ANDORRA DIGITAL

ANC-AD
Agència Nacional
de Ciberseguretat d'Andorra

Informe de Ciberintel·ligència

L'amenaça dels ciberatacs de suplantació d'identitat mitjançant hipertrucatges

TLP:WHITE

MAIG 2024

FITXA DEL DOCUMENT

Versió	Redactat/Revisat per	Aprovat per	Data aprovació	Data publicació
1.0	ANC-AD	ANC-AD	20/05/2024	22/05/2024

Registre de canvis			
Versió	Pàgines	Data Modificació	Motiu del canvi

Propietari del document	ANC-AD

ÍNDEX

1. METODOLOGIA **4**

2. INTRODUCCIÓ **5**

3. EVOLUCIÓ DE LA SUPLANTACIÓ DE LA IDENTITAT: DEL MAN-IN-THE-MIDDLE A L'HIPERTRUCATGE **6**

3.1. Tipus d'atacs clàssics de suplantació de la identitat 6

3.2. Nous atacs a la suplantació mitjançant els LLM i els hipertrucatges 7

4. L'IMPACTE DE LA IA A LES TÈCNIQUES DE SUPLANTACIÓ DE LA IDENTITAT **9**

5. ELS SERVEIS D'IA EN TEMPS REAL, UNA AMENAÇA EMERGENT PER A LES EMPRESES **10**

5.1. El desafiament de l'atac «frau del CEO» amb ús d'hipertrucatges 10

5.2. Casos d'estudi d'estafes a empreses mitjançant l'ús de la IA 11

6. CONSELLS I MESURES DE SEGURETAT PER IDENTIFICAR I NO CAURE EN UNA ESTAFA DE SUPLANTACIÓ DE LA IDENTITAT MITJANÇANT L'ÚS D'HIPERTRUCATGES **13**

7. CLÀUSULA DE CONFIDENCIALITAT **14**

1. METODOLOGIA

Aquest informe aplica els principis de Traffic Light Protocol (TLP). És un esquema creat per fomentar un intercanvi més bo d'informació delicada (però no classificada) en l'àmbit de la seguretat de la informació.

A través d'aquest esquema, d'una manera àgil i senzilla, s'indica fins on pot circular la informació més enllà del receptor immediat, i aquest ha de consultar l'Agència Nacional de Ciberseguretat d'Andorra quan cal distribuir la informació a tercers.

Codi	Com es fa servir	Com es comparteix
TLP: RED	S'ha de fer servir **TLP:RED** quan la informació està limitada a persones concretes, i podria tenir impacte en la privacitat, la reputació o les operacions si es fa servir malament.	Els receptors no han de compartir informació designada com a **TLP:RED** amb cap tercer fora de l'àmbit on va ser exposada originalment.
TLP: AMBER	S'ha de fer servir **TLP:AMBER** quan la informació ha de ser distribuïda de manera limitada, però suposa un risc per a la privacitat, la reputació o les operacions si és compartida fora de l'organització.	Els receptors poden compartir informació indicada com a **TLP:AMBER** només amb membres de la seva pròpia organització que necessiten conèixer-la, i amb clients, proveïdors o associats que necessiten conèixer-la per protegir-se a si mateixos o evitar danys. L'emissor pot especificar restriccions addicionals per compartir aquesta informació.
TLP: GREEN	S'ha de fer servir **TLP:GREEN** quan la informació és útil per a totes les organitzacions que hi participen, com també amb tercers de la comunitat o el sector.	Els receptors poden compartir la informació indicada com a **TLP:GREEN** amb organitzacions afiliades o membres del mateix sector, però mai a través de canals públics.
TLP: WHITE	S'ha de fer servir **TLP:WHITE** quan la informació no suposa cap risc de mal ús, conforme a les regles i procediments establerts per a la seva difusió pública.	La informació **TLP:WHITE** pot ser distribuïda sense restriccions, únicament subjecta a controls de copyright.

2. INTRODUCCIÓ

Probablement, la intel·ligència artificial sigui la fita tecnològica dels últims anys que està tenint més impacte en qualsevol àmbit de la vida. Els beneficis i les oportunitats que ha aportat han suposat tota una revolució i no generen cap dubte. Tanmateix, també hi ha qui no ha deixat escapar l'oportunitat de fer servir tot el seu potencial amb finalitats poc o gens ètiques. Per tant, en contrapartida, **la IA ha possibilitat el sorgiment de noves ciberamenaces, que es caracteritzen per una sofisticació de la qual no existeixen precedents.**

Des que hi ha Internet, la suplantació d'identitat ha estat una de les tècniques més emprades pels ciberdelinqüents per dur a terme els seus atacs. Els objectius poden ser diversos, tot i que a la majoria dels casos les motivacions solen ser econòmiques o el robatori de dades confidencials. I sí, la IA s'ha convertit en una eina poderosa per als ciberdelinqüents, que els permet personalitzar i perfeccionar els seus atacs de manera més eficient i efectiva que mai.

