Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **published version** of the bachelor thesis:

Milà Solé, Quim; Borrego Iglesias, Carlos, dir. És el nostre temps entre tasques una dada vulnerable en termes de privacitat?. 2024. (Enginyeria Informàtica)

This version is available at https://ddd.uab.cat/record/290090

under the terms of the BY-NC-ND license

És el nostre temps entre tasques una dada vulnerable en termes de privacitat?

Quim Milà Solé

8 de febrer de 2024

Resum– En aquest treball s'ha investigat la relació entre el nostre temps entre tasques i el tràfic de xarxa. Concretament, s'ha buscat comprendre si els Proveïdors de Serveis d'Internet, vigilant la nostra activitat en línia, podrien predir els nostres temps entre tasques. S'ha dissenyat i desenvolupat un sistema capaç de generar els bancs de dades semblant dels quals disposaria un Proveïdor de Serveis d'Internet. Posteriorment, aquest sistema gestiona el format de les dades generades per a la preparació eficient per a l'entrenament d'una Xarxa Neuronal Recurrent (RNN). Amb aquest enfocament, la nostra missió ha estat aprofundir en una possible utilitat afegida de les nostres dades, en aquest cas la nostra activitat de xarxa.

Paraules clau– Tasca, Traces, Dataset, Privacitat, Xarxa Neuronal, Proveïdor de Serveis d'Internet

Abstract– In this work, the relationship between our task-switching intervals and network traffic has been investigated. Specifically, the aim was to understand if Internet Service Providers, monitoring our online activity, could predict our task-switching intervals. A system capable of generating dataset similar to those available to an Internet Service Provider has been designed and developed. Subsequently, this system manages the data format generated for efficient preparation for training a Recurrent Neural Network (RNN). With this approach, our mission has been to delve into a potential additional utility of our data, in this case, our network activity.

Keywords– Task, Traces, Dataset, Privacy, Neuronal Networks, Internet Service Provider

1 Introducció

En el món digital actual, les nostres dades personals, s'han convertit en una mena de tresor digital de gran valor. Cada clic que fem, cada cerca que realitzem i cada acció que prenem en línia es converteix en una petita joia que les empreses anhelen. Moltes d'aquestes empreses prometen oferir-nos els millors productes i serveis personalitzats, però sovint ho fan sense tenir en compte la nostra privacitat.

Immersos en aquest món digital, ens trobem sotmesos a un flux constant d'estímuls i interrupcions. La nostra atenció, un bé cada cop més preciós, està sotmesa a un assalt constant de notificacions, correus electrònics i distraccions virtuals. Aquestes interrupcions poden tenir un impacte devastador en la nostra concentració, fragmentant el nostre pensament i afectant la nostra productivitat. La investigació científica ens revela que saltar de tasca en tasca, una pràctica cada cop més comuna, no només allarga el temps necessari per acabar les tasques, sinó que també augmenta els errors i limita la nostra creativitat i capacitat de recordar.

Aquest treball de final de grau pren consciència dels desafiaments en el món digital actual i se centra a comprendre a fons la relació entre les interrupcions digitals i la privacitat de l'usuari. L'objectiu principal és explorar si els proveïdors de xarxa poden posar en perill la privacitat dels temps entre tasques dels usuaris a través de la monitoratge de la seva activitat de xarxa.

Per assolir aquest objectiu, s'ha desenvolupat un sistema amb la capacitat d'analitzar amb detall els intervals entre tasques de l'usuari i de generar les dades de tràfic de xarxa a les quals el nostre Proveïdor de Serveis té accés. Aquest enfocament integral permet identificar patrons i possibles amenaces a la privacitat. El processament eficient de les dades ha estat assegurat mitjançant un mòdul de gestió de

• E-mail de contacte: quim.mila@autonoma.cat
• Menció realitzada: Tecnologies de la Informació
• Treball tutoritzat per: Carlos Borrego Iglesias (Departament de la Enginyeria de la Informació i Comunicació)
• Curs 2023/24

dades, preparant-les per a l'entrenament d'una xarxa neuronal recurrent (RNN), que preveurà els temps entre tasques dels usuaris.

El document està estructurat de la següent manera: en la secció 2 trobem estat de l'art d'aquest camp d'investigació. En la secció 3 els objectius del treball. En la secció 4 la metodologia que s'ha seguit pel desenvolupament del projecte. A partir de la secció 5 es detalla el disseny del sistema, el desenvolupament dels mòduls i el funcionament de la xarxa neuronal en les seccions posteriors. Els resultats i les conclusions es troben a partir de la secció 10, en la secció final del document, acompanyades dels agraïments i la bibliografia.

2 Estat de l'art

L'estat de l'art d'un projecte fa referència al punt més avançat i contemporani de coneixement, tecnologia o recerca en una àrea específica relacionada amb el projecte. Constitueix una revisió exhaustiva i actualitzada dels progressos, descobriments, mètodes i pràctiques rellevants en el domini temàtic en el qual es desenvolupa el projecte. S'han trobat diferents estudis que mesuren l'atenció relacionant, també, les interrupcions i com poden afectar a la concentració durant l'aprenentatge o durant el temps laboral.

En la literatura científica s'ha parlat enormement de privacitat i comportament en entorns informàtics. Concretament, en l'estudi *Time to focus* [1] examina com proporcionar una retroalimentació sobre la durada de les interrupcions influeix en el comportament d'auto interrupció en tasques informàtiques. S'observa que les interrupcions físiques es posposen, mentre que les digitals es gestionen immediatament. L'eina que es desenvolupa té l'objectiu de mostrar la durada de les interrupcions per augmentar la consciència del temps dedicat a elles.

Amb un objectiu similar, una proposta per aquest problema és un altre estudi [2] que ens parla de les estratègies per reduir aquestes possibles interrupcions, com ara auriculars o una porta d'oficina tancada. En aquesta recerca s'ha desenvolupat una eina que combina una llum LED similar a un semàfor físic amb una mesura automàtica de la interruptibilitat basada en dades d'interacció amb l'ordinador.

