Anàlisi de trànsit de xarxa Bitcoin

Ot Altimira López

Resum– Bitcoin és una criptomoneda que va sorgir l'any 2009. Aquesta moneda funciona a través d'una xarxa de nodes distribuïts i interconnectats que treballen conjuntament com a entitat reguladora verificant les transaccions i duent a terme un control de la moneda mitjançant la tecnología blockchain. En una xarxa com aquesta, on hi han molts diners movent-se contínuament per ella, la robustesa i seguretat del sistema són imprescindibles. L'objectiu d'aquest projecte és analitzar el tràfic de la xarxa Bitcoin per tal de poder extreure patrons de comportament d'aquest i, així, poder detectar malfuncionaments o comportaments anòmals de la xarxa. Per dur a terme aquesta anàlisi, es pretén desenvolupar un software capaç d'extreure informació del tràfic que circula per la xarxa i poder-la vizualitzar de forma que pugui ser analitzada fàcilment.

Paraules clau– Bitcoin, xarxa, anàlisi, python

Abstract– Bitcoin is a cryptocurrency that appeared in 2009. This currency functions through a distributed network of nodes that are interconnected and work together as a regulatory entity verifying the transactions and carrying out a control of the currency through the blockchain technology. In a network like this, where there is a lot of money moving through it, the security and robustness of the system is a must. The objective of this project is to analyze the traffic of the Bitcoin network in order to be able to extract behavior patterns of it to detect malfunctions or abnormal behaviors of the network. To carry out this analysis, it is intended to develop a software which is capable of extracting traffic information that flows through the network with the objective to facilitate an easy way to visualize the information to analyze.

Keywords– Bitcoin, network, analysis, python

1 INTRODUCCIÓ

Bitcoin és una criptomoneda que va aparèixer l'any 2009, quan un usuari anònim anomenat Satoshi Nakamoto va publicar un article definint el funcionament d'aquesta i les seves característiques. Aquesta, permet tenir control absolut sobre els bitcoins per part del propietari (usuari), és descentralitzada i és anònima. Per aconseguir totes aquestes característiques, Bitcoin fa ús de la tecnologia blockchain, un "llibre de comptabilitat" descentralitzat on s'apunten totes les transaccions i és immutable un cop s'ha escrit en ella, de tal manera que es té un registre complet de totes les transaccions que s'han dut a terme des de l'inici de Bitcoin al 2009. Aquesta tecnologia té un potencial increïblement elevat de revolucionar la manera en què les empreses, clients, governs, ciutadans i tot tipus de dispositius interactuaran en el futur degut a la seva naturalesa distribuïda i segura.

Per tal de poder dur a terme les seves funcions, Bitcoin funciona a través de la xarxa fent us del seu propi protocol de comunicació. Aquest protocol intercomunica els nodes, que treballen conjuntament per validar transaccions i mantenir el registre de la blockchain actualitzat, permetent transmetre la informació necessària entre ells.

En aquest document s'analitza i explica el protocol que utilitza Bitcoin per intercomunicar els nodes de la xarxa i es proposa una solució per poder recavar informació del tràfic d'aquesta de tal manera que es pugui analitzar fàcilment per extreure coneixement sobre la xarxa i poder detectar mal funcionaments o anomalies en aquesta.

Aquest document s'estructura de la següent forma: en la Secció 2 es presenten els objectius i la motivació del projecte i en la Secció 3 la metodologia. La Secció 4 presenta el context i l'estat de l'art. La Secció 5 explica els conceptes bàsics de Bitcoin, les característiques del Bitcoin i protocol Bitcoin. La Secció 6 presenta l'anàlisi de la xarxa amb la problemàtica i la solució proposada. La Secció 7 són els resultats del projecte, que estan dividits en 3 apartats: anàlisi per sessions, anàlisi global i histograma de temps. La Secció 8 planteja treballs futurs. Finalment la Secció 9 són les conclusions seguides dels agraïments.

• E-mail de contacte: ot.al@hotmail.com
• Menció realitzada: Tecnologies de la Informació
• Treball tutoritzat per: Cristina Pérez Solà, DEIC
• Curs 2017/18

2 OBJECTIUS I MOTIVACIÓ

L'objectiu d'aquest projecte és estudiar la criptomoneda Bitcoin i el seu protocol en profunditat, entenent el funcionament i aplicació a la xarxa. L'objectiu d'aquest estudi es poder arribar a analitzar el tràfic de la xarxa Bitcoin per tal d'intentar definir un comportament normal d'aquesta i detectar anomalies o mal funcionaments. Per dur a terme aquesta anàlisi s'implementarà una aplicació per extreure dades dels missatges del protocol Bitcoin i generar gràfiques que podran ser analitzades per trobar comportaments anòmals.

La motivació a l'hora de dur a terme aquest projecte ha estat poder entendre en profunditat la criptomoneda Bitcoin i el seu funcionament a la xarxa, així com el comportament i la seguretat que aquesta criptomoneda té a nivell de protocol de xarxa. Aquest tipus de moneda, si no Bitcoin una altre amb característiques semblants i a arrel de Bitcoin, pot ser un canvi revolucionari per a la humanitat si és adoptat per la majoria de la població mundial de tal manera que s'ha de posar èmfasi en la seguretat d'aquesta xarxa ja que no es pot permetre cap tipus de fallada. És per això que aquest projecte pretén desenvolupar una aplicació capaç d'analitzar el tràfic d'aquesta xarxa buscant possibles anomalies o malfuncionaments.