La IA ha permès que els cibercriminals tinguin la capacitat de crear i distribuir correus electrònics i missatges de textos falsos de manera massiva, i adaptar-los a les característiques específiques de cada víctima potencial. A més a més, amb els algoritmes d'aprenentatge automàtic també tenen la capacitat d'analitzar el comportament dels seus objectius i generar **atacs cada vegada més complexos i convincents, i alhora fan summament difícil discernir entre què és legítim i què és fraudulent**.

I, tot i que el que s'acaba d'esmentar ja suposa un desafiament enorme per a usuaris comuns, empreses i professionals de la ciberseguretat, els ciberdelinqüents estan explotant una altra aplicació de la IA que els atorga capacitats noves i millorades per aconseguir que els seus atacs tinguin una altíssima taxa d'èxit: poder utilitzar la imatge o la veu de qualsevol persona. **Els hipertrucatges han provocat que el concepte tradicional de suplantació d'identitat hagi adquirit una dimensió nova i s'hagi convertit en una amenaça emergent**.

Aquest informe busca explorar en profunditat l'impacte que la IA ha tingut en l'evolució dels ciberatacs de suplantació de la identitat, i per això s'analitzaran les diverses maneres en les quals estan fent servir la IA per millorar l'efectivitat dels seus atacs, com també les implicacions ètiques i de seguretat que sorgeixen d'aquesta intersecció entre tecnologia i delicte cibernètic.

A més a més, examinarem les estratègies i les eines que les organitzacions i els usuaris poden emprar per mitigar els riscos associats amb aquests atacs, i proposarem possibles vies per abordar aquest desafiament en el futur.

3. EVOLUCIÓ DE LA SUPLANTACIÓ DE LA IDENTITAT: DEL MAN-IN-THE-MIDDLE A L'HIPERTRUCATGE

Les tècniques fetes servir per dur a terme ciberatacs de suplantació de la identitat han experimentat una evolució notable des dels seus inicis, **on els atacants se centraven primordialment a suplantar paràmetres tècnics, com ara les adreces IP o MAC**, amb l'objectiu d'enganyar els sistemes informàtics fent-los creure que les comunicacions procedien o s'enviaven d'una font legítima quan, en realitat, no era així. Aquest tipus d'atacs es van estendre ràpidament, **i també van abastar la suplantació de correus electrònics, números de telèfons o targetes SIM**.

Amb l'avenç de la tecnologia i la interconnexió creixent dels dispositius, **els ciberdelinqüents han perfeccionat les seves tècniques, i han adoptat estratègies cada cop més sofisticades per manipular i enganyar les seves víctimes**. Actualment, els atacs de suplantació de la identitat han assolit nivells de sofisticació impressionants, gràcies a l'ús d'eines d'intel·ligència artificial (IA).

Un dels aspectes més destacats d'aquesta evolució és el sorgiment dels hipertrucatges, una aplicació de la IA que permet crear imatges, vídeos i àudios falsos, indistingibles dels originals, mitjançant el reemplaçament de cares i veus.

3.1. Tipus d'atacs clàssics de suplantació de la identitat

Tot seguit, s'exposen els tipus d'atacs de suplantació de la identitat que han fet servir els ciberdelinqüents tradicionalment:

- **Suplantació (*Spoofing*) d'IP, MAC i DNS**: la falsificació d'adreces IP, MAC o registre DNS s'han utilitzat des dels inicis de la ciberdelinqüència per enganyar els sistemes informàtics i **redirigir el tràfic a servidors maliciosos controlats pels atacants**. Permet, entre altres coses, **interceptar comunicacions**, fer atacs Man-in-the-Middle (MITM), o desviar els usuaris a llocs web falsos dissenyats per robar informació o infectar equips.

- **Suplantació de correu electrònic**: consisteix a manipular els encapçalaments dels correus electrònics per fer que sembli que han estat enviats des d'una adreça de correu electrònic legítima quan en realitat procedeixen d'una font maliciosa. Això es pot fer servir per **enganyar els destinataris i fer que confiïn en els correus electrònics falsos**, cosa que facilita l'execució d'**atacs de pesca** o altres activitats malicioses.