Fins al moment, no s'ha realitzat cap estudi exhaustiu que mesuri de manera precisa la vulnerabilitat de les dades relatives als temps d'interrupcions, tenint en compte les dades que inadvertidament compartim amb els proveïdors de serveis de xarxa. La necessitat d'un enfocament més crític i analític en aquest àmbit és evident, ja que la nostra dependència creixent dels serveis digitals requereix una comprensió més profunda dels riscos associats a la transmissió i emmagatzematge d'aquestes dades, particularment en relació amb els temps d'interrupcions.

3 Objectius

En aquesta secció, s'aborden els objectius fonamentals del projecte, destacant el que s'espera aconseguir i de quina manera es durà a terme. Com s'ha esmentat a la Secció d'Introducció, l'objectiu principal d'aquest projecte consisteix a investigar si el temps entre tasques pot ser considerat una dada vulnerable en termes de privacitat.

És essencial contextualitzar el marc on es produeix la pèrdua de privacitat. En el món digital, els Proveïdors de Serveis d'Internet tenen accés a l'activitat de xarxa de tots els seus clients, que inclou les comunicacions i les interaccions amb recursos en línia. No obstant això, és crucial destacar que aquest accés no s'estén als processos interns dels sistemes operatius, salvaguardant així certa capa de privacitat associada a les accions realitzades directament en els dispositius.

Concretament, s'ha plantejat com a objectiu dissenyar i desenvolupar un sistema que coordinarà diferents mòduls amb objectius específics. En primer lloc, s'ha creat un mòdul específic dissenyat per complir l'objectiu de mesurar el temps entre tasques d'un usuari, amb una atenció especial a les interaccions amb el sistema operatiu Windows. Aquesta aplicació proporciona una visió detallada del comportament temporal de l'usuari, essencial per a la comprensió dels patrons d'activitat.

En segon lloc, s'ha desenvolupat un altre mòdul destinat a l'objectiu de capturar i analitzar, el tràfic de xarxa generat durant les interaccions digitals. Aquest mòdul ha estat instrumental per a la creació de conjunts de dades que, en conjunció amb l'anàlisi del temps entre tasques, permeten explorar i identificar relacions rellevants pel que fa a la privacitat de l'usuari.

En tercer lloc, s'ha desenvolupat un altre mòdul amb l'objectiu de gestionar els conjunts de dades generats i normalitzar els formats per al posterior entrenament de la xarxa neuronal.

Finalment, s'ha realitzat un altre mòdul amb l'objectiu d'oferir una primera versió d'una xarxa neuronal que seria l'encarregada de predir els temps entre tasques dels usuaris, mitjançant els bancs de dades generats prèviament.

A través d'aquest enfocament, el projecte no només busca proporcionar respostes a la pregunta central sobre la vulnerabilitat del temps entre tasques en termes de privacitat, sinó que també pretén contribuir a la comprensió més àmplia dels riscos i les dinàmiques de privacitat en l'entorn digital actual.

4 Metodologia

Per dur a terme aquest projecte, s'ha considerat que la millor metodologia possible és la Iterativa i Incremental [3]. El desenvolupament Iteratiu i Incremental és una metodologia en la qual el projecte es divideix en petites parts, i cada part és desenvolupada i millorada mitjançant iteracions successives. Aquest enfocament permet una adaptabilitat més gran a canvis durant el procés de desenvolupament i proporciona resultats parcialment funcionals en cada iteració. Les iteracions són cicles repetitius de planificació, desenvolupament, prova i avaluació, i cada iteració construeix sobre la base de les anteriors. Incremental significa que el projecte es desenvolupa en petites parts addicionals o increments successius, afegint funcionalitats noves o millorades en cada iteració. Aquesta metodologia facilita una millora contínua i una millor adaptació als canvis en els requisits del projecte. S'ha optat per aquesta metodologia, ja que el projecte implica la creació d'un sistema que comptarà amb diferents mòduls: un per a la mesura dels temps entre tasques, un altre per a l'anàlisi de l'activitat de xarxa, un altre per modificar les dades obtingudes i l'últim mòdul

que serà la xarxa neuronal que un cop entrenada s'encarregarà de predir els temps entre tasques. Separar el sistema en mòduls aconsegueix reduir la complexitat d'aquest, i una metodologia iterativa permet abordar-los en etapes més petites i manejables.

5 Disseny del sistema

En aquesta secció, es detalla la interacció dels mòduls desenvolupats durant el projecte i com es combinen per formar el sistema final. El sistema proporcionat a l'usuari final consta dels mòduls de "Canvi de Tasques" i "Captura de Tràfic de Xarxa". Aquests mòduls són responsables de generar els dos conjunts de dades fonamentals necessaris per al posterior entrenament de la xarxa neuronal. És important destacar que, un cop generats els conjunts de dades, el mòdul de "Gestió de Dades" realitzarà les modificacions pertinents per assegurar la coherència en el format d'ambdós conjunts de dades. I finalment, el mòdul "Xarxa Neuronal" serà l'encarregat de predir, si és possible, el temps entre tasques dels usuaris mitjançant el banc de dades d'activitat de xarxa. L'esquema de la figura següent il·lustra el funcionament global del sistema:

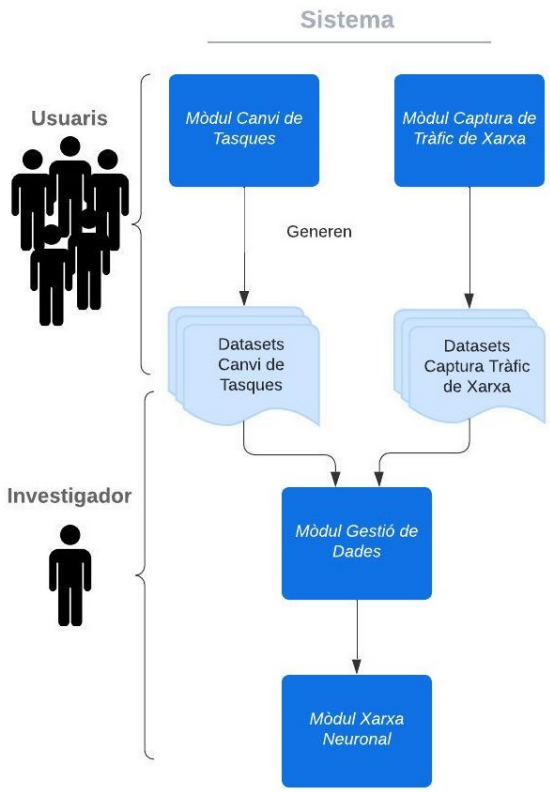

Fig. 1: Sistema

5.1 Definició Tasca i temps entre Tasques

Abans d'explicar com s'ha desenvolupat el mòdul és necessari definir que s'ha entès com a tasca a l'hora de desenvolupar el sistema. Generalment, s'utilitza el terme tasca per referir-nos a una feina o activitat específica que es realitza com a part d'una tasca encara més àmplia o com a part d'un procés. Per dur a terme aquesta investigació s'ha usat el terme tasca com a programa informàtic, entenent així que el canvi entre tasques és el canvi entre dos programes o dues finestres. Per tant, el temps entre tasques és el temps que prenem per canviar d'una aplicació a un altre.

5.2 Traces de Xarxa

Abans d'explicar com s'ha desenvolupat el mòdul és necessari definir que s'ha entès com a traça a l'hora de desenvolupar el sistema. En el context de les comunicacions de xarxa, el terme traça es refereix al rastre digital que registra la ruta i la interacció dels paquets de dades entre els diversos nodes de la xarxa. Aquesta activitat pot incloure paquets TCP[1]/UDP[2], així com altres tipus de trànsit de xarxa, com ara ICMP[3] o altres protocols. Una traça de xarxa és una seqüència temporal de paquets que es registren per a anàlisi o control.

5.3 Proveïdors de Serveis d'Internet i la nostra privacitat

Cal especificar, que els ISP[4] poden monitorar i registrar detalladament la nostra activitat de xarxa, incloent-hi les pàgines web que visitem, les aplicacions que utilitzem i altres comportaments en línia. Això crea una gran quantitat d'informació personal i pot revelar detalls íntims de la nostra vida en línia. Basant-se en les dades recollides, els ISP poden crear perfils d'usuari detallats. Aquests perfils poden incloure preferències, hàbits de navegació, interessos personals i informació delicada. La creació de perfils pot afectar la privacitat i ser emprada amb fins de segmentació publicitària. Fins i tot, en alguns casos, els ISP poden vendre o compartir les dades dels seus usuaris amb tercers, com ara empreses de màrqueting o altres societats comercials. Totes aquestes accions comporten una pèrdua de privacitat considerable respecte les nostres dades.

5.4 Com es relacionen els canvis de tasques i les aplicacions

Les aplicacions en un dispositiu digital són la peça fonamental que impulsa les diverses tasques realitzades pels usuaris. Cada aplicació, ja sigui un navegador web, una aplicació de missatgeria o una plataforma de mitjans socials, està dissenyada per oferir funcionalitats específiques que satisfacin les necessitats de l'usuari. El concepte de task switching juga un paper crucial en aquest escenari, ja que els usuaris freqüentment canvien entre diferents aplicacions per acomplir les seves tasques diàries. A mesura que l'usuari alterna entre aplicacions, commuta de manera

[1] TCP: significa "Protocol de Control de Transmissió", un protocol fonamental de la capa de transport a la xarxa TCP/IP. Aquest protocol assegura una comunicació fiable i ordenada entre dispositius a través de la xarxa, essencial per a aplicacions com connexions web i transferència de fitxers.

[2] UDP: correspon a "Protocol de Datagrama d'Usuari" en català. És un protocol de la capa de transport a l'arquitectura de xarxa TCP/IP. A diferència del TCP, l'UDP no ofereix una comunicació fiable ni orientada a la connexió.

[3] ICMP: ICMP correspon a "Protocol de Missatges de Control d'Internet" en català. Aquest protocol és part de la capa de xarxa del model TCP/IP i s'utilitza principalment per enviar missatges d'error i control associats a la comunicació de dades a través d'Internet.

[4] ISP: L'acrònim prové de l'anglès, Internet Service Provider, i la traducció al català és Proveïdor de Serveis d'Internet.

fluïda entre diferents funcionalitats i tasques, ja sigui llegir correus electrònics, navegar per internet o interactuar en xarxes socials.

5.5 Relació aplicacions i traces

Les aplicacions a Internet estableixen connexions a través d'*endpoint* identificats per adreces IP i ports específics. Aquestes connexions generen traces de xarxa, un rastre digital que registra la ruta i la interacció dels paquets de dades entre els diversos nodes de la xarxa. Els ISP són essencials per a aquest procés, ja que faciliten la infraestructura i els recursos necessaris per al funcionament de les aplicacions i les comunicacions en línia, gestionant el trànsit de dades i assegurant la connectivitat eficient entre els diversos *endpoints*.

6 MÒDUL "CANVI DE TASQUES"

El mòdul "Canvi de Tasques" té com a objectiu principal registrar tots els canvis de finestra de l'usuari en una màquina amb sistema operatiu Windows. S'ha optat per crear el sistema específicament per a aquest entorn, atès que és àmpliament utilitzat per aquells que fan servir els seus ordinadors per tasques de productivitat i utilitat quotidiana. Aquesta elecció es fonamenta en la popularitat general d'aquest sistema operatiu en àmbits com l'ofimàtica, sense abordar consideracions més tècniques relacionades amb el desenvolupament, disseny o multimèdia.