3 METODOLOGIA

Per a la realització i desenvolupament d'aquest projecte s'han definit un seguit de fases a seguir:

1. **Realització d'un estudi del software:** On s'han identificat i estudiat els diversos llenguatges de programació per dur a terme el projecte. S'han analitzat a nivell de llibreries i frameworks disponibles per a poder programar els diversos procediments que duu a terme l'analitzador de la forma més concisa i clara possible. Com s'explicarà més endavant, el llengüatge escollit ha sigut python.

2. **Realització d'un estudi del protocol Bitcoin:** s'ha dut a terme una anàlisi del protocol Bitcoin i les seves tecnologies per tal d'entendre'l en profunditat i poder extreure els requisits més importants a l'hora d'implementar l'analitzador de xarxa Bitcoin. També s'han estudiat els tipus de paquets i la diversa informació que contenen.

3. **Programació contínua i cíclica:** S'ha desenvolupat l'analitzador duent a terme cicles d'implementació. Primerament s'ha desenvolupat un nucli que permet extreure els valors més importants del tràfic i capaç de llegir arxius .pcap realitzant un filtratge bàsic de les dades. Un cop el nucli ha estat implementat, s'ha hagut d'anar polint i millorant duent a terme diversos cicles d'implementació on s'ha modificat i corregit codi d'aquest i s'han implementat més funcionalitats com la generació de gràfics.

4. **Testeig:** Al llarg de l'implementació de l'analitzador s'han dut a terme diverses proves per corroborar que les dades extretes i el funcionament dels mòduls implementats fóssin correctes.

4 CONTEXT I ESTAT DE L'ART

En aquesta secció s'introduirà la història de Bitcoin i el seu estat de l'art respecte a projectes amb semblança a aquest.

4.1 Història

Al novembre de 2008 es registrà un article en una llista de distribució de criptografia titulat "Bitcoin: A Peer-to-Peer Electronic Cash System"[1] i firmat amb el nom de Satoshi Nakamoto. En aquest article es detallava com utilitzar una xarxa peer-to-peer (xarxa distribuïda on tots els nodes són iguals) i la criptografia de clau pública per a crear un sistema de transaccions electròniques que no depenia de terceres entitats per dur a terme una transacció. En aquest article, es detallava el funcionament de la moneda, de la xarxa Bitcoin i d'una nova tecnologia, la blockchain.

La blockchain és una base de dades distribuïda que basa el seu funcionament en l'encadenament de blocs de dades i, un cop les dades són encadenades al bloc anterior, aquestes, no poden ser modificades. Aquesta tecnologia, aplicada al Bitcoin, permet mantenir un registre immutable de totes les transaccions que s'han dut a terme. El 3 de gener de 2009, Satoshi Nakamoto registra el primer bloc (bloc de gènesi), un bloc especial que creà els primers bitcoins i va iniciar la xarxa Bitcoin. Aquest bloc es va crear utilitzant un text molt especial i significatiu denotant les intencions del Bitcoin: "The Times 03/Jan/2009 Chancellor on brink of second bailout for banks", la portada del diari The Times del 3 de gener de 2009, on la notícia principal era "El canceller a la vora del segon rescat per als bancs".

4.2 Estat de l'art

Bitcoin és una de les primeres implementacions d'una criptomoneda. Actualment la capitalització del mercat de criptomonedes és de 245.740 milions de dòlars [2]. Bitcoin té una capitalització de 104.500 milions de dòlars, sent la que té més capital i seguida per ethereum amb 43.558 millions de dòlars. Amb aquestes xifres, ha despertat gran interès mediàtic. Acutalment s'estan duent a terme molts projectes d'investigació en el món de les criptomonedes i, sobretot de la tecnologia blockchain. Pel que fa a projectes i informació respecte a l'anàlisi de la xarxa Bitcoin, hi ha molts pocs recursos útils per al desenvolupament del projecte. S'ha trobat un mòdul del programa *Scapy* [3] (programa que permet tractar paquets del tràfic de la xarxa) que intenta aportar funcionalitats a aquest per permetre llegir el protocol Bitcoin podent extreure tota la informació de forma estructurada, però el mòdul no té en compte la partició dels missatges Bitcoin en la capa de transport en el protocol TCP i falla a l'hora d'estructurar les dades.

Un exemple a gran escala és el projecte satoshi[4]. Aquest és de codi obert i programat pels desenvolupadors de Bitcoin amb l'objectiu de dur un seguiment de nodes de la xarxa, presentar tot tipus d'estadístiques i protegir Bitcoin mitjançant la transparència de l'activitat de la xarxa. Aquest projecte permet recompilar dades dels nodes en el que s'executa com el consum de CPU o l'ús de memòria. En aquest sentit, satoshi arriba a processar moltes més gràfiques a part de les de xarxa.

5 CONCEPTES BÀSICS DE BITCOIN

En aquesta secció s'exposaran les característiques més destacades del Bitcoin i es parlarà del protocol Bitcoin, explicant els tipus de missatges més importants a l'hora de realitzar el projecte.

5.1 Característiques del Bitcoin

La criptomoneda Bitcoin es caracteritza per oferir unes propietats que les monedes convencionals no ofereixen.

