



UNIVERSITÀ
DEGLI STUDI
FIRENZE

UNIVERSITÀ DEGLI STUDI DI FIRENZE
Scuola di Ingegneria

Corso di Laurea Magistrale in
INGEGNERIA DELLE TELECOMUNICAZIONI

**Identificazione dei frodatori nel mercato
della telefonia IP internazionale: un
approccio basato sull'analisi della
reputazione**

**Fraudsters detection in the international IP
telephony market: an approach based on
analysis of reputation**

Tesi di Laurea di
Francesco Ermini

Relatori:

Ing. Francesco Chiti

Ing. Tommaso Pecorella

Indice

| | | |
|----------|--|-----------|
| 1 | Introduzione | 4 |
| 2 | Scenario | 5 |
| 2.1 | La rete PSTN | 6 |
| 2.1.1 | Modelli di interconnessione regolamentati ITU | 9 |
| 2.1.2 | Modelli di interconnessione nel mercato telefonico deregolamentato | 12 |
| 2.2 | La rete Internet | 13 |
| 2.2.1 | Modelli di interconnessione nella rete Internet | 15 |
| 2.3 | La rete della telefonia su IP | 17 |
| 2.3.1 | Modello di interconnessione nel mercato della telefonia IP | 18 |
| 2.3.2 | Least Cost Routing | 20 |
| 3 | Problema | 22 |
| 3.1 | Le frodi | 22 |
| 3.2 | I frodatori | 24 |
| 3.3 | Rivelazione delle frodi | 25 |
| 3.4 | Rivelazione dei frodatori | 26 |
| 4 | Soluzione | 28 |
| 4.1 | Idea | 28 |
| 4.2 | Architettura | 33 |
| 4.3 | Ricostruzione della traccia | 36 |
| 4.3.1 | Individuazione della frode | 37 |
| 4.3.2 | Ricostruzione parziale della traccia | 38 |
| 4.4 | Gestione dei feedback | 41 |
| 4.4.1 | Generazione dei feedback | 41 |
| 4.4.2 | Creazione della matrice dei feedback | 43 |
| 4.4.3 | Elaborazione a posteriori della matrice dei feedback | 43 |
| 4.4.4 | Aggiornamento della matrice dei feedback con i valori precedentemente ottenuti | 44 |
| 4.5 | Calcolo della reputazione | 45 |
| 4.5.1 | TNASL | 45 |
| 4.6 | Mitigazione della frode | 50 |
| 4.7 | Considerazioni | 51 |
| 4.7.1 | Contrasto alle strategie maligne dei frodatori | 51 |
| 4.7.2 | Onestà dei providers telefonici | 52 |
| 4.7.3 | Considerazioni su scelte progettuali | 52 |
| 5 | Validazione | 54 |
| 5.1 | Uso del simulatore | 54 |
| 5.2 | UML Concettuale | 55 |
| 5.3 | Funzionamento | 56 |
| 5.3.1 | Generazione dei nodi | 57 |
| 5.3.2 | Generazione tracce | 58 |
| 5.3.3 | Creazione della matrice dei feedback | 58 |
| 5.3.4 | Somma dei feedback precedentemente ottenuti | 59 |
| 5.3.5 | Calcolo della reputazione | 60 |

| | |
|--|-----------|
| 6 Risultati | 61 |
| 6.1 Acquisizione e stima dei dati | 61 |
| 6.1.1 Providers, intermediari e minuti di chiamate | 62 |
| 6.1.2 Frodi e frodatori | 64 |
| 6.2 Simulazione tramite scenari | 65 |
| 6.2.1 <i>Scenario ideale</i> | 66 |
| 6.2.2 <i>Scenario variabile</i> | 70 |
| 6.2.3 <i>Scenario applicativo</i> | 75 |
| 6.2.4 <i>Scenario attuativo</i> | 77 |
| 6.3 Analisi costi benefici | 79 |
| 7 Conclusioni | 81 |
| 8 Appendice A | 83 |

1 Introduzione

Nel panorama della telefonia globale a lunga distanza le frodi agli operatori di telecomunicazioni rimangono un problema in gran parte irrisolto.

Secondo il Global Fraud Loss Survey di CFCA^[4] nel 2017 gli operatori di telefonia hanno subito frodi per 29.2 bilioni di dollari, che è il 1.27% dei ricavi globali nel settore. Le statistiche mostrano chiaramente che le frodi sono in aumento dal 2013. Gli strumenti utilizzati oggigiorno per individuare i frodatori non sono efficaci.

I motivi di questa inefficacia sono da ricercarsi nelle carenze dei protocolli di comunicazione, nell'assenza di regolamenti e nelle dinamiche concorrenziali del mercato. La complessità e la globalità dell'infrastruttura rendono impossibile sistemare le cause di insicurezza.

Nell'ultimo decennio i ricercatori hanno sperimentato modelli per l'individuazione delle frodi^{[6][7][8][9]}. I modelli analizzati si basano sull'elaborazione dei dati contenuti nei CDR, Call data record o nelle tracce audio delle chiamate. *Grazie a questi metodi gli operatori sono in grado di riconoscere le chiamate frodate nelle loro tracce, ma non riescono ad individuare il colpevole della frode in modo altrettanto preciso e diretto.*

Infatti l'instradamento di una chiamata segue un principio simile al *passa-parola*. L'operatore d'origine non parla direttamente con quello di terminazione ma viene mediato da un certo numero di operatori intermediari, detti operatori di transizione.

Né l'operatore d'origine né quello di transizione conoscono tutta la catena formata dagli operatori di transito che hanno instradato la chiamata. L'operatore fraudolento si nasconde nell'opacità della catena di comunicazione. Trovarlo è come trovare un ago in pagliaio. La ricerca nella letteratura scientifica di metodologie per fronteggiare il problema dell'individuazione dei frodatori è scarsa e non risulta al momento nessuna soluzione valida candidata^[1].

L'innovazione che verrà apportata a questa tesi è quella di studiare, per la prima volta nello scenario internazionale degli operatori telco, un approccio per l'individuazione degli operatori fraudolenti basandosi sulla reputazione dei nodi.

L'idea è quella di attribuire un indice di reputazione ai nodi che gestiscono le chiamate internazionali sulla base del comportamento, onesto o fraudolento, rivelato per quel nodo. L'approccio seguito sarà quello utilizzato negli studi relativi alle *Network of Trust* in reti *P2P* dove l'individuazione del nodo maligno avviene tramite la cooperazione dei nodi stessi della rete.