6.1 API Utilitzada i funcions específiques

Primerament, s'ha determinat que la millor manera d'abordar aquest problema és utilitzar l'API de Windows [4] i la biblioteca de temps de C [5]. Aquesta elecció ofereix una solució robusta per enregistrar els esdeveniments de canvi d'enfocament de finestra o aplicació en el sistema operatiu, o també dit, els canvis de tasca.

El punt central de la implementació és la funció *WinEventProc*, que actua com un callback i que s'executa quan es produeix un esdeveniment de sistema específic, concretament l'esdeveniment EVENT_SYSTEM_FOREGROUND, que indica un canvi de finestra activa en el sistema.

```
void CALLBACK WinEventProc(
HWINEVENTHOOK hHook, DWORD event,
LONG idObject, DWORD dwmsEventTime)
```

En aquest punt, s'obté el títol de la finestra actual mitjançant la funció *GetWindowText*, que retorna una cadena de caràcters.

```
GetWindowTextW(hwnd, windowTitle,
sizeof(windowTitle))
```

Per facilitar la visualització d'aquests canvis de tasca, s'ha dissenyat la funció *PrintTimestamp*, la qual és responsable de generar i registrar les marques de temps en format [AAAA-MM-DD HH:MM:SS:SSS]. En ser cridada, utilitza l'API de Windows per obtenir l'hora local del sistema mitjançant la funció GetLocalTime.

```
void PrintTimestamp()
{
    SYSTEMTIME st;
    GetLocalTime(&st);
    //...
    Resta de codi
    ...//
}
```

6.2 Generació arxiu del *dataset* de Canvis de Tasca

Garantir una escriptura segura a l'arxiu del *dataset* és crucial quan es treballa en un entorn amb múltiples fils d'execució o processos que podrien accedir simultàniament al mateix fitxer. Sense les precaucions adequades, això podria conduir a problemes de concurrència que afectarien la consistència i la integritat de les dades emmagatzemades.

Per garantir una escriptura segura a l'arxiu del *dataset*, s'ha implementat l'ús de l'objecte *lockguard* associat a la llibreria *mutex* [6] en aquest context específic de C++ ajuda a prevenir aquests problemes. El *mutex* (o semàfor) actua com una barrera que permet que només un fil tingui accés al fitxer alhora. Quan un fil obtingui el bloqueig del *mutex*, això indica que cap altre fil podrà accedir al fitxer fins que el bloqueig sigui alliberat.

Finalment, cal destacar que l'objectiu d'aquest mòdul és la generació del *dataset* on queden registrats els timestamps dels canvis de tasca.

7 MÒDUL "ANÀLISIS DE XARXA"

El mòdul "Anàlisis de Xarxa" té com a objectiu principal capturar el tràfic de xarxa de l'usuari i generar un *dataset* que contingui els paquets capturats. Aquest mòdul ha estat desenvolupat amb un enfocament integral per garantir una funcionalitat eficient i accessible dins del sistema global.

En el disseny del sistema, aquest mòdul està estretament relacionat amb el mòdul de "Canvi de Tasques". En la fase de planificació, s'ha considerat crucial assegurar que els dos primers mòduls puguin ser executats simultàniament, per la qual cosa s'ha implementat mitjançant *threads* que faciliten la seva execució coordinada.

7.1 Generador de Traces de Xarxa i Anàlisi de Dades

Pel desenvolupament d'aquest mòdul, l'elecció del programari *TShark* s'ha revelat fonamental. *TShark* destaca per la seva capacitat d'utilitzar comandes per iniciar i configurar la captura i anàlisi de xarxa, la qual cosa concorda amb els requisits del projecte. El codi s'ha escrit en Python, seleccionat per les seves facilitats en projectes relacionats amb la gestió i l'ús de la xarxa.

7.2 Funcionalitats del mòdul

Les funcionalitats d'aquest mòdul es despleguen en un ordre específic. Primerament, es verifica si l'usuari té instal·lat el programari *TShark*, essencial per la captura de tràfic. En cas que no estigui instal·lat, el mòdul procedeix automàticament a la descàrrega i instal·lació.

Seguidament, s'executa la comanda de *TShark* per iniciar la captura de tràfic, utilitzant els flags adequats per configurar l'execució del programa segons les necessitats del projecte. Els flags es representen a la Fig. 2.

Flag		Descripció
-i	interficie	S'estableix la interficie per la cual capturar el tràfic
-Y	tcp || udp	S'especifiquen els protocols a capturar
-T	fields	A partir d'aquest flag s'especifiquen els diferents camps que tindrà l'arxiu de sortida
-e	frame.number	Número de datagrama
-e	frame.time	Timestamp del datagrama
-e	ip.src	IP origen del datagrama
-e	ip.dst	IP destí del datagrama
-e	ip.proto	Protocol del datagrama
-e	tcp.srcport	Port TCP origen
-e	tcp.dstport	Port TCP destí
-e	udp.srcport	Port UDP origen
-e	udp.dstport	Port UDP destí
-e	frame.len	Longitud del datagrama
-E	header=y	A la primera fila de l'arxiu de sortida hi hauran títols per cada columna de dades
-E	separator=;	S'especifica el símbol amb el qual es volen separar les dades
-E	quote=d	S'utilitzen cometes dobles (") per les dades a l'arxiu de sortida
-E	occurrence=f	Especifica com s'agrupen les dades a l'arxiu de sortida

Fig. 2: Flags TShark

I la línia de codi que s'encarrega d'executar la comanda és la següent:

```
subprocess.run(comando, shell=True,
cwd=path)
```

S'ha utilitzat el mòdul de Python *subprocess* [7] per executar la comanda de *TShark* en un nou procés, el procediment manual equivalent per reproduir aquesta funció seria obrir una Command Prompt, CMD, i introduir la comanda de *TShark* per començar a capturar l'activitat de xarxa.