- **Pesca tradicional**: és una de les tècniques de suplantació de la identitat més comunes i que es fan servir més sovint, on s'empra el correu electrònic per aconseguir enganyar la víctima. L'objectiu és fer passar un correu fraudulent per un de legítim, **i aconseguir, d'aquesta manera, que l'usuari acabi fent alguna mena d'acció de la qual se'n beneficia l'atacant**.

- **Pesca dirigida (*Spear-phishing*)**: és una **variant més dirigida i personalitzada**. En aquest cas, els ciberdelinqüents investiguen a fons les seves víctimes potencials i personalitzen els correus electrònics maliciosos per tal que semblin més convincents. Aquesta tècnica pot implicar l'ús d'informació personal obtinguda de fonts públiques o violacions de dades anteriors, cosa que augmenta la probabilitat que la víctima caigui a la trampa.

- **Pesca grossa (*Whaling*)**: també conegut com a **pesca de perfil alt, se centra a atacar** individus d'alt nivell d'una organització, com ara **executius o directors**. Els ciberdelinqüents aprofiten l'autoritat i l'accés a la informació confidencial d'aquests individus per perpetrar atacs més sofisticats, com ara el robatori de dades corporatives o la transferència de fons fraudulents. **El frau del CEO és el cas de pesca grossa més popularitzat**.

- **Pesca per SMS (*Smishing*) i pesca per veu (*Vishing*)**: aquestes variants es caracteritzen per **l'ús de missatges de text o trucades telefòniques** en lloc de correus electrònics per dur a terme l'engany. Els ciberdelinqüents poden enviar missatges de text falsificats que contenen enllaços maliciosos o fer trucades telefòniques automatitzades per enganyar les víctimes i obtenir informació confidencial.

3.2. Nous atacs a la suplantació mitjançant els LLM i els hipertrucatges

Actualment, no es pot parlar de tècniques de suplantació de la identitat sense fer referència a la intel·ligència artificial i la seva aplicació de la generació de textos, mitjançant els models de llenguatge LLM, i dels hipertrucatges.

Els models de llenguatge LLM, entre els quals destaca el famós ChatGPT, han guanyat una rellevància especial en l'elaboració de continguts amb finalitats malicioses. De fet, segons l'informe de l'empresa SlashNext, publicat a finals del 2023 passat, els ciberatacs de pesca, pesca dirigida i les campanyes BEC havien experimentat un augment increïble de 1.265 % des de l'aparició de la famosa aplicació d'OpenAI.

Els hipertrucatges són els altres grans protagonistes en l'àmbit de la suplantació de la identitat. El terme «Deepfake» (hipertrucatge) fusiona «fake» (fals) amb «deep» de «deep learning», una aplicació de la IA que empra algoritmes avançats capaços de discernir entre contingut real i manipulat, i millorar d'aquesta manera la precisió de la falsificació sense que calgui la intervenció humana directa, i fa referència a vídeos, imatges o àudios alterats amb el propòsit d'aparentar autenticitat i veracitat, i induir a l'engany.

Es categoritzen en dues tipologies entre les quals **es diferencien les *deepfaces***, on allò que se suplanta és el rostre o l'aparença física d'una persona, i les ***deepvoices***, en el qual allò que es suplanta és la veu.

D'aquesta manera**, textos, cares i veus creats mitjançant IA s'han convertit en eines que els ciberdelinqüents estan fent servir** amb finalitats diferents. Han servit **per desestabilitzar governs i manipular processos electorals** mitjançant campanyes de notícies falses, també s'han fet servir per **generar contingut pornogràfic** amb cares de personalitats famoses i, fins i tot, de ciutadans normals i corrents. A més a més, han tingut rellevància especial **com ara la**

tècnica d'atac d'enginyeria social mitjançant la suplantació de la identitat, que és l'ús que s'analitza en profunditat en aquest informe.

I quan es parla d'hipertrucatges i suplantació de la identitat és obligatori abordar els riscos que es desprenen dels **ciberatacs contra les empreses**, entre els quals caldria destacar les campanyes de:

- **Pesca, pesca dirigida i BEC** mitjançant l'ús dels LLM.
- **Pesca per veu** amb l'ús de *deepvoice*.
- **Videotrucades** amb l'ús de *deepfaces*.

4. L'IMPACTE DE LA IA A LES TÈCNIQUES DE SUPLANTACIÓ DE LA IDENTITAT

Per entendre la relació entre intel·ligència artificial i suplantació de la identitat, cal abordar **els hipertrucatges i les xarxes neuronals de la IA mitjançant els quals es creen, conegudes com les GAN** (Generative Adversarial Networks o Xarxes generatives adversàries, en català).