- **Està descentralitzada:** és una moneda que no és controlada per cap estat, banc, institució financiera o empresa tot i que molts governs han intentat regular el seu ús. Al ser descentralitzada, fa que no sigui possible generar inflació mitjançant la creació de més moneda (com es pot fer amb el dòlar o l'euro) ja que Bitcoin es regula mitjançant la generació de bitcoins cada cop que un bloc de la blockchain és creat. És la propia xarxa que, duent a terme la mineria, gestiona l'emisió de bitcoins de forma descentralitzada.
- **És anònima:** no es necessari revalar la identitat de la persona que realitza una transacció per a dur-la a terme. Aquesta és una de les principals característiques que va inspirar la creació del Bitcoin.
- **Es té la propietat íntegra del diner:** la propietat del diner és 100% de l'usuari. No hi ha terceres entitats com podrien ser els bancs. Això també te els seus inconvenients, si es perd la clau pertanyent a un cert nobre de bitcoins, aquesta no pot ser recuperada de cap manera. És per això que hi han carteres online i pàgines d'intercanvi de monedes que permeten a l'usuari guardar els seus bitcoins de forma que l'usuari no controla les claus privades d'aquests.
- **No és falsificable:** No és possible la falsificació o duplicació gràcies a la tecnologia de la blockchain, que manté un registre de totes les transaccions i amb la que és possible saber exactament d'on provenen els diners. Els usuaris compten amb les seves pròpies carteres, protegides per ells mateixos. Cada usuari és propietari dels seus diners i de la seva protecció.
- **No hi ha intermediaris:** les transaccions es fan directament entre dues entitats. El funcionament peer to peer de la xarxa permet que les transaccions siguin quasi instantànies comparat amb les transferències bancaries i amb uns costos molt baixos de processament.
- **Les transaccions són irreversibles:** un cop s'ha dut a terme una transacció, aquesta no pot ser modificada ni anul·lada.
- **Canvi de divisa:** és possible canviar de divisa com si fós una moneda normal.

5.2 Protocol Bitcoin

El protocol Bitcoin és un protocol de la capa d'aplicació utilitzat pels nodes que participen en la xarxa Bitcoin. Aquest protocol consta de 30 tipus de missatges diferents que són utilitzats pels nodes per dur a terme la comunicació entre ells [5]. Seguidament s'explicaran els més rellevants pel projecte.

- **version:** quan un node crea una connexió amb un altre immediatament envia la versió. Fins que no rep una resposta amb la versió de l'altre no hi pot haver comunicació.
- **verack:** aquest missatge s'envia en resposta a *version* acceptant la versió.
- **addr:** proveeix informació sobre els nodes conneguts de la xarxa.
- **getblocks:** permet als nodes demanar els blocs que els hi falten a la seva còpia de la blockchain.
- **inv:** permet als nodes informar sobre el seu coneixement d'un o més objectes (transaccions o blocs). Pot ser rebut sense haver sigut solicitat o com a resposta de *getblocks*. Pot conternir com a màxim 50.000 entrades.
- **getdata:** s'utilitza en resposta a un *inv* per a recuperar el contingut d'un objecte específic.
- **tx:** en resposta a *getdata*, descriu una transacció Bitcoin.
- **block:** s'envia en resposta a un *getdata* que demana la informació d'un bloc.

6 ANÀLISI DE LA XARXA

Analitzar la xarxa Bitcoin no és senzill degut a la dimensió i naturalesa distribuïda d'aquesta. Per aquest projecte l'anàlisi s'ha dut a terme capturant el tràfic des d'un node cap a la xarxa. Aquestes captures s'han guardat en arxius *.pcap* de gran tamany que han sigut particionats en arxius de 500.000 missatges. No tots els missatges són tràfic del protocol Bitcoin.

6.1 Problemàtica

La xarxa Bitcoin ha de garantir uns nivells de seguretat i resistència a mal funcionaments de forma consistent ja que per a què una criptomoneda com Bitcoin sigui adoptada per la població ha de demostrar primerament que és completament segura i resistent a fallades. Assegurar que una xarxa sigui completament segura no és una tasca senzilla tot i així el que es pot fer és analitzar-la. Es pot intentar establir un patró de comportament "normal" i, a partir d'aquest, detectar quan hi ha moviments de missatges anòmals en la xarxa per tal de prevenir o evitar un possible atac o mal funcionament. Per establir un comportament "normal"es necessita analitzar moltíssima informació de forma constant per tal que aquest sigui representatiu i això requereix de molts recursos computacionals per poder-ho dur a terme en un temps raonable.

6.2 Solució proposada

La solució que es proposa en aquest projecte és un analitzador de la xarxa Bitcoin que permet extreure informació rellevant d'aquesta per tal de que es pugui analitzar.

6.2.1 *Analyzer*

El llenguatge escollit per desenvolupar l'*analyzer* ha sigut python [6] i s'han fet ús del programa *Scapy* [7](que es pot importar en format llibreria) i la llibreria *matplotlib*[8]. *Scapy* proporciona funcionalitats per a la lectura de fitxers *.pcap* i la manipulació de sessions i paquets. *matplotlib* proporciona funcionalitas per processar les dades i dibuixar gràfiques d'aquestes. Amb els mètodes que proporciona *Scapy* per a llegir arxius *.pcap* i tractar la informació que contenen s'ha desenvolupat un lector de sessions que, un cop les ha llegit totes processa els paquets de cada sessió i permet extreure'n les dades. *Matplotlib* proporciona els mètodes necessaris per generar gràfiques i poder editar-les, de tal forma que puguin ser presentades de forma entenedora.