A l'hora d'executar el sistema, en diferents entorns, proporcionat als usuaris, s'han trobat certes qüestions a solucionar per un funcionament correcte en diversos entorns d'execució. Els desafiaments tècnics en el desenvolupament d'aquest mòdul es centren especialment en l'ajustament dels PATHs dels arxius durant la conversió del codi Python a un executable per a sistemes operatius Windows. L'assegurament de la coherència del directori d'execució, especialment en la configuració de l'executable, ha requerit una revisió detallada per assegurar un comportament fiable i consistent en diversos entorns d'execució.

7.3 Interfície gràfica Tkinter

Finalment, l'ús de la llibreria Tkinter [9] per crear una interfície gràfica senzilla ha estat un punt clau. Aquesta interfície millora significativament l'experiència de l'usuari final, proporcionant una manera intuïtiva d'interactuar amb les funcionalitats de captura de tràfic i anàlisi. Així, el sistema es presenta no només com una solució tècnica eficaç, sinó també com una amigable i accessible per a tota classe d'usuaris.

Com es pot observar en la Fig. 3, la interfície gràfica disposa de dos quadres de text per introduïr el nom i l'edat, dos botons tipus ràdio per escollir el gènere i la interfície de xarxa que s'utilitza per a la captura i un botó per iniciar la sessió de l'experiment i un altre per aturar-la quan l'usuari necessiti.

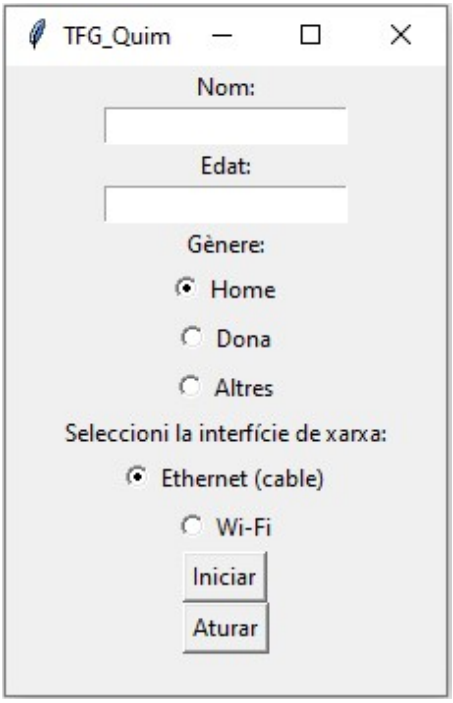

Fig. 3: Interfície gràfica Tkinter

8 Mòdul "Gestió de Dades"

El propòsit fonamental d'aquest mòdul és realitzar les modificacions necessàries als *datasets* per garantir un processament eficient per part de la xarxa neuronal. Aquest procés implica la homogeneïtzació del format de les dades en ambdós *datasets* i l'addició d'un identificador d'usuari (ID) que facilitarà la realització de prediccions específiques per a cada usuari.

La primera tasca del mòdul consistirà en la homogeneïtzació dels formats de dades dels dos *datasets*. Això inclourà l'ajustament de les columnes, la conversió de formats de temps per garantir la consistència i integritat dels conjunts de dades.

La següent funció s'encarrega de convertir el format de temps perquè sigui idèntic en ambdós datsets per facilitar el posterior entrenament de la Xarxa Neuronal. La funció rep una línia del *dataset* i la retorna amb les modificacions pertinents.

```
def convertir_data(text_data)
```

Per a garantir la concordança adequada entre els timestamps de les traces de xarxa i les dades temporals de l'usuari, s'afegirà un identificador d'usuari (ID) a les dades. Aquesta incorporació facilitarà a la xarxa neuronal la distinció i predicció específica per a cada usuari, millorant així la precisió de les prediccions.

La següent funció s'encarrega de generar el ID per a cada usuari, el ID es genera en vers a un hash del nom del fitxer, que és el nom de l'usuari, i li afegeix un nombre aleatori per assegurar que no hi ha dos IDs iguals encara que hi hagi dos arxius amb el mateix nom.

```
def generar_id_usuari(nom_fitxer)
```

Finalment, el mòdul inclourà una fase de test i validació per assegurar-se que les modificacions realitzades no introdueixen errors i que les dades preprocessades són adequades per a l'entrada a la xarxa neuronal. Aquesta fase implicarà la comprovació de la consistència, l'exploració visual de les dades i, si és necessari, ajustaments addicionals.

9 MÒDUL "XARXA NEURONAL"

L'últim mòdul a desenvolupar per completar la implementació del sistema és un prototip d'una xarxa neuronal. Aquesta xarxa neuronal tindrà la responsabilitat de predir, si és possible, els temps entre tasques dels usuaris basant-se en l'arxiu que conté la informació del tràfic de xarxa. L'objectiu principal d'aquest mòdul és emular l'entorn d'un ISP, que disposa d'una quantitat massiva de dades. En el seu context, és factible entrenar una xarxa neuronal amb èxit, gràcies a l'abundància de dades i als recursos computacionals i de temps disponibles. En cas que aquesta xarxa neuronal fos capaç de predir amb èxit els temps entre tasques dels clients, podríem declarar que el nostre temps entre tasques és una dada vulnerable en termes de privacitat.

Per a aquest propòsit específic, s'opta per una xarxa neuronal de tipus recurrent, coneguda com a RNN [10]. Aquesta elecció es basa en la seva capacitat per capturar relacions a llarg termini en les dades. Veure Fig. 4.

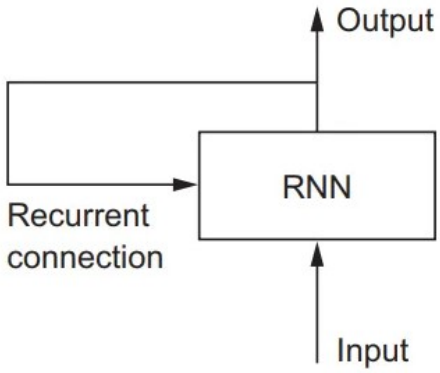

Fig. 4: Esquema RNN

En primer lloc, les RNNs destaquen per la seva habilitat per processar seqüències de dades amb una memòria interna, permetent-los capturar relacions a llarg termini. Aquesta característica és crítica en el context dels temps entre tasques, ja que aquests comportaments es desenvolupen en el temps i poden mostrar dependències complexes entre les diferents observacions.

