafa que se ha vuelto común en la red social 'X' (anteriormente conocida como 'Twitter'), la estafa del soporte técnico de carteras de criptomonedas.

La estafa consiste en usar *bots* para hacerse pasar por personas realies, que vienen a ayudar a un usuario que realiza una publicación pidiendo ayuda al encontrarse con problemas técnicos relacionados con su cartera de criptomonedas. Los atacantes hacen entonces que la víctima se pase a otro canal de comunicación en el que comienza la estafa: o bien les engañan para obtener sus claves privadas y así robar el dinero de sus carteras, o bien les piden dinero por prestar servicios que no acaban realizando.

Los autores del artículo han creado perfiles falsos en la red social y, mediante el uso de una herramienta que han desarrollado, realizado publicaciones para atraer a estos estafadores e interactuar con ellos, recabando automáticamente información de las interacciones y de los perfiles. Con ello, han logrado analizar su *modus operandi*, así como la prevalencia de la estafa en la red social. Finalmente, los autores realizan algunas recomendaciones a las plataformas afectadas a fin de prevenir la estafa.

2. Contribuciones principales

Al analizar los datos recolectados, los autores pudieron examinar la forma en la que empezaba la interacción. Las principales formas de interacción eran las respuestas y las *citas*, ya que permiten enviar un mensaje con el que continuar la estafa. Los autores también observaron otras formas de interacción como los *likes*, los *reposts*, y los *follows*, los cuales los autores presuponen que son usados para dar mayor veracidad a las interacciones.

También se analizaron las fotos de perfil de los perfiles, observando que los estafadores tendían a usar imágenes de NFTs, logos de plataformas de intercambio de criptomonedas, o personas reales.

Los atacantes incluían en los mensajes de respuesta links a otras redes de mensajería, como Instagram, Telegram, o WhatsApp, y también direcciones de correo electrónico. Los perfiles se hacían pasar o bien por soporte técnico de las plataformas de criptomonedas, o por personas reales.

Después de que la víctima 'mordiera el anzuelo', se producían una de dos tipo de estafas.

El primer tipo de estafa consiste en engañabar a la víctima para obtener la clave privada de su cartera, ya sea a través de mensajería, o un formulario. Una vez obtenida la clave, los fondos de la cartera eran transferidos a otra cartera. En algunos casos, al ver que los fondos de las carteras eran bajos, los estafadores pedían a la víctima que incluyera más dinero en la cartera. Los autores también analizaron el movimiento del dinero

en las carteras, observando que el dinero acababa en unas pocas carteras finales. También se observó que entre esas carteras de estafadores, sólo alrededor de un tercio recibieron pagos. Finalmente, los autores encontraron que, entre esas cartera se movían una cantidad de criptomonedas por valor de alrededor del millón de dólares.

El segundo tipo se enfocaba en pedir dinero por adelantado por servicios de asistencia técnica, los cuales nunca acababan siendo realizados. Éstos se hacian mediante criptomonedas o mediante la plataforma PayPal, y rondaban entre los $150 y $2,550.

Los autores también analizaron cómo reaccionaban las plataformas a esta estafa. Twitter era acababa bloqueando a la mayoría de estos perfiles, aunque estos perfiles acababan durando unos 90 días de media. Los autores contactaron directamente con PayPal para obtener información de las cuentas de los estafadores, y pudieron comprobar que la amplia mayoría de ellas ya habían sido detectadas y restringidas por la plataforma. Por otro lado, Gmail había bloqueado las cuentas de los estafadores, mientras que las plataformas de formularios apenas habían tumbado algunos de los distintos formularios usados para la estafa.

3. Opinión crítica

Hay varios puntos del artículo que me gustaría comentar.

Lo primero, es posible que la información que provee el artículo esté algo desactualizada. Los datos fueron recopilados entre octubre de 2022 y enero de 2023, y el artículo fue publicado en enero de 2024. Entre la recolección de los datos y la publicación del artículo ha pasado suficiente tiempo como para que cambie la metodología de la estafa, o la prevalencia de ella. Es común que los propios usuarios y propietarios de la plataforma se den cuenta de estas estafas y luchen contra ellas, haciéndola más difícil.