El protocol Bitcoin, al ser un protocol d'aplicació, va encapsulat dins de paquets TCP. A l'hora de dur a terme el projecte, això ha sigut un problema ja que molts paquets Bitcoin queden particionats en la capa de transport (on s'implementa el protocol TCP) i la lectura d'aquests es dificulta. Aquest problema s'ha pogut solucionar buscant explícitament les capçaleres Bitcoin per extreure quin tipus de missatge contenien. Un altre problema ha sigut no disposar de suficients recursos de processament hardware, ja que l'anàlisi dels arxius *.pcap* s'ha dut a terme amb un ordinador convencional i es realitza carregant aquests a memòria i, si l'arxiu es molt gran, aquest procés és lent i pot utilitzar tots els recursos RAM de l'ordinador i col·lapsar-lo. És per això que prèvia execució, s'ha de particionar l'arxiu *.pcap* si aquest és de gran tamany. Per particionar-lo s'ha fet ús d'una eina que incorpora *Wireshark* [15]: *editcap*[16], que permet particionar arxius *.pcap* parametritzant quants missatges per partició es volen.

L'*analyzer* rep per paràmetres dos valors, primer un directori d'on llegir i segon un altre on guardar els resultats. Si en el directori d'on llegeix hi ha més d'un arxiu, es crearà un directori de resultats per a cadascun d'ells. Llegeix cada arxiu extraient dos parells de diccionaris. En un es guarda un comptador per cada tipus de missatge, en l'altre es guarda una llista de *timestamps* (marques de temps) de cadascun dels missatges. Per poder aconseguir aquestes dades realitza una anàlisi de les sessions (connexions entre un parell de nodes) per extreure cadascun dels paquets TCP d'aquestes. Un cop s'han extret tots els paquets TCP es llegeix el *payload* (contingut del paquet), on es busca una capçalera del protocol Bitcoin. Si la troba, es guarda el tipus de missatge Bitcoin que pugui conenir i el *timestamp*, emplenant els diccionaris.

6.2.2 Processament de les dades

Els diccionaris que extreu l'*analyzer* es processen mitjançant la llibreria *[*matplotlib] que permet generar gràfiques. L'implementació de l'*analyzer* permet extreure dues gràfiques diferents, un gràfic de barres on es comptabilitzen quants missatges de cada tipus hi ha i un histograma on per un període de temps donat representa els missatges d'un tipus que han passat pel node analitzat. Aquest histograma es genera per cada tipus de missatge. Es poden dur a terme diversos processaments. Aquestes dues gràfiques es poden generar per sessions, on cada sessió té les seves pròpies gràfiques que representen el tràfic entre un parell de nodes i també de forma que s'engloben totes les dades analitzades del fitxer sent una gràfica general sense tenir en compte les sessions.

Generar gràfiques és un procés senzill però generar gràfiques que representessin bé les dades i quedéssin bé estèticament ha representat un repte a l'hora de dur a terme aquest apartat del projecte.

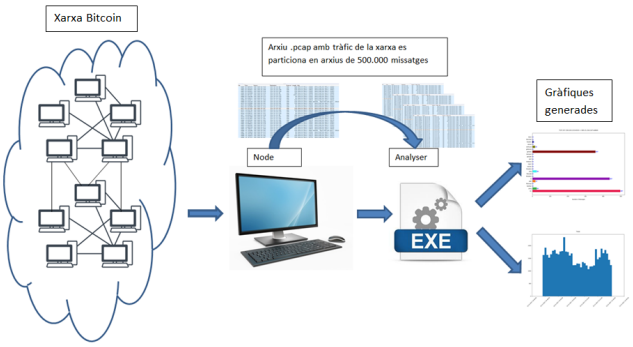

Fig. 1: Diagrama del procés

En la Figura 1 es pot veure el procediment que es segueix a l'hora de dur a terme l'anàlisi de les dades. Primerament hi ha la xarxa bitcoin on, des d'un node, es captura tràfic d'aquesta. Si el fitxer capturat es de gran tamany, aquest es particiona. Els fitxers resultants són llegits per l'*analyzer* i aquest extreu les dades i genera les gràfiques per poder dur a terme l'anàlisi.

7 Resultats

Per a dur a terme proves i extreure resultats amb aquest software, s'han utitzat 3 tipus d'arxius diferenciats pel tamany. Els primers, arxius de tamany reduit, amb 500 missatges de xarxa capturats. Aquests arxius han sigut molt útils a l'hora d'evaluar i comprovar que les dades dels paquets TCP eren extretes correctament. Al ser arxius molt petits, les gràfiques processades no mostraven massa informació ni tenien una forma definida. És per això que s'han utilitzat arxius de tamany mitjà per fer les proves a l'hora de crear les gràfiques i optimitzar-les de forma que quedessin bé. *matplotlib* proporciona eines molt útils per generar gràfiques de forma ràpida però fer que les gràfiques generades siguin clares ha requerit dedicar recursos a l'edició d'aquestes. Finalment s'ha tractat un fitxer de tràfic capturat de 20GB que s'ha particionat en arxius més petits de 500.000 missatges cadascun. Amb aquest arxius s'ha dut a terme una anàlisi dels resultats.