Les GAN són una mena d'algoritme d'aprenentatge profund que consta de dues xarxes neuronals enfrontades entre si: un generador i un discriminador. El generador s'encarrega de crear dades noves, que, en aquests cas, són rostres sintètics, mentre que el discriminador té la tasca de distingir entre dades reals i sintètiques. Aquestes dues xarxes treballen en conjunt de manera competitiva: el generador intenta produir rostres que siguin prou realistes per enganyar el discriminador, mentre que el discriminador busca millorar la seva capacitat per distingit entre rostres reals i sintètics.

Gràcies a aquest procés de retroalimentació constant, **les GAN poden generar rostres sintètics que no existeixen a la vida real**, però que semblen autèntiques a simple vista. O, al contrari, es poden fer servir per **suplantar identitats de manera efectiva mitjançant la generació d'un rostre sintètic a partir de la imatge d'una persona real**.

L'ús de les GAN per a la creació de rostres sintètics té diverses aplicacions, des de la generació de contingut per a pel·lícules i videojocs, fins a la creació de perfils falsos a les xarxes socials o la manipulació d'imatges o vídeos amb finalitats malicioses o delictives.

Aquest avenç tecnològic planteja desafiaments ètics i de seguretat importants, atès que **la seva capacitat de generar identitats falses amb gran precisió fa que sigui més difícil verificar l'autenticitat de la informació** i detectar possibles fraus.

5. ELS SERVEIS D'IA EN TEMPS REAL, UNA AMENAÇA EMERGENT PER A LES EMPRESES

Els delinqüents cibernètics han trobat en els hipertrucatges i els models de llenguatge LLM unes eines que els permeten escometre atacs amb un grau de sofisticació mai vist.

Si a això se li afegeix el fet que **cada vegada hi ha més programaris i eines que no exigeixen excessius coneixements tecnològics o informàtics per part dels usuaris per generar hipertrucatges**, el desafiament que representa l'accessibilitat fàcil el converteix en un perill molt més gran.

És important tenir en compte que tradicionalment les tècniques de suplantació de la identitat més sofisticades es basaven en una recopilació d'informació exhaustiva de la víctima, per fer l'atac més personalitzat i dirigit. Ara, a més a més, pot ser la mateixa veu o la cara d'un alt representant d'una empresa la que demani a un empleat que escometi una acció determinada.

Cal destacar que **tant els textos creats pels LLM, com les *deepvoices* i les *deepfaces* es poden generar al mateix moment que es produeix l'atac, cosa que els converteix en unes armes versàtils i adaptables a múltiples circumstàncies.**

5.1. El desafiament de l'atac «frau del CEO» amb ús d'hipertrucatges

El frau del CEO és una variant de la pesca amb la qual moltes empreses estan familiaritzades. **Aquest atac s'adreça a algun empleat o directiu amb capacitat per prendre decisions i executar operacions de caire econòmic.**

D'aquesta manera, mitjançant correu electrònic o trucada telefònica, l'atacant es fa passar per un superior jeràrquic (CEO, CFO o homòlegs) i insta la víctima a fer una transferència bancària o qualsevol altra mena de pagament, i li fa creure que és una operació legítima.

Tradicionalment, i davant dels possibles dubtes que pogués generar l'acció a la víctima, l'atacant s'assegurava de disposar de la informació suficient (factures, operacions de l'entitat, dinàmiques internes, noms d'empleats i responsables o superiors) que poguessin dotar de veracitat la seva petició.

Però els hipertrucatges han transformat per complet aquesta mena d'atac, i fan que sigui molt més complex poder detectar una acció il·legítima. Els estafadors ara, a més de la informació que siguin capaços de recopilar i fer servir durant l'atac, **poden exercir una pressió sense precedents sobre la víctima, i convèncer-la amb una trucada o videotrucada fent servir la veu o la cara del CEO real.**

5.2. Casos d'estudi d'estafes a empreses mitjançant l'ús de la IA

Estafa de 4 milions d'euros mitjançant el frau del CEO amb *deepvoices* a l'Empresa Municipal de Transports de València

L'Empresa Municipal de Transports de València es va convertir en notícia l'any 2019 perquè va ser objecte d'aquesta mena d'atac. Però no es tractava d'un ciberatac més de suplantació de la identitat.

En aquell cas, a més de suplantar identitats a través de correus electrònics, també es van fer servir *deepvoices* a les trucades telefòniques.

Finalment, van aconseguir que la directora d'administració de l'EMT ordenés fins a vuit transferències per un total de 4 milions d'euros, amb el pretext d'una suposada adquisició a la Xina.