También cabe notar los grandes cambios que ha sufrido la plataforma de Twitter desde la adquisición por parte de Elon Musk a finales de 2022, los cuales empezaron a verse a mediados de 2023. Estos cambios afectaron profundamente a la plataforma, y probablemente hayan hecho lo mismo para la estafa. El artículo menciona el cambio sufrido en las cuentas verificadas, las cuales han excluído del estudio, y que ahora se pueden simplemente comprar. Es bien sabido que muchos perfiles de *bots* o de estafadores pagan por la verificación para destacar y parecer más reales.

Por último, hubiera sido interesante estudiar los orígenes de la estafa, ya que no se menciona en el estudio cuándo pudo aparecer, o cómo se propagó en el tiempo.

Otro punto interesante para el análisis sería un análisis más profundo de los perfiles de los estafadores. Mientras que el artículo se centra exclusivamente en las respuestas recibidas por los perfiles creados por los autores, hay mucha más información que se puede extraer de estos perfiles. Dado que las respuestas y las citas son públicas, se podría haber recabado mucha más información de cómo atacan a distintas víctimas. ¿Las respuestas

son siempre las mismas, o usan algún tipo de patrones como las usadas por los *honey pots*? ¿Es posible que usaran IA generativa? Y, ¿qué tienen en común las publicaciones de las víctimas? ¿Por qué palabras clave, o *hastags* podrían estar filtrando los estafadores?

Las recomendaciones ofrecidas por los autores a las plataformas también son algo vagas, o algo ingenuas. Por ejemplo, al recomendar que Gmail muestre una notificación de seguridad al mencionar claves privadas; la cantidad de falsos positivos que esto causaría sería molesto para los usuarios, aparte de que implicaría que Google monitorizara los contenidos de los correos electrónicos privados entre distintas cuentas.

Por último, también habría sido interesante investigar algo más a fondo a los individuos que llevan a cabo estas estafas. ¿Forman parte de alguna organización? ¿Son personas aisladas? Los autores recogieron información de localización de los perfiles, pero esta información estaba basada en la que los propios perfiles (falsos) daban, no en la localización real.

4. Opinión personal

Éste tipo de estafas han existido desde la antigüedad. Las redes sociales simplemente han hecho la comunicación más fácil, lo que ha aumentado también la facilidad de hacerlas. Y se han adaptado a los tiempos y las modas actuales.

Pero también cabe destacar que las criptomonedas facilitan este tipo de estafas, por varios motivos. El primero es el de siempre, la promesa del dinero fácil, el hacerte rico sin esfuerzo. No entraré a valorar si las criptomonedas son la nueva filatelia, pero es cierto que este tipo de 'negocios' atraen tanto a gente sin apenas conocimientos sobre el tema como a gente dispuesta a estafar a otras personas. También, el hecho de que sea un tipo de moneda descentralizado, donde sea prácticamente imposible 'deshacer' una transacción, o controlar este tipo de estafas, lo hace aún más sencillo. Aunque se hable de 'robo', las transacciones que se realizan desde las carteras de las víctimas hacia las de los estafadores son 'legales'. Lo ilícito es el robo de la clave, pero aún así, hay poca legislación, y poco control.

Y las plataformas también tienen muchas dificultades para controlarlo. Con miles de millones de usuarios, y millones de publicaciones, mensajes, y respuestas por segundo, es como buscar microbios dentro del millones de pajares. Especialmente Twitter, que llevaba años lidiando con el problema de los *bots*, y que ahora se dispara en su propio pie con decisiones erráticas, cada vez menos empleados, y peor moderación.

Por último, la entrada de la IA generativa, capaz de hacerse pasar por seres humanos convincentes, hace que la moderación sea aún más difícil, y que el coste de la estafa sea aún más barato. En las siguientes décadas habrá que discutir qué posibles soluciones puede tener este problema o, al menos, qué medidas podemos llegar a tomar para disminuírlo.