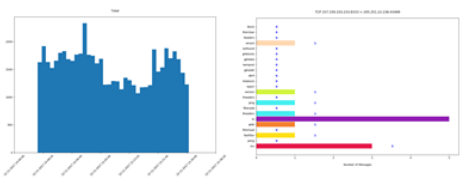

Fig. 2: Exemples de gràfica temporal i de missatges

Les gràfiques com les de la Figura 2 són les que genera

l'*analyzer* i permeten visualitzar tota la informació extreta. Primerament, per sessions (connexions entre dos nodes), es pot distingir la quantitat de missatges enviats i de quin tipus eren els missatges, podent detectar comportaments anòmals com, per exemple, l'enviada massiva de pings. Això és un punt de partida ja que depentent de la durada de la sessió entre els dos nodes, un nombre elevat de missatges no és motiu d'alerta. És per això que s'extreuen les gràfiques temporals, per visualitzar de forma temporal quan han sigut enviats aquests missatges podent veure si s'ha dut a terme l'enviament en un temps prolongat o si s'han enviat de forma massiva en poc temps. Tot i així, per a les sessions s'ha evitat generar gràfiques temporals degut a que la majoria d'aquestes sessions són de curta duració i no es pot extreure massa informació. Només es pot aprofitar quan genera gràfiques de sessions llargues, ja que llavors si que aporten informació. Sobretot s'han evitat perquè generar les gràfiques requereix de temps sense els recursos hardware adecuats,

Les gràfiques totals, permeten veure el tràfic complet que ha sigut capturat en el node, sense tenir en compte les sessions. Aquestes donen una imatge general del que està passant a la xarxa i també permeten veure comportaments anòmals de la mateixa forma que amb les gràfiques de sessió però sense saber entre quins nodes està passant. En aquest cas les gràfiques temporals sí que tenen sentit i valor degut a què per cada tipus de missatge es genera una gràfica temporal on es pot veure el nombre de missatges que s'han enviat per períodes de temps.

7.1 Anàlisi per sessions

Aquestes gràfiques permeten analitzar com està funcionant i què està passant a la xarxa bitcoin entre parells de nodes. En l'anàlisi de les sessions, es troben tres tipus de comportaments diferenciats En el primer cas, que es dóna en moltes de les sessions, ens trobem amb gràfiques com les de la Figura 3. Aquestes sessions són curtes i consisteixen

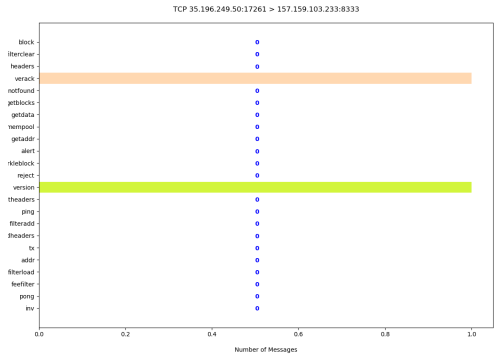

Fig. 3: Node que es connecta i desconnecta

d'un missatge *version* i un missatge *verack* i ocasionalment també algun altre missatge. Això vol dir que la sessió s'ha establert però no s'ha dut a terme cap procediment. És per això que es considera que aquests nodes s'han connectat i seguidament desconnectat sense realitzar res.

En el segon cas, es troben gràfiques com la de la Figura 4. En aquestes, s'estableix sessió i es duen a terme uns quants intercanvis de missatges però sense gran activitat.

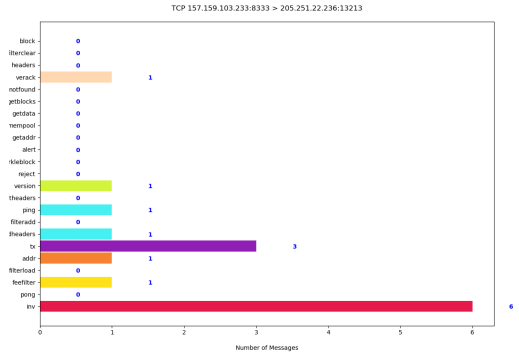

Fig. 4: Node que envia pocs missatges

Amb aquestes gràfiques es pot deduir que aquestes sessions s'han dut a terme quan un usuari ha vulgut fer un procediment, per exemple una transacció, en la xarxa i seguidament desconnectar-se d'ella.

En el tercer cas, hi han certes sessions que tenen un patró comú entre elles. Aquestes envien molts missatges *inv*, *tx* i *getdata*. Tots aquests missatges són per transmetre i anunciar informació. Mirant les sessions que pertanyen a gràfics com la Figura 5, es troba un mateix node enviant per més d'un port informació molt similar a tots els seus peers. Això fa pensar que en la xarxa Bitcoin també hi ha nodes connectats durant periodes llargs de temps que tenen molta informació i la transmeten constantment.