La sfida iniziale sarà capire in che modo sia possibile utilizzare l'approccio usato nelle *Network of Trust* in una rete, quella delle telecomunicazioni, differente rispetto alle reti *P2P* solitamente utilizzate per quell'approccio. La sfida finale sarà quella di validare il sistema di reputazione, tramite simulazione, per capire limiti, efficacia e prestazioni del sistema ideato.

2 Scenario

Lo scenario delle telecomunicazioni che oggi conosciamo è il punto di arrivo di una evoluzione storica iniziata nei primi anni del 900'. Era l'inizio di una nuova industria, altamente redditizia, che porterà in poco tempo alla posa di migliaia di chilometri di cavi sotterranei e sottomarini e alla costruzione di edifici, detti *carrier hotel*, dove custodire le apparecchiature di interconnessione.

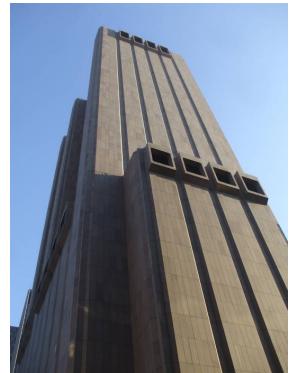
Alcuni dei palazzi più famosi che si trovano a New York[12] sono illustrati in figura 1.



(a) Western Union Headquarters, 1930. 60-hudson-street, NY.



(b) "The Hub" AT&T Long Lines Headquarters, 1932. 32-avenue-of-the-americas, NY.



(c) AT&T Long Lines Building, 1974. 33-thomas-street, NY.



(d) Telehouse America, colocation data center (1996).

Figura 1: Carrier hotels

In poco tempo l'industria delle telecomunicazioni vide affermarsi pochi grandi colossi che monopolizzarono l'intero mercato, tra cui *AT&T* per il mercato internazionale e *Verizon* per quello interno.

Oltre a ciò il settore era altamente regolamentato dalle leggi del governo federale. Nel 1934 il presidente Franklin D. Roosevelt firmò il *Communications Act* che recita: *For the purpose of regulating interstate and foreign commerce in communication by wire and radio [...] there is hereby created a commission to be known as the 'Federal Communications Commission'*.

Intorno ai primi anni 90 il sistema della telefonia globale era diventato un sistema complesso e costoso. L'intero sistema era gestito da poche compagnie che avevano il monopolio del mercato. Fu così che nel 1996, al fine di abbassare le tariffe e incentivare l'evoluzione della rete, il presidente Bill Clinton firmò il *Telecommunication Act*, liberalizzando così il mercato delle telecomunicazioni: *"let anyone enter any communications business – to let any communications business compete in any market against*

any other. "[2]" Oggi, come mostrato in immagine 2, quasi tutto il mercato globale delle telecomunicazioni è gestito da aziende private in diretta concorrenza tra loro.

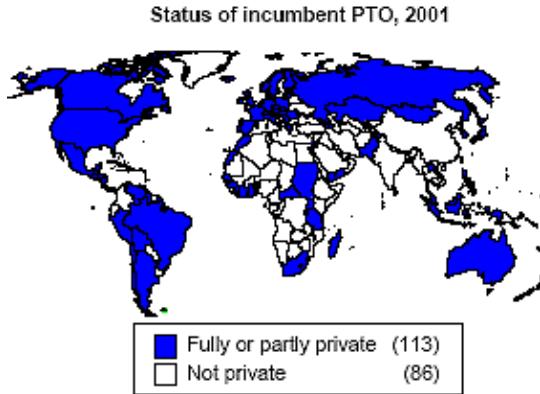


Figura 2: Liberalizzazione del mercato telco globale

Sul piano tecnologico la diffusione del *VoIP*, introdotto nel 1995, complicò ulteriormente le cose. La trasmissione della voce tramite la rete internet, soprattutto per le chiamate internazionali, portò alla nascita di un nuovo mercato e alla proliferazione di nuove compagnie. La facilità di entrata nel mercato "internet" consentì l'ingresso nello scenario della telefonìa internazionale a molte piccole aziende che, grazie ai bassi costi di esercizio, entrarono in competizione con i colossi della telefonìa internazionale, come *AT&T*, che operavano nel vecchio sistema telefonico detto *PSTN*.

Tra queste compagnie alcune si specializzarono nel mercato *retail*, offrendo vendita e installazione di servizi telefonici *VOIP* agli utenti (in prevalenza aziende) mentre altre si specializzarono nell'intermediazione del traffico telefonico a lunga distanza. Le prime, i *VOIP Provider*, non sono di particolare interesse per questo lavoro di tesi mentre lo sono le seconde, i *VOIP Wholesale provider (o carrier)*. Quest'ultimi in pratica svolgono lo stesso lavoro dei carrier internazionali ma al posto di operare tramite l'infrastruttura telefonica, operano tramite l'infrastruttura internet. Gli operatori *VOIP wholesale* offrono servizi di interconnessione tra operatori telefonici "classici", operatori *VOIP* e gli stessi *VOIP wholesale*. Queste compagnie operano al confine tra la rete internet e la rete telefonica, tra regolamenti non definiti ed accordi strategici tra operatori, in un settore in cui la concorrenza è molto forte e le dinamiche di mercato spingo le piccole compagnie a fare di tutto pur di aumentare i profitti.

Tra i *VOIP wholesale providers* si nascondono gli operatori fraudolenti oggetto di questa tesi. Per questo motivo sarà necessario analizzare il complesso scenario in cui operano queste compagnie: lo scenario della telefonìa *VOIP* internazionale. L'intento di questo capitolo è quello di evidenziare chi sono gli attori che partecipano a questo scenario, come interagiscono sul piano tecnologico e quali interessi economici ne ricavano. Per migliorare la chiarezza espositiva la descrizione dello scenario *VOIP* verrà preceduta dalla descrizione dello scenario (1) della telefonìa classica *PSTN* e (2) della rete internet.

2.1 La rete PSTN

L'infrastruttura, intesa come il mezzo dove fisicamente passa l'informazione, è un insieme di "cavi" e "commutatori" (oltre ai collegamenti radio, satellitari..etc) dislocati in

tutto il pianta. L'infrastruttura telefonica, diversamente da quella internet, non nacque come una rete globale bensì come una rete nazionale, "proprietà privata" dello stato sul cui territorio si estendevano quei cavi e quei commutatori. Tutt'oggi la rete telefonica internazionale è organizzata in modo geografico, suddivisa per stati e nel caso della telefonia fissa anche per regioni.