Estafa de 220.000 euros mitjançant el frau del CEO amb *deepvoices* a una empresa energètica del Regne Unit

El 30 d'agost de 2019 el Wall Street Journal es va fer eco d'un cas que va titular aleshores com un ciberatac inusual.

Uns mesos abans, al març, una empresa havia denunciat a les autoritats que havia fet una transferència bancària a conseqüència d'una estafa.

Tot i que no se'n van fer públics els detalls, sí que es va donar visibilitat a l'ús d'allò que aleshores era una tècnica innovadora: l'ús de la suplantació de la veu del CEO.

Estafa de 9,7 milions d'euros mitjançant el frau del CEO amb *deepvoices* a una empresa farmacèutica del Regne Unit

Un any més tard, el 2020, durant la crisi del coronavirus, la farmacèutica Zendal va ser víctima d'un altre frau del CEO. En aquesta ocasió les pèrdues provocades van ascendir als 9,7 milions d'euros.

El modus operandi dels delinqüents va ser similar: van suplantar la identitat del CEO per instruir un executiu financer per tal que fes transferències en nom de l'empresa, i, a més, es van fer passar per professionals de KPMG per proporcionar ordres de pagament i factures falses que justificaven les transaccions.

Estafa de 23,7 milions d'euros mitjançant el frau del CEO amb *deepface* en una videotrucada a una empresa a Hong Kong.

Van enganyar un empleat que va fer una transferència creient que anava adreçada a la filial de la seva empresa al Regne Unit. Els atacants van fer servir tecnologies d'hipertrucatge per

suplantar la identitat del CFO i altres directius durant la videoconferència en la qual instaven el treballador a dur a terme l'operació econòmica.

6. CONSELLS I MESURES DE SEGURETAT PER IDENTIFICAR I NO CAURE EN UNA ESTAFA DE SUPLANTACIÓ DE LA IDENTITAT MITJANÇANT L'ÚS D'HIPERTRUCATGES

La proliferació de tecnologies d'hipertrucatge planteja un desafiament significatiu per a la seguretat de les empreses i les organitzacions, i augmenta el risc de suplantació de la identitat i de fraus. Per abordar aquesta amenaça, és crucial implementar mesures de seguretat proactives i promoure la conscienciació entre el personal. Per a això, les organitzacions que vulguin fer front a aquesta amenaça emergent haurien de valorar les recomanacions i bones pràctiques següents:

- **Conscienciació i formació del personal** per identificar els hipertrucatges i la comprensió dels riscos associats és fonamental. El principi de «conèixer el seu client» (KYC) s'ha d'estendre a la identificació de possibles hipertrucatges en qualsevol interacció digital.

- **Adopció de tecnologia antifalsificació**: els algoritmes criptogràfics per a la inserció de valors *hash* a vídeos, pot ajudar a detectar manipulacions. La utilització de la intel·ligència artificial (IA) i la cadena de blocs (*blockchain*) per registrar les empremtes digitals a prova de manipulacions també pot ser eficaç.

- **Implementació de programes de detecció d'artefactes digitals**: els programes dissenyats per inserir artefactes digitals en vídeos poden dificultar la creació d'hipertrucatges d'alta qualitat, i reduir d'aquesta manera el risc d'èxit dels intents d'estafes mitjançant la suplantació de la identitat.

- **Polítiques de seguretat Zero-Trust**: és imprescindible disposar de procediments i protocols d'actuació per definir com es duen a terme les dinàmiques de qualsevol àrea d'una organització. Per altra banda, l'automatització de controls, com els que es poden dissenyar per a casuístiques especialment delicades, com els processos financers, també poden ser de gran ajuda per prevenir fraus. Tot i que també seria recomanable complementar tot el que s'acaba d'esmentar fins ara amb:

    - Educar els empleats i familiars sobre la detecció dels hipertrucatges.

    - Promoure l'ús de fonts d'informació confiables.

    - Adoptar una actitud escèptica i verificar l'autenticitat de les comunicacions, especialment aquelles que impliquen transaccions financeres.

En resum, la combinació de conscienciació, tecnologia avançada i procediments de seguretat sòlids és fonamental per mitigar el risc de suplantació de la identitat mitjançant hipertrucatges en entorns empresarials i organitzatius.

7. CLÀUSULA DE CONFIDENCIALITAT

Aquest document és propietat de l'Agència Nacional de Ciberseguretat d'Andorra. Tota la informació que conté és confidencial, aquesta informació s'actualitzarà regularment per reflectir els possibles canvis dels productes i no podrà ser copiada o revelada a terceres persones sigui totalment o en part, sense consentiment previ exprés de l'Agència Nacional de Ciberseguretat d'Andorra.