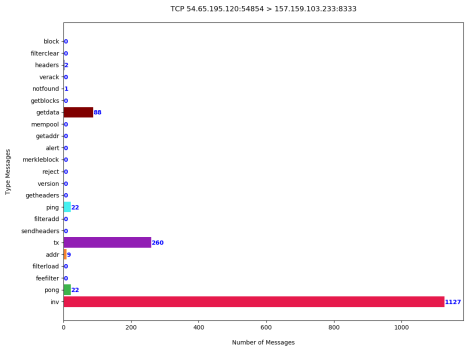

Fig. 5: Node que envia molts missatges

7.2 Anàlisi global

Les Figures 6 i 7 pertanyen a dues de les anàlisis de les 45 particions del fitxer analitzat de 20GB. Mirant-los tots, hi ha un patró que segueixen. Totes les anàlisis tenen una repartició molt semblant. Això és degut a què les sessions com les vistes en la figura 5 són predominants en les gràfiques totals. Aquestes gràfiques aporten una visió general del trànsit de la xarxa. Majoritàriament es duen a terme *inv*, *tx*, *getdata*, *ping* (per comprovar si un node està en funcionament) i *pong* (resposta al ping).

Obviant els missatges *ping* i els *pong*, el tràfic de la captura de 20GB denota que la xarxa té un trànsit elevat de tres tipus de missatges que són els que s'utilitzen per demanar, compartir i anunciar informació. Aquest comportament sembla ser l'usual en la xarxa i vol dir que s'està transmetent informació entre els peers per tal de que tots puguin dur

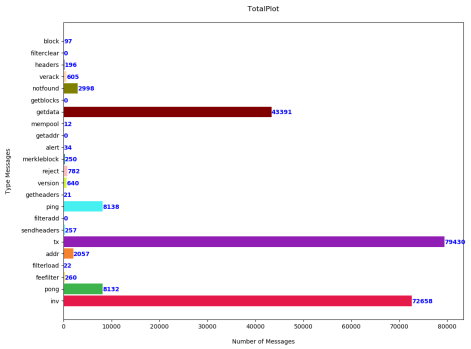

Fig. 6: Gràfica total 1

a terme les seves funcions i transaccions. Si es pensa en què la xarxa Bitcoin suporta un sistema de moneda electrònica, té sentit el tràfic analitzat ja que els events que més es duran a terme en una xarxa com aquesta seran intercanvis d'informació de les transaccions i blocs generats, que és el que denoten les gràfiques.

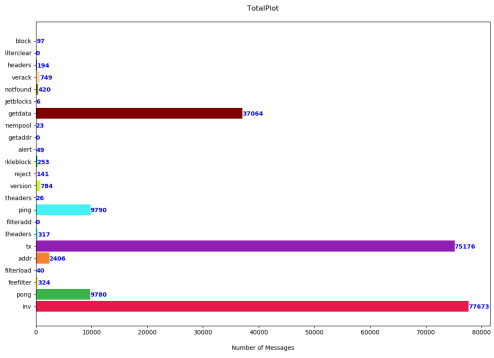

Fig. 7: Gràfica total 2

En les gràfiques generades s'aprecia com la comunicació entre els nodes és en gran part per transmetre informació de transaccions i anunciar informació. Amb menys mesura, es reben peticions d'informació tot i ser un tràfic menys predominant.

7.3 Histogrames de temps

En aquesta secció es parla dels histogrames de temps generats. Aquest tipus de gràfica només s'ha generat pel tràfic total, sense contemplar les sessions. Això és degut a què moltes sessions de la xarxa Bitcoin duren un temps curt i aquestes gràfiques perden el sentit. Si es disposesin de recursos hardware suficients, es podrien generar igualment però despreciant les sessions sense dades rellevants. Al no tenir aquests recursos i al generar una gràfica temporal per cada tipus de missatge (30 gràfiques per sessió) moltes de les quals no contenen informació s'ha decidit no generar-les per sessions i fer-ho només per el total de les dades. Que hi hagin tantes sessions de curta duració fa pensar que molts nodes es connecten per realitzar una transacció i seguidament es desconnecten de la xarxa.

En les Figures 8 i 9 apareixen dues gràfiques d'un periode de 40 minuts dels missatges *version* i *verack* pertanyents a un dels arxius de la partició de 20GB. En aquestes gràfiques es pot apreciar com el patró de les dues és semblant. Analitzant més fitxers i les seves gràfiques es troba que sempre es compleix aquest patró. Això vol dir que la majoria de les sessions s'estableixen satisfactoriament i molt poques vegades no s'acorda una versió entre dos nodes.

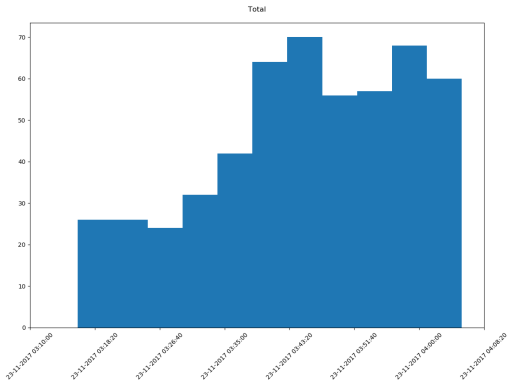

Fig. 8: Histograma de missatges version

En les gràfiques d'aquests dos tipus de missatges no es contemplen sessions que ja havien iniciat abans de què es capturés el tràfic. També són útils per vizualitzar les franges horàries on es connecten més nodes si es generen gràfiques suficients.