La capacitat de les RNNs per mantenir una memòria de llarg termini fa que siguin adequades per entendre i predir seqüències temporals. En aquest escenari, on es vol simular l'entorn d'un ISP amb una gran quantitat de dades, aquesta característica s'ajusta bé a l'objectiu de processar dades seqüencials relacionades amb el tràfic de xarxa dels usuaris.

En aquest cas, s'està treballant amb dades temporals, ja que, el més rellevant d'ambdós *datasets* és el temps exacte en el qual es genera cada línia d'informació. La capa LSTM és una elecció raonable per a aquest tipus de dades, ja que és capaç de mantenir una memòria de llarg termini, la qual és útil per capturar relacions complexes en seqüències temporals.

Pel que fa al nombre de neurones a la capa LSTM (50 en aquest cas), la tria d'aquest paràmetre és una qüestió empírica i pot dependre de la complexitat de les dades i de la tasca. En general, és útil experimentar amb diferents valors per veure com afecta el rendiment del model. Típicament, una major quantitat de neurones és beneficiosa per capturar patrons més complexos. No obstant això, cal tenir precaució amb la possibilitat de sobreajustament, especialment en conjunts de dades més petits. La tria del nombre de neurones també pot dependre de la naturalesa de la tasca específica i de la quantitat de dades disponibles. En algunes situacions, es podria considerar l'ús de tècniques com la validació creuada o l'ajust automàtic de paràmetres per trobar la millor configuració.

En aquest context, la implementació s'ha realitzat en Python per la seva flexibilitat i ampli suport per a tasques d'anàlisi de dades i aprenentatge profund. Les llibreries *TensorFlow* [11] i *scikit-learn* [12] s'han utilitzat per simplificar la construcció i entrenament del model, així com per a la preparació i processament de les dades. *TensorFlow* és especialment coneguda per les seves capacitats en el camp dels models d'aprenentatge profund, mentre que *scikit-learn* ofereix eines efectives per a la preparació i avaluació de dades en tasques d'aprenentatge automàtic. La combinació d'aquestes llibreries proporciona una plataforma robusta per a la implementació i experimentació amb models de xarxes neuronals en Python.

En resum, la tria d'un model RNN amb una capa LSTM i el nombre de neurones és adequada per a la tasca de predicció temporal basada en seqüències de timestamps de traces de xarxa.

10 RESULTATS

A continuació, es detallen els *datasets* que s'han obtingut després de dur a terme els experiments amb usuaris reals. Es troba necessari remarcar que els resultats que es mostren són exemples dels *datasets* que ha generat el sistema i utilitzaria el mòdul de la xarxa neuronal posteriorment, cal especificar que la xarxa neuronal no ha estat entrenada per manca de grans quantitats de dades, temps i capacitat computacional. Entrenar-la amb poques dades i amb temps insuficient hauria donat una resposta incompleta que no es considera suficient com a resposta d'aquesta investigació.

10.1 *Dataset* de les Traces de Xarxa

Primerament, el mòdul de "Anàlisis de xarxa" ha generat un fitxer CSV[5] que conté tota l'activitat de xarxa generada per l'usuari durant l'experiment, també cal afegir que aquest CSV ha sigut modificat pel mòdul de "Gestió de Dades" per afegir l'ID d'usuari i per normalitzar el format i les dades. La Fig. 5 mostra un petit exemple del *dataset* de traces:

10.2 *Dataset* dels Canvis de Tasques

Simultàniament, el mòdul de "Canvi de tasques" ha generat un altre fitxer CSV que conté tots els canvis de tasca que ha realitzat l'usuari durant la sessió de l'experiment. De la mateixa manera que l'anterior *dataset*, aquest també ha sigut modificat pel mòdul de "Gestió de Dades" per afegir l'ID d'usuari i per assegurar que els formats coincideixin. La Fig. 6 mostra un exemple del *dataset* de canvi de tasques:

[5]CSV: Els fitxers CSV (de l'anglès comma-separated values) són un tipus de document en format obert senzill per representar dades en forma de taula, on les columnes se separen per comes (o punt i coma) i les files per salts de línia.