Le aree in cui è suddivisa la rete telefonica sono dette *Local access and transport area code (LATA)*. All'interno di un area *LATA* operano le compagnie telefoniche locali, dette *Local Exchange Carrier (LEC)*. Ciascuna compagnia locale gestisce l'infrastruttura telefonica (e internet) operando tramite un struttura ramificata di *central office (CO)* (*class 5 switch*) che permettono alla compagnia locale di raggiungere le singole utenze o dei "convogliatori" (?) di più linee detti *Private Branch Exchange (PBX)*. Il traffico telefonico tra operatori diversi dentro una stessa *LATA*, ovvero tra *local exchanges*, viene fatto passare tramite i *Tandem Exchange* e non viene sottoposto a tariffazioni aggiuntive. I collegamenti tra i *CO* delle compagnie locali e le utenze finali sono detti *local loop*.

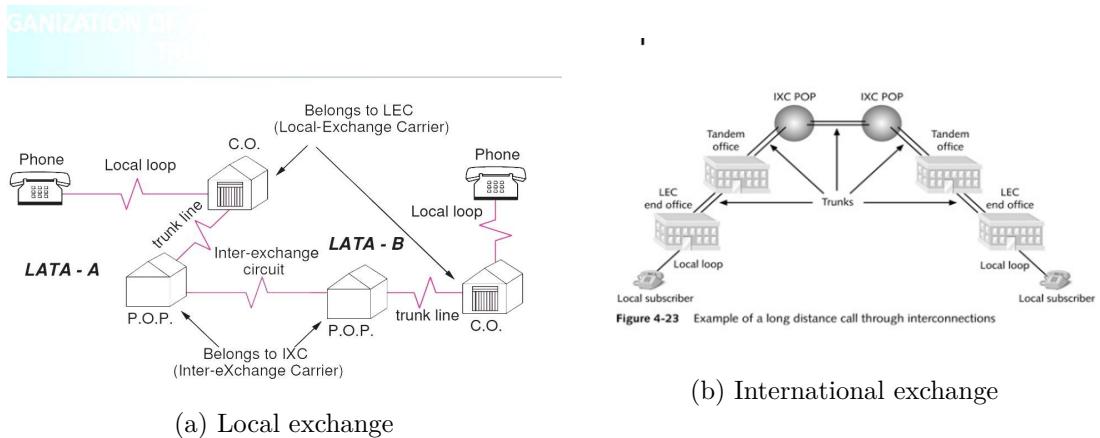


Figura 3: PSTN

Quando una chiamata deve traversare aree diverse si classifica come *inter-LATA*. Una chiamata *inter-LATA*, spesso detta chiamata a lunga distanza o (impropriamente) chiamata internazionale, viene gestita in modo diverso da una locale. Le compagnie che gestiscono l'instradamento del traffico internazionale sono dette *Inter Exchange carrier (IXC)* o più comunemente *long-distance carrier*. L'instradamento delle chiamate internazionali avviene tramite i class-4 switches. Già dal 1980 i class-4 switches furono collegati a cavi da 4-fili e convertiti pienamente alla trasmissione digitale a time-division multiplexing (TDM) per modulari insieme molteplici chiamate. Tra questi: T1, T3, OC-3, etc. Questi commutatori sono detti anche tandem switches, toll switching trunk or toll connecting trunk.

L'instradamento di una chiamata a lunga distanza nella rete *PSTN* avviene in modo gerarchico. L'operatore locale consegna la chiamata all'operatore internazionale di terminazione o di transito (se non c'è collegamento diretto). Il traffico dati della chiamata è preceduto da un fase di setup detta segnalazione. La fase di setup della chiamata nella rete *PSTN* avviene tramite il sistema di segnalazione SS7. L'operatore internazionale di terminazione viene spesso chiamata "gateway internazionale" perché rappresenta il punto d'accesso esterno agli operatori locali che operano in un certo paese.

The PSTN

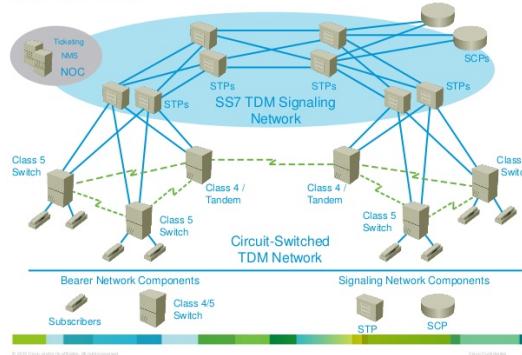


Figura 4: Class-5 switches e class-4 switches

Il problema dell’interconnessione non è solo legato alla realizzazione di una infrastruttura e al rispetto di protocolli standardizzati di comunicazione ma è anche (e soprattutto) legato ai profitti che incentivano le compagnie di telecomunicazioni a mantenere ed evolvere l’infrastruttura esistente.

La sfida nell’interconnessione di operatori telco è trovare regole ed accordi che permettano l’evoluzione della rete e l’abbassamento delle tariffe. Questa sfida si concretizza nel capire (1) in che modo gli operatori possono accordarsi per instradare una chiamata lungo le rispettive infrastrutture e (2) come condividere i profitti tra gli operatori che hanno partecipato ad instaurare la chiamata.

Esistono differenti modelli di interconnessione tra operatori. Comprendere questi modelli vuol dire comprendere il flusso di denaro tra operatori e questo risulterà fondamentale nell’identificazione del frodatore, perché naturalmente il frodatore trae beneficio da quei modelli di interconnessione in cui massimizza i profitti e non viene scoperto.

I modelli verranno analizzati in questo ordine:

1. I modelli di interconnessione regolamentati da ITU per gli operatori telefonici nazionali, detti *ILEC* (monopolisti), nella rete telefonica PSTN.
2. I modelli di interconnessione non regolamentati divenuti popolari dopo la deregolamentazione del mercato delle telecomunicazioni nella rete telefonica PSTN.
3. I modelli di interconnessione utilizzati dagli operatori nella rete internet.

Il principio del *calling party pay* (il chiamante paga) è il principio su cui si sono costruiti i modelli di interconnessione che verranno analizzati in seguito. Filosoficamente parlando, una chiamata è un servizio che permette a due persone di dialogare a distanza. Entrambe le persone che dialogano beneficiano di questo servizio, sia il chiamato che il chiamante. Tuttavia questo principio non è sempre vero, basti pensare alle fastidiose chiamate ricevute dai call-center. Nel principio del *calling party pay* si assume che chi effettua la chiamata, essendo il responsabile di quella azione, si prende anche la responsabilità del costo della chiamata.

Il principio del *calling party pay* si riflette sulle modalità di condivisione dei profitti tra operatori. Infatti il chiamante è abbonato all’operatore che ha originato la chiamata.