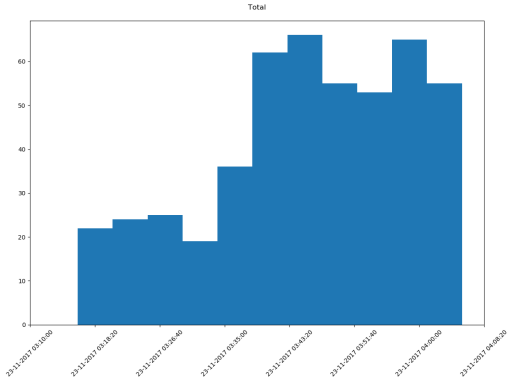

Fig. 9: Histograma de missatges verack

La Figura 10 representa el flux de missatges *inv* més habitual trobat en les diverses gràfiques generades.

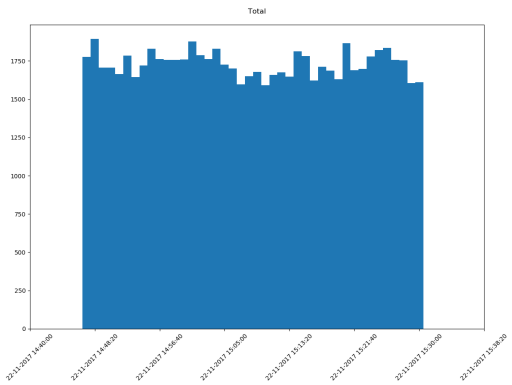

Fig. 10: Histograma de missatges inv

Aquesta mostra un flux elevat d'aquest tipus de missatge al llarg de totes les diverses franges de temps. Totes les

gràfiques generades, en més o menys mesura tenen una certa forma d'ona. Aproximadament cada deu minuts hi ha uns minuts on es generen més missatges *inv*. Una possible explicació és que la generació d'un nou bloc de la blockhain provoca que els nodes retransmetin la informació per a què el nou bloc arribi a tothom i per això es generin aquests petits pics cada cert temps.

La figura 11 representa el flux de missatges *block* predominant en les captures. En aquest, s'accentúa el patró de comportament on cada cert període de temps (10 minuts aprox.) hi han pics de missatges transmesos. Això és degut a què aquest tipus de missatge es transmet per enviar informació sobre blocs i, aquests, es generen aproximadament cada deu minuts i s'ha dinformar a la xarxa del nou bloc provocant el pic de tràfic.

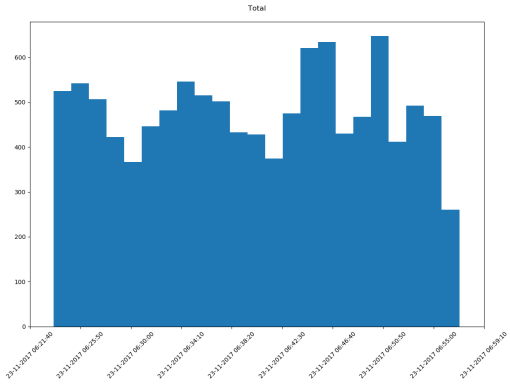

Fig. 11: Histograma de missatges block

La figura 12 mostra el comportament general dels missatges *tx*. Al llarg de tot el tràfic analitzat es troba un flux constat d'aquest tipus de missatge amb pics aleatoris que varien en el temps. Donat que aquests missatges descriuen una transacció Bitcoin és normal que tinguin un comportament més aleatori ja que el nombre de transaccions que es duen a terme a la xarxa és variant en el temps i en funció d'aquestes, el tràfic d'aquest tipus de missatge augmenta o disminueix. Aquest comportament aleatori també s'aplica als missatges *inv* degut a què anuncien informació de les transaccions, però al ser utilitzat per més propòsits fà que l'aleatorietat no afecti tant en el tràfic d'aquest tipus de missatge i no sigui molt apreciat en les gràfiques d'*inv*.

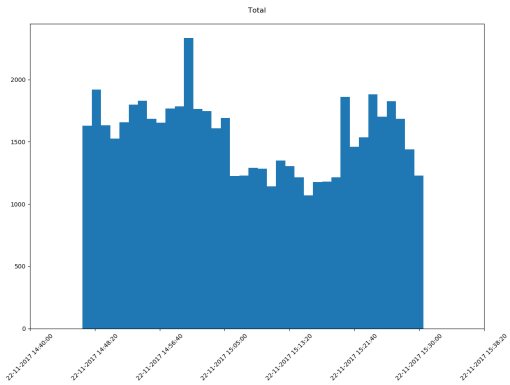

Fig. 12: Histograma de missatges tx

Les gràfiques de temps generades han sigut molt útils a l'hora d'analitzar el comportament dels tipus de missatges. Aquestes permeten veure més enllà de la quantitat de missatges transmesos oferint una vista on es veuen els missatges transmesos en períodes de temps. Com s'ha explicat, la generació per sessions d'aquestes gràfiques és possible però es necessiten molts recursos hardware per poder dur a terme la generació d'aquestes en un temps raonable. És per això que en aquest projecte només s'han extret de la totalitat dels missatges dels diversos arxius.