Nom_Exemple	24	Home								
frame.number	frame.time	ip.src	ip.dst	ip.proto	tcp.srcport	tcp.dstport	udp.srcport	udp.dstport	frame.len	user.id
2	[2024-02-06 10:52:56:925]	142.250.201.74	192.168.0.16	6	443	51210			139	e77f43edf835abdc9cc7355a2fbf1a0d9879
3	[2024-02-06 10:52:56:973]	192.168.0.16	142.250.201.74	6	51210	443			54	e77f43edf835abdc9cc7355a2fbf1a0d9879
4	[2024-02-06 10:52:56:986]	192.168.0.17	239.255.255.250	17			52125	1900	143	e77f43edf835abdc9cc7355a2fbf1a0d9879
5	[2024-02-06 10:52:57:717]	40.101.137.18	192.168.0.16	6	443	49446			102	e77f43edf835abdc9cc7355a2fbf1a0d9879
6	[2024-02-06 10:52:57:719]	192.168.0.16	40.101.137.18	6	49446	443			96	e77f43edf835abdc9cc7355a2fbf1a0d9879
7	[2024-02-06 10:52:57:719]	192.168.0.16	40.101.137.18	6	49446	443			96	e77f43edf835abdc9cc7355a2fbf1a0d9879
8	[2024-02-06 10:52:57:747]	40.101.137.18	192.168.0.16	6	443	49446			60	e77f43edf835abdc9cc7355a2fbf1a0d9879
9	[2024-02-06 10:52:58:591]	192.168.0.17	192.168.0.255	17			50910	32414	63	e77f43edf835abdc9cc7355a2fbf1a0d9879
10	[2024-02-06 10:52:58:592]	192.168.0.17	192.168.0.255	17			57828	32412	63	e77f43edf835abdc9cc7355a2fbf1a0d9879
14	[2024-02-06 10:53:03:591]	192.168.0.17	192.168.0.255	17			50910	32414	63	e77f43edf835abdc9cc7355a2fbf1a0d9879
15	[2024-02-06 10:53:03:592]	192.168.0.17	192.168.0.255	17			57828	32412	63	e77f43edf835abdc9cc7355a2fbf1a0d9879
16	[2024-02-06 10:53:04:730]	192.168.0.16	142.250.184.3	6	51213	443			147	e77f43edf835abdc9cc7355a2fbf1a0d9879
17	[2024-02-06 10:53:04:730]	192.168.0.16	142.250.184.3	6	51213	443			147	e77f43edf835abdc9cc7355a2fbf1a0d9879
18	[2024-02-06 10:53:04:744]	142.250.184.3	192.168.0.16	6	443	51213			60	e77f43edf835abdc9cc7355a2fbf1a0d9879
19	[2024-02-06 10:53:04:744]	142.250.184.3	192.168.0.16	6	443	51213			60	e77f43edf835abdc9cc7355a2fbf1a0d9879
20	[2024-02-06 10:53:04:765]	142.250.184.3	192.168.0.16	6	443	51213			163	e77f43edf835abdc9cc7355a2fbf1a0d9879
21	[2024-02-06 10:53:04:765]	142.250.184.3	192.168.0.16	6	443	51213			128	e77f43edf835abdc9cc7355a2fbf1a0d9879
22	[2024-02-06 10:53:04:765]	192.168.0.16	142.250.184.3	6	51213	443			54	e77f43edf835abdc9cc7355a2fbf1a0d9879
23	[2024-02-06 10:53:04:765]	142.250.184.3	192.168.0.16	6	443	51213			93	e77f43edf835abdc9cc7355a2fbf1a0d9879
24	[2024-02-06 10:53:04:766]	192.168.0.16	142.250.184.3	6	51213	443			93	e77f43edf835abdc9cc7355a2fbf1a0d9879
25	[2024-02-06 10:53:04:767]	142.250.184.3	192.168.0.16	6	443	51213			140	e77f43edf835abdc9cc7355a2fbf1a0d9879

Fig. 5: *Dataset* traces

window.swap	swap.time	window.name	user.id
1	[2024-2-6 10:53:11:393]	Window Focus Changed: Privacitat digital: riscos i mesures - Brave	e77f43edf835abdc9cc7355a2fbf1a0d9879
2	[2024-2-6 10:54:9:594]	Window Focus Changed: Correu - Quim Mila Sole - Outlook - Brave	e77f43edf835abdc9cc7355a2fbf1a0d9879
3	[2024-2-6 10:54:35:108]	Window Focus Changed: YouTube â€" Mozilla Firefox	e77f43edf835abdc9cc7355a2fbf1a0d9879
4	[2024-2-6 10:54:43:215]	Window Focus Changed: Correu - Quim Mila Sole - Outlook - Brave	e77f43edf835abdc9cc7355a2fbf1a0d9879
5	[2024-2-6 10:54:52:763]	Window Focus Changed: YouTube â€" Mozilla Firefox	e77f43edf835abdc9cc7355a2fbf1a0d9879
6	[2024-2-6 10:55:7:720]	Window Focus Changed: C:\Users\Quim\Desktop\TFG_QUIM\bin\Timestamps.e	e77f43edf835abdc9cc7355a2fbf1a0d9879
7	[2024-2-6 10:55:18:974]	Window Focus Changed: Correu - Quim Mila Sole - Outlook - Brave	e77f43edf835abdc9cc7355a2fbf1a0d9879
8	[2024-2-6 10:55:24:806]	Window Focus Changed: YouTube â€" Mozilla Firefox	e77f43edf835abdc9cc7355a2fbf1a0d9879
9	[2024-2-6 10:55:31:632]	Window Focus Changed: TFG_Quim.py - TFG_QUIM - Visual Studio Code	e77f43edf835abdc9cc7355a2fbf1a0d9879
10	[2024-2-6 10:55:42:468]	Window Focus Changed: TFG_Quim	e77f43edf835abdc9cc7355a2fbf1a0d9879
11	[2024-2-6 10:56:43:169]	Window Focus Changed: Friends - Discord	e77f43edf835abdc9cc7355a2fbf1a0d9879
12	[2024-2-6 10:56:50:110]	Window Focus Changed: YouTube â€" Mozilla Firefox	e77f43edf835abdc9cc7355a2fbf1a0d9879
13	[2024-2-6 10:56:51:586]	Window Focus Changed: YouTube â€" Mozilla Firefox	e77f43edf835abdc9cc7355a2fbf1a0d9879
14	[2024-2-6 10:56:51:588]	Window Focus Changed: Friends - Discord	e77f43edf835abdc9cc7355a2fbf1a0d9879
15	[2024-2-6 10:56:56:358]	Window Focus Changed: Nom_Exemple.csv - TFG_QUIM - Visual Studio Code	e77f43edf835abdc9cc7355a2fbf1a0d9879
16	[2024-2-6 10:57:1:942]	Window Focus Changed: Friends - Discord	e77f43edf835abdc9cc7355a2fbf1a0d9879
17	[2024-2-6 10:57:2:37]	Window Focus Changed: Nom_Exemple.csv - TFG_QUIM - Visual Studio Code	e77f43edf835abdc9cc7355a2fbf1a0d9879
18	[2024-2-6 10:57:8:321]	Window Focus Changed: TFG_Quim	e77f43edf835abdc9cc7355a2fbf1a0d9879

Fig. 6: *Dataset* canvi de tasques

10.3 Relació dels *datasets*

Com s'ha esmentat a la Subsecció 5.4 i la Subsecció 5.5 d'aquest document, es pot establir una relació entre els canvis de tasca i l'activitat de xarxa generada per les aplicacions. Aquesta relació no resulta fàcil d'intuir amb un enfocament simplista; per aquest motiu, aquesta investigació proposa la implementació d'una Xarxa Neuronal. A través dels conjunts de dades de Traces, aquesta xarxa té com a objectiu predir el temps entre tasques per a cada usuari. S'entrena de manera recursiva, verificant la precisió de les prediccions mitjançant la comparació amb les dades del conjunt de Canvi de Tasques. Malauradament, el prototip de la Xarxa Neuronal suggerit en aquest treball no ha pogut ser entrenat a causa de limitacions de recursos.