Quindi per ogni chiamata internazionale l'operatore di origine addebita al chiamante un costo noto come "collection charge". Una parte di questa cifra sarà ceduta dall'operatore di origine verso gli operatori coinvolti nella chiamata. Questo costo prende il nome di *accounting rate*. Il costo dell'*accounting rate*, naturalmente scorrelato rispetto al *collection charge*, dipende da molti fattori: la concorrenza del mercato, i regolamenti, gli accordi, la "dimensione" delle compagnie coinvolte e la qualità di servizio, ma quello che incide veramente sul costo di *accounting rate* è la tariffa di terminazione.

Generalmente l'operatore che consente l'uso della propria rete per la terminazione richiede il pagamento di una tassa, detta tariffa di terminazione. Ciascun operatore decide la tariffa di terminazione in modo autonomo. Il costo della tariffa di terminazione è legato all'area telefonica in cui opera, che prende il nome di *LATA* (*Local Access and Transport Area*).

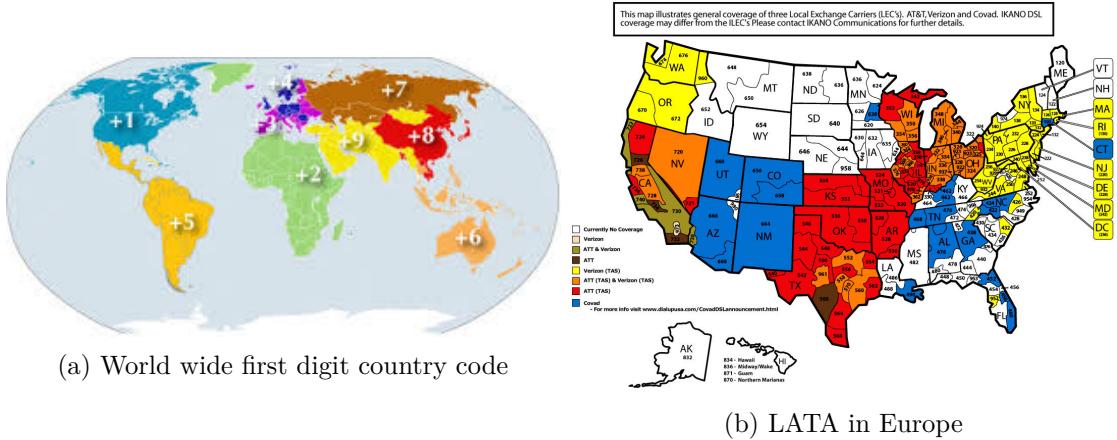


Figura 5: LATA

Ogni LATA ha un codice che viene ricavato dal numero chiamato. Infatti nel traffico internazionale ciascun numero ha la forma di NXX-XX-. Ad esempio si consideri il numero +1 345 123456. La prima cifra, +1, indica il continente di destinazione, ovvero il nord America. La seconda tariffa indica la regione.

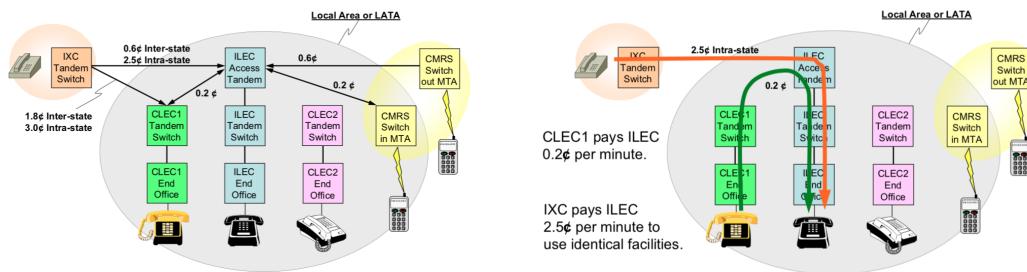


Figura 6: Termination rate prices

2.1.1 Modelli di interconnessione regolamentati ITU

La necessità di effettuare chiamate internazionali risale la metà del novecento con la posa dei cavi sottomarini (TAT-1, 1956). Nel 1970 fu effettuata la prima chiamata

transcontinentale da NY (prefisso 212) a Londra (prefisso 01) senza l'ausilio di un operatore umano per la commutazione. L'istituto internazionale di telecomunicazione ITU fu fondato nel 1865 si occupò di standardizzare procedure e modalità di suddivisione dei ricavi tra gli operatori telefonici internazionali.

Il modello noto come *Accounting Rate Regime* fu definito da ITU-T nel 1988. Questo modello definisce i regolamenti di interconnessione per le chiamate internazionali tra due paesi in cui operano compagnie telefoniche nazionali (ILEC). Il modello bilaterale prevede che due operatori, quello di origine e quello di terminazione, si accordino su una tariffa per lo scambio reciproco del traffico. In seguito all'accordo la tariffa che l'uno paga all'altro è generalmente la metà della tariffa ufficiale. Questo perché nel modello bilaterale il costo della chiamata, invece di essere completamente a carico di un solo operatore (quello che chiede accesso alla rete dell'altro), è suddiviso in modo equo tra le due parti che hanno sottoscritto l'accordo e di conseguenza la tariffa che l'uno paga all'altro risulta dimezzata. Per facilitare l'interconnessione i due operatori considerano la somma del traffico accumulato nelle due direzioni in certo periodo, chiamato *settlement period*. In questo modo la tariffa viene applicata solo alla differenza di traffico. L'operatore che ha inviato più traffico di quanto non ne abbia ricevuto paga la tariffa stabilita al partner dell'accordo bilaterale. Questo traffico prende il nome di *net traffic settlement*.

L'esempio in figura 7 semplifica il concetto. L'operatore A ha 100 minuti di traffico vero l'operatore B, l'operatore B ha 180 minuti di traffico verso l'operatore A. Quindi la differenza di traffico soggetta all'accordo bilaterale è di 80 minuti da B verso A. La tariffa di terminazione di A è di 50 cents/minuto, ma poiché esiste tra i due l'accordo bilaterale la tariffa è di 25cents/min. Alla fine A ha collezionato e trattenuto 1000 cents, B ha collezionato 1400 cents ma ne ha trattenuti 1333 per cui ha guadagnato 200 euro.

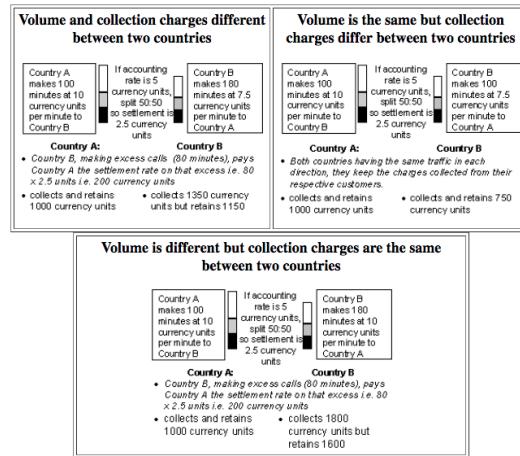


Figura 7: Bilateral System

Tuttavia il modello di accordi bilaterali non funziona bene quando fra i due operatori c'è uno sbilanciamento del traffico in una delle due direzioni. Il costo che un operatore paga per inoltrare la chiamata all'interno della propria rete è considerato nullo perché non varia proporzionalmente al traffico telefonico generato ma dipende solo da costi fissi (infrastruttura, personale, uffici..etc). Di conseguenza se un operatore ha un deficit di traffico entrante rispetto a quello uscente, il costo del traffico di *settlement* verso altri operatori limita la sua capacità competitiva; diminuiscono i margini economici necessari

ad offrire tariffe vantaggiose ai propri clienti rispetto alla concorrenza.

Quando invece i volumi di traffico tra i due operatori sono praticamente uguali nelle due direzioni un accordo di *accounting* risulterebbe inutile. In questo caso i due operatori possono accordarsi per terminare le rispettive chiamate a costo zero. Questo modello di interconnessione prende il nome di *Bill & Keep*. *Bill & Keep* è un accordo bilaterale in cui il *net traffic settlement* è zero. Questa tipologia di accordo è usata tra gli operatori locali dei paesi del Southern African Development Council (i.e. Malawi, Zambia, Zimbabwe, South Africa, Botswana, Mozambique) in cui non ci sono sostanziali differenze di traffico tra operatori. *Bill & Keep* è un modello piuttosto semplice da implementare (non ci sono movimenti di denaro) ed è molto vantaggioso per il fornitore del servizio telefonico il quale trattiene tutti i soldi guadagnati dai propri abbonati. Tuttavia il modello di *Bill & Keep* non può essere utilizzato su scala globale. Infatti l'interconnessione di operatori locali, tramite l'infrastruttura telefonica, è gestita (e dominata) dai *carrier internazionali*. Il guadagno dei *carrier internazionali* deriva interamente dalle tasse di transito. Per questo motivo il modello di *Bill & Keep*, che prevede di non pagare tasse di transito, non può essere usato nello scenario delle chiamate internazionali.

Il sistema di accordi bilaterali definito da ITU-T non si limita a definire l'aspetto economico tra le due parti coinvolte, O e D , ma definisce anche le procedure da seguire nel caso ci sia un operatore di transizione T , situazione piuttosto frequente quando manca un collegamento diretto tra origine e destinazione. I rapporti tra O, D e T sono di due tipologie, *star accounting* e *cascade accounting*. Ogni trimestre l'operatore O dichiara a D il traffico inviato. Se lo dichiara sia a T che a D allora usa la tipologia di *star accounting* mentre se O dichiara il traffico solo a T , il quale poi lo dichiara a D , usa la tipologia di *cascade*. In entrambi i casi tutti gli operatori sono consapevoli delle intenzioni degli altri ed operano con estrema trasparenza. Difatti nel sistema di *Accounting Rate Regime*:

1. si considera sempre uno e un solo operatore di transizione; le catene di operatori di transizioni sono proibite in questo modello di interconnessione.
2. l'operatore di transizione viene scelto in base al principio del collegamento diretto tra i due operatori, non c'è concorrenza tra operatori di transito.
3. le tariffe ufficiali di terminazione di un operatore sono dichiarate apertamente a tutti gli altri. Non esistono accordi segreti.
4. ogni operatore che partecipa alla chiamata conosce gli altri operatori coinvolti. Tramite un sistema di dichiarazione un operatore invia il CDR agli operatori interessati. Eventuali controversie tra operatori sono risolte analizzando i rispettivi CDR.

Il modello di *Accounting rate regime* è il primo ed unico modello regolamentato per l'interconnessione di operatori. In questo modello l'interconnessione tra operatori è stabilita dalle relazioni "geografiche" tra paesi. Ad esempio tutte le chiamate che originano dall'Inghilterra e terminano in Australia sono gestite da *Telstra* (compagnia telefonica australiana) mentre tutte le chiamate che originano dall'Australia e terminano in Inghilterra sono gestite da *BT* (compagnia telefonica britannica). I collegamenti da un paese verso gli altri paesi stranieri sono chiamati *routes*. Nel sistema di *Accounting rate regime* si parla di *settlement routes*. Prima della liberalizzazione del mercato

delle telecomunicazioni, iniziata negli anni 80', tutte le chiamate internazionali tra paesi stranieri avvenivano tramite le *settlement routes*. Queste rotte, collegate attraverso i *tandem switches* della rete *PSTN*, garantivano controllo, trasparenza e assenza di frodi.

Successivamente con la liberalizzazione del mercato delle telecomunicazioni gli operatori di transito divennero competitivi ed iniziarono ad individuare nuovi modelli di business per accrescere i profitti. Nel corso degli anni 90' si consolidarono nuovi modelli di interconnessione noti come (1) confidential rate, (2) arbitrage, (3) refiling o hubbing , (4) re-originating o re-seller.

2.1.2 Modelli di interconnessione nel mercato telefonico deregolamentato

Negli anni 90' il traffico telefonico internazionale crebbe e con questo le tariffe di *net settlement rate* pagate dai paesi sviluppati per terminare le chiamate verso i paesi in via di sviluppo. Le tariffe soggette ad *accounting rate regime* rimanevano elevate. Governi ed istituzioni avevano interesse nel diminuire le tariffe, sia per tutelare le tasche dei cittadini sia per rilanciare il commercio internazionale.

Nel 1995 i paesi membri del WTO firmarono un accordo, il General Agreement on Trade in Services, per lo scambio di merci e servizi che sostituiva il sistema di *accounting rate regime*. Nonostante la maggior parte dei paesi membri dell'ITU non permetteva il libero mercato nel settore delle telecomunicazioni, i 3/4 del traffico internazionale erano generati dai paesi membri del WTO (i più ricchi).

Nel corso di quegli anni si consolidarono modelli di interconnessione tra operatori diversi rispetto a quello di *accounting rate regime*. Questi nuovi modelli di interconnessione verranno discussi in questa sezione, ripartendo dal funzionamento del sistema di *accounting rate regime*. Nel sistema di *accounting rate regime* l'operatore di transizione dichiarava la tariffa di transizione, che tipicamente era di 0.42 SDR. L'operatore di terminazione dichiarava una tariffa ufficiale per la terminazione del traffico, ad esempio di 0.8 SDR. L'operatore di origine O che aveva stipulato un accordo bilaterale con quello di terminazione D passando tramite l'operatore di transito T pagava a D la metà della differenza tra il costo di terminazione ufficialmente dichiarato da D ed il costo di transito, ovvero 0.19 SDR (la metà della differenza tra 0.8 e 0.42). Quindi ricapitolando O spendeva 0.61 SDR di cui 0.19 SDR andavano all'operatore di terminazione e 0.42 SDR a quello di transito.

Il *confidential rate* fu la prima pratica non trasparente adottata dagli operatori. L'operatore di transizione T si accordava con O per far transitare il traffico a 0.05 SDR invece che a 0.42 SDR. In questo modo O spendeva 0.24 SDR, risparmiando più della metà del costo della chiamata. Anche T guadagnava perché, in virtù del fatto che transitare tramite T era più conveniente che stabilire un collegamento diretto tra origine e destinazione, T catturava più traffico di prima e complessivamente incrementava i propri profitti. Solo l'operatore D veniva penalizzato da questo accordo segreto perché la tariffa per terminare la chiamata tramite T risultava più bassa della tariffa che avrebbe ricevuto se la chiamata fosse stata terminata da O direttamente. Nonostante questo il sistema era ancora regolamentato dagli accordi bilaterali. L'operatore D era consapevole che la chiamata originata da O era transitata tramite T.

I breve la ricerca delle tariffe sempre più basse di terminazione portò ad una nuova pratica per l'instradamento della chiamata nota col nome di *hubbing*. In modalità *hubbing* l'operatore di origine e destinazione non hanno nessun accordo. L'operatore di transito si comporta come un aggregatore di chiamate, ovvero un *hub* (da cui il nome *hubbing*) , che convoglia più chiamate provenienti da vari operatori verso l'operatore del chiamato. L'operatore di transizione offre a quello di origine un prezzo conveniente per terminare la chiamata nel paese di destinazione. L'operatore di origine paga quello di transizione, e quello di transizione paga quello di terminazione come se fossero due chiamate separate al posto di una. L'operatore di origine non ha nessun controllo né visibilità su quello che ha fatto l'operatore di transizione. Tuttavia il minor costo per la terminazione della chiamata porta l'operatore di origine ad optare per questa scelta, ovvero affidare la chiamata internazionale ad un carrier internazionale invece che connettersi direttamente all'operatore di terminazione e pagare la tariffa stabilita dagli accordi bilaterali. Le tariffe di terminazione vengono pubblicate dai carrier internazionali mensilmente o volte anche settimanalmente. L'operatore di origine cercherà sempre di inoltrare la chiamata tramite la rotta che costa meno, detta *LCR - Least Cost Routing*, e per questo sceglierà il carrier che ha la tariffa più bassa. I carrier a loro volta cercheranno di avere la tariffa più bassa per una certa destinazione per attirare maggior traffico. in questo modo, anche se i margini di profitto per chiamata sono molto bassi, i profitti complessivi derivanti dalla somma di tutte le chiamate portate a destinazione saranno molto elevati. Questa compravendita di minuti è quella che caratterizza il mercato *wholesale* della telefonia internazionale ed è una pratica legale nei paesi in cui il settore delle telecomunicazioni non è stato de-regolamentato.

L'uso del modello a *hub* permise di introdurre una tecnica chiamata *Re-originating o Refiling* che permetteva ad un operatore internazionale di terminare una chiamata inter-LATA al costo (ridotto) della tariffa locale facendo apparire la chiamata come proveniente dalla stessa LATA del chiamato. Questa tecnica sfruttava le funzionalità di *SS7* per sostituire il *Calling Line Identity CLI*. In molti paesi l'uso di questa tecnica fu considerato illegale. Tuttavia l'operatore di terminazione non ha modo di sapere in quale switch sia stato cambiato il CLI e per questo l'individuazione del operatore furbo non è possibile. Oggi giorno i paesi che hanno ridotto la differenza tra tariffe locali e tariffe internazionali hanno attenuato il fenomeno.

2.2 La rete Internet

Nel corso degli anni 90 si iniziò a diffondere un nuovo tipo di commutazione a pacchetto; stava nascendo la rete internet. Gli stessi operatori che gestivano la *core network* dell'infrastruttura rete telefonica iniziarono a commutare il traffico dati sulla rete internet. Oggigiorno il traffico voce e dati nella *core network* avviene quasi interamente tramite la rete internet. La rete internet sarà trattata in questa sezione. L'utilità di questa trattazione sarà quella di (1) marcire le differenze tra rete internet e rete telefonica e (2) scoprire "chi" c'è dentro la nuvola spesso usata per rappresentare la *core network* della rete internet.

La *core network* è una rete formata da *backbones*, i cavi in fibra ottica che collegano i continenti e le città più importanti di tutto il mondo, illustrati in figura ??, e da *Co-Location Center oppure Internet Exchange Point*, grandi edifici collocati ai capi

delle *backbones* che contengono *Servers*, *Routers*, *Switch Ethernet* e *MPLS* ed anche *digital/analog converter*.

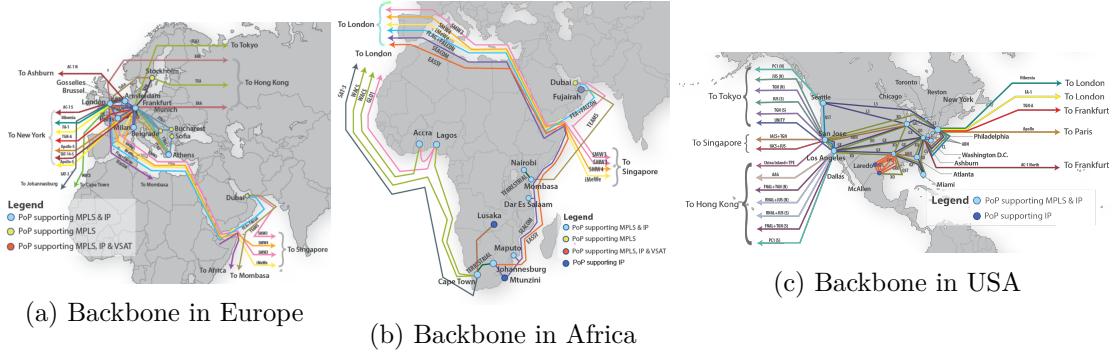


Figura 8: World wide backbone

Co-Location Center. E' uno particolare tipo di *data center* che fornisce in affitto lo spazio, gli apparati di rete e i meccanismi di raffreddamento, continuità energetica e sicurezza anti-intrusione nella struttura. Tipicamente molti *co-location center* sono situati presso le sedi delle compagnie che avevano il monopolio di stato prima della de-regolamentazione del mercato.

Internet eXchange point (IXP). Un IXP è una struttura che interconnette reti di operatori diversi (Autonomous System, AS). L'interconnessione tra operatori all'interno di un IXP, detta peering, viene realizzata tramite VLAN condivise tra gli ISP collegati (peering pubblico) oppure tramite VLAN dedicate tra coppie di ISP (peering privato). Le dinamiche commerciali legati al *peering* verranno discusse in successivamente. Un *IXP* può essere un "operatore neutrale" (*neutral IXP*) oppure un'azienda privata. In Europa i *neutral IXP* sono quelli prevalenti. Si tratta di organizzazioni no-profit in cui gli operatori sono parimenti membri della società, ciascuno con un voto. Le decisioni tecniche e di governance sull'evoluzione dell'infrastruttura vengono stabilite collettivamente tramite riunioni. In questo modo non si creano conflitti di interesse e l'accesso alla rete internet è parimenti concesso a tutti gli operatori membri di un *IXP*, indipendentemente dalla potenza politica o commerciale dei singoli operatori.

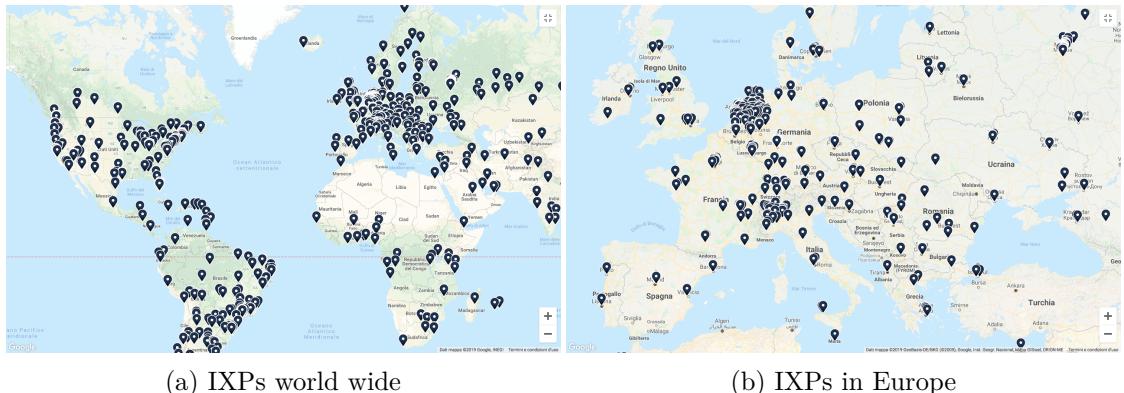


Figura 9: Internet Exchange Map by TeleGeography (2019)

In Italia il maggior IXP è il *Milan Internet eXchange (MIX)* raffigurato in figura

[10\(a\)](#). Il *MIX* è in grado di collegare in *peering* 257 *ASNs* dislocati in tutto il mondo, con una prevalenza di *ASNs* italiani[\[14\]](#). Il *MIX* "noleggia" lo spazio ad una cinquantina di *carriers*. I *carrier* presenti nel *MIX*[\[14\]](#) potranno terminare le chiamate verso gli operatori italiani o far transitare le chiamate verso altri *ASNs*. Ciascun *carrier* cercherà di posizionarsi su più di un *IXP*, ognuno dei quali prede il nome di *Point of Present (PoP)*. Nel *MIX* sono presenti due carrier globali, Cogent e Interoute.

Dalle homepage dei rispettivi siti si può vedere la mappa di tutti i *PoP* dei due carrier, illustrata in figura [40](#). La lista dei maggiori *IXP*, circa una cinquantina, è disponibile online[\[13\]](#).

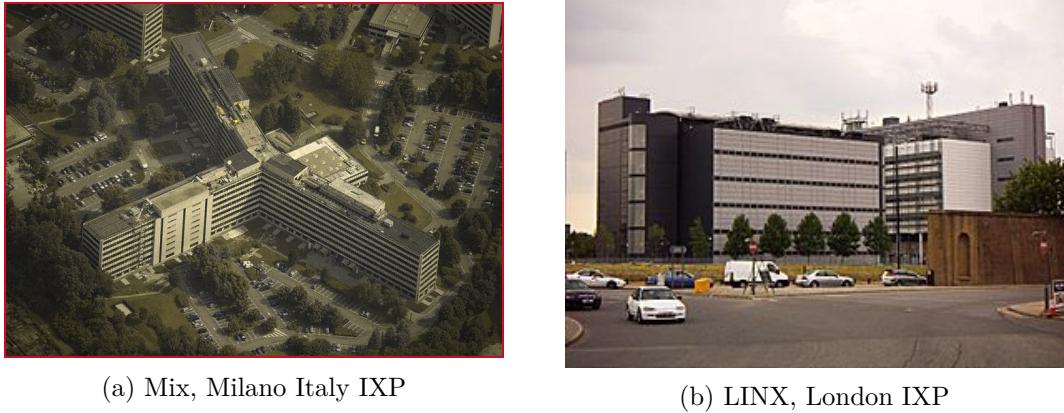


Figura 10: IXP infrastructures

Point of Presence (PoP). E' un punto di demarcazione artificiale che identifica un luogo di interconnessione tra un carrier internazionale *IXC* ed un area in cui può terminare una chiamata *LEC* (che verranno descritti più avanti). Alcuni *Long distance Carrier (IXC)* hanno migliaia di posizioni *POP* di solito situate presso Internet Exchange Point (*IXP*) e centri di colocation. Queste posizioni fisiche consentono a *IXC* di terminare le chiamate in quel posto e di fornire connettività vero il *local telephone network (LATA)* del posto. Tipicamente le infrastrutture fisiche dei *local Exchange carrier (LEC)* sono situate presso dei sistemi più "piccoli" dei centri di scambio internazionale detti central office.

2.2.1 Modelli di interconnessione nella rete Internet

Gli operatori internet sono suddivisi per *Autonomous System (AS)*. Ogni *AS* ha un numero che lo identifica e un blocco di indirizzi IP a lui riservati assegnati dai registri internet regionali, tra cui RIPE, ARIN, APNIC, LACNIC e AFRINIC.

L'interconnessione tra due *end-users* nella rete internet richiede un meccanismo di interconnessione che permette di attraversare molti *AS*. Esistono due tipologie di accordi commerciali tra *AS* chiamati *transit* e *peering*.

- **Peering:** Due o più *AS* si accordano tra loro per instradare il traffico, reciprocamente l'uno verso l'altro, senza pagare alcuna tariffa per i dati scambiati.
- **Transit:** l'operatore di un *AS* instrada il traffico ad un *transit provider* capace di instradare il traffico a qualsiasi altro *AS*. Il *transit provider* prende una percentuale sul traffico scambiato.

Il traffico di *transit* richiede di pagare un "tassa di pedaggio" al *transit provider* proporzionale ai dati trasferiti. L'instradamento tramite un *transit provider* spesso risulta una scelta obbligata perché tramite *peering* non sempre è possibile raggiungere l'*AS* di destinazione.

Gli operatori, per risparmiare le spese di transito, sono incentivati a effettuare accordi di *peering*. Tipicamente gli operatori si interconnettono in grandi edifici, detti *Co-Location*. Quando il numero di operatori presenti diventa abbastanza elevato questi luoghi prendono il nome di *Internet Exchange Point (IXP)*.

Il numero elevato di operatori (*AS*) connessi in un *IXP* attrae gli operatori di transito che creeranno accordi di *peering* (pagando) con gli operatori presenti e venderanno il servizio di transito ad operatori di transito più estesi. L'aumento degli operatori di transito causerà un abbassamento dei costi di transito e, a cascata, un aumento di operatori interconnessi in quel *IXP* per l'elevato numero di *peers* ed i bassi costi di *transito*.

In Europa questo fenomeno è molto accentuato. Praticamente quasi tutti i medi e grandi operatori internet sono presenti nei *IXPs* di Amsterdam, London, Frankfurt, and Paris. I piccoli operatori, a causa del costo molto basso delle tariffe di transazione, non sempre sono presenti negli *IXP* perché il costo delle tariffe di terminazione è inferiore al costo per l'affitto di spazio e apparecchiature nel *CO*.

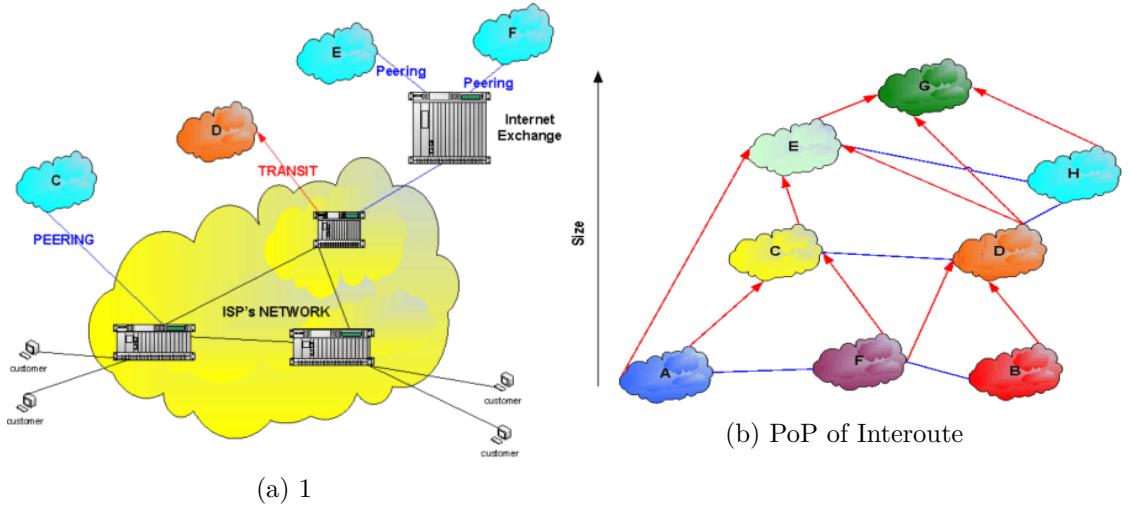


Figura 11: source: Arstechnica[16]

Le politiche di *peering* e di *transit* portano gli operatori nello scenario in figura 11 (b). Gli operatori di transito più grandi verranno utilizzati, quindi pagati, dagli operatori di transito più piccoli o dagli altri operatori. Questi operatori, che possono instradare l'informazione verso praticamente tutti gli *AS* dislocati nel mondo senza dover pagare i costi di transizione, sono detti operatori *T1*. La lista di questi operatori *T1* è disponibile online[15]

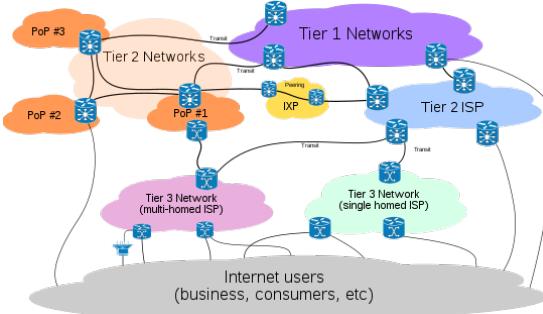


Figura 12: Tier

| | | Peering | Transit | RRMS | | |
|---|--------------------------------|-----------------------------|----------------------------------|---------|------------------------------------|--|
| AT&T ^[14] | United States | 7018 | 2,228 | 410,000 | 660,000 ^[15] | AT&T Pe policy ^[16] |
| CenturyLink (formerly Level 3, Qwest, Savvis, Global Crossing, TW Telecom and Exodus) ^{[14][15]} | United States | 209 3356 3549 4323 | 1,888 4,076 2,538 2,028 | 750,000 | 885,139 ^{[16][17]} | North Am Internatc Level 3 F Policy ^[18] |
| Deutsche Telekom AG (ICSS) ^[18] | Germany | 3320 | 581 | ? | ? | DTAG Pe Details ^[19] |
| GTT Communications, Inc. (formerly Tinet & nLayer) ^[19] | United States | 3257 (4436) | 1,576 | 100,000 | 160,934 ^[20] | GTT Peet Policy ^[21] |
| KPN International ^[21] | Netherlands | 286 | 278 | 75,000 | 120,000 ^{[22](dead link)} | KPN Peet Policy ^[23] |
| Liberty Global ^{[23][24]} | United Kingdom ^[25] | 6830 | 777 | 500,000 | 800,000 ^[26] | Peering ^[27] |
| NTT Communications (America) (formerly Verio) ^[27] | Japan | 2914 | 1,714 | ? | ? | North Am |
| Orange (OpenTransit) ^[28] | France | 5511 | 181 | ? | ? | OTI peer |
| PCCW Global | Hong Kong | 3491 | 680 | ? | ? | Peering ^[29] |
| Sprint (SoftBank Group) ^[29] | Japan | 1239 | 392 | 26,000 | 42,000 ^[30] | Peering ^[31] |
| Tata Communications India Limited (Acquired