8 Conclusions

Aquest projecte m'ha permès entendre en profunditat el funcionament de la criptomoneda Bitcoin i les tecnologies subjacents, així com la comunicació entre els nodes d'aquesta xarxa distribuïda. Ha sigut una oportunitat per a desenvolupar un software d'anàlisi de la xarxa mitjançant python que fos capaç d'extreure informació de fitxers *.pcap* i generar gràfiques de la informació extreta. Ha sigut un projecte desafiant degut a la manca d'informació a l'hora de desenvolupar el software i l'aparició de problemes com la partició dels missatges Bitcoin dins de paquets TCP o la lectura dels fitxers *.pcap* de grans dimensions. El desenvolupament d'aquest projecte ha aportat una visió simple de les dades de la xarxa Bitcoin de forma que són fàcilment analitzables. Un projecte amb aquestes característiques és fàcilment ampliable amb noves funcionalitatats però el desenvolupament d'aquestes pot requerir de molts coneixements i temps. Aquest projecte presenta una base per a futurs desenvolupaments d'anàlitzadors de xarxa Bitcoin.

8.1 Treballs futurs

El projecte realitzat és tan sols una petita part de l'anàlisi exhaustiu que es podria dur a terme en la xarxa bitcoin. Aquesta xarxa té moltes dades i informació que a l'hora de dur a terme el projecte s'han hagut d'obviar per la complexitat que podria arribar a tenir el desenvolupament d'un software capaç d'analitzar completament la xarxa. El que s'ha dut a terme és un anàlisi de les caracterísitques més bàsiques del tràfic de la xarxa, que són la quantitat de missatges transmesos de cada tipus i els períodes de temps en què han estat transmesos. El projecte estableix una base per a poder analitar arxius .pcap de tràfic de la xarxa Bitcoin tot i així, es podria ampliar molt, treient moltes més estadístiques de, per exemple, el tamany dels missatges o la replicació d'aquests. Amb suficients recursos hardware aquest projecte podria ampliar-se per a analitzar grans volums de dades.

Per poder dur a terme aquesta anàlisi s'haurien de realitzar optimitzacions al codi per tal de fer-lo el més eficient possible i aixi permetre que l'execució tardés un temps raonable. Un objectiu futur per aquest projecte és que llegeixi els fitxers i els processi al mateix temps, de tal manera que llegeixi un flux de dades en comptes de carregar arxius sencers a memòria. El problema a l'hora de llegir un flux de dades és que el processament d'aquestes per sessions no es pot dur a terme ja que no es té tota la informació de les sessions de cop. És per això que la modificació de la forma de lectura dels fitxers no s'ha dut a terme.

AGRAÏMENTS

Vull agraïr a la meva tutora Cristina Pérez Solà per guiarme en la realització d'aquest projecte, ajudar-me cada vegada que ho he necessitat proporcionant-me coneixements i informació molt útil i aconsellar-me a l'hora de desenvolupar l'*analyzer*. També a en Joaquin García-Alfaro per donar accés a les dades.

REFERÈNCIES

[1] "Bitcoin: A peer-to-peer electronic cash system", bictoin.org, 2018. [Online] Available: https://bitcoin.org/bitcoin.pdf

[2] "Cryptocurrency Market Capitalizations — CoinMarketCap", Coinmarketcap.com, 2018. [Online]. Available: https://coinmarketcap.com/.

[3] "jbcayrou/scapy-bitcoin", GitHub, 2018. [Online]. Available: https://github.com/jbcayrou/scapy-bitcoin/blob/master/README.md.

[4] "Grafana", Statoshi.info, 2018. [Online]. Available: https://statoshi.info/.

[5] "Bitcoin Wiki", En.bitcoin.it, 2018. [Online]. Available: https://en.bitcoin.it/wiki.

[6] "Welcome to Python.org", Python.org, 2018. [Online]. Available: https://www.python.org/.

[7] "Welcome to Scapy's documentation! ", Scapy.readthedocs.io, 2018. [Online]. Available: https://scapy.readthedocs.io/en/latest/.

[8] "Matplotlib: Python plotting — Matplotlib 2.2.2 documentation", Matplotlib.org, 2018. [Online]. Available: https://matplotlib.org/.

[9] A. Antonopoulos, Mastering Bitcoin. Sebastopol, CA: O'Reilly Media Inc., 2015.

[10] "Bitcoin Developer Reference - Bitcoin", Bitcoin.org, 2018. [Online]. Available: https://bitcoin.org/en/developer-reference#message-headers.

[11] M. Steinkirch, "Black Hat Python: Infinite possibilities with the Scapy Module", Bt3gl.github.io, 2018. [Online]. Available: http://bt3gl.github.io/black-hat-python-infinite-possibilities-with-the-scapy-module.html.

[12] "Scapy: Sessions (or Streams)", The IT Geek Chronicles, 2018. [Online]. Available: https://itgeekchronicles.co.uk/2014/12/02/scapy-sessions-or-streams/.

[13] "How can I filter a pcap file by specific protocol using python?", Stack Overflow, 2018. [Online]. Available: https://stackoverflow.com/questions/2247140/how-can-i-filter-a-pcap-file-by-specific-protocol-using-python/2248149#2248149.

[14] "The Bitcoin Protocol – 4 – Network Messages 1 – Version", Coin Logic, 2018. [Online]. Available: https://coinlogic.wordpress.com/2014/03/09/the-bitcoin-protocol-4-network-messages-1-version/.

[15] "Wireshark Go Deep", Wireshark, 2018. [Online]. Available: https://www.wireshark.org

[16] "editcap - The Wireshark Network Analyzer 2.6.1", Wireshark.org, 2018. [Online]. Available: https://www.wireshark.org/docs/man-pages/editcap.html.