11 Conclusions

El treball presentat ha explorat la complexa relació entre els canvis de tasca i l'activitat de xarxa generada per les aplicacions, amb l'objectiu principal de saber si el nostre temps entre tasques és una dada vulnerable amb termes de privacitat, però també per millorar la comprensió i la predicció del temps entre tasques per a cada usuari.

Primerament, s'ha posat de manifest la intrínseca complexitat de la relació entre els canvis de tasca i l'activitat de xarxa. La diversitat i dinàmica d'aquesta relació ressalten la necessitat d'abordar aquesta problemàtica amb mètodes analítics avançats.

En segon lloc, la implementació proposada d'una Xarxa Neuronal ha sorgit com a eina prometedora per abordar la complexitat esmentada. Mitjançant l'ús de *datasets* de Traces, la Xarxa Neuronal busca predir de manera precisa el temps entre tasques, oferint una proposta innovadora, ja que fins al moment no es coneix cap investigació que ho tracti. L'ús d'un enfocament recursiu en l'entrenament de la Xarxa Neuronal s'ha destacat com a element clau. La verificació progressiva de les prediccions amb el *dataset* de Canvi de Tasques ressalta la necessitat d'adaptar el model a mesura que s'obtenen més dades.

En tercer lloc, no obstant això, les limitacions de recursos han impedit la completa realització de l'entrenament de la Xarxa Neuronal. Aquesta limitació posa en relleu la importància de consideracions pràctiques en la implementació d'aquest tipus d'investigació. Malgrat les limitacions, la investigació apunta a una potencial aplicació pràctica en la millora de la gestió del temps en entorns amb elevada activitat de xarxa.

En conclusió, aquest treball ha proporcionat una visió aprofundida de la relació entre canvis de tasca i activitat de xarxa, i ha plantejat la implementació d'una Xarxa Neuronal com a mitjà per abordar aquesta problemàtica. Les

limitacions actuals presenten oportunitats per a futures investigacions, amb l'objectiu de millorar encara més la precisió i l'aplicabilitat del model proposat.

AGRAÏMENTS

Vull expressar el meu sincer agraïment a totes aquelles persones que han estat fonamentals en la realització d'aquest Treball de Final de Grau.

En primer lloc, vull donar les gràcies al meu tutor, Carlos Borrego Iglesias, per la seva guia i suport constants al llarg d'aquesta jornada acadèmica. Els seus suggeriments, consells i dedicació han estat inestimables, contribuint de manera significativa al desenvolupament i èxit d'aquest projecte.

A més a més, desitjo agrair als valuosos participants de l'experiment, sense els quals aquest treball no hauria estat possible. La seva participació i col·laboració han enriquit les dades i aportat perspectives essencials a la recerca.

Finalment, vull estendre el meu agraïment al meu entorn més proper, família i amics, per haver estat a la meva vora durant aquesta fase final del meu grau. El vostre suport emocional i encoratjament han estat crucials i han fet aquesta etapa més significativa i assolible.

A tots, moltes gràcies per la vostra inestimable contribució i suport.

REFERÈNCIES

[1] *Borghouts, Judith, Duncan P. Brumby, and Anna L. Cox. "TimeToFocus: Feedback on interruption durations discourages distractions and shortens interruptions."ACM Transactions on Computer-Human Interaction (TOCHI) 27.5 (2020): 1-31. [En línia].* Recuperat de: https://dl.acm.org/doi/abs/10.1145/3396044

[2] *Züger, Manuela, et al. "Reducing interruptions at work: A large-scale field study of flowlight."Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017. [En línia].* Recuperat de: https://dl.acm.org/doi/abs/10.1145/3025453.3025662

[3] *"What is Iterative and Incremental Development? — Wrike". Versatile and Robust Project Management Software — Wrike. [En línia].* Recuperat de: https://www.wrike.com/project-management-guide/faq/what-is-iterative-incremental-development/

[4] *"Windows API - C++ Articles". [En línia].* Recuperat de: https://cplusplus.com/articles/winapi/

[5] *C++, "ctime lib". [En línia].* Recuperat de: https://cplusplus.com/reference/ctime/

[6] *C++, "mutex lib". [En línia].* Recuperat de: https://cplusplus.com/reference/mutex/

[7] *"subprocess - Subprocess management". Python documentation. [En línia].* Recuperat de: https://docs.python.org/3/library/subprocess.html

[8] *"Tshark — tshark.dev". Tshark — tshark.dev. [En línia].* Recuperat de: https://tshark.dev/

[9] *"tkinter - Interface de Python para Tcl/Tk". Python documentation. [En línia].* Recuperat de: https://docs.python.org/es/3/library/tkinter.html

[10] *W. Koehrsen. "Recurrent Neural Networks by Example in Python". Medium. [En línia].* Recuperat de: https://towardsdatascience.com/recurrent-neural-networks-by-example-in-python-ffd204f99470

[11] *"Keras: la API de alto nivel para TensorFlow — TensorFlow Core". TensorFlow. [En línia].* Recuperat de: https://www.tensorflow.org/guide/keras?hl=es

[12] *"scikit-learn: machine learning in Python — scikit-learn 1.4.0 documentation". scikit-learn: machine learning in Python — scikit-learn 0.16.1 documentation.[En línia].* Recuperat de: https://scikit-learn.org/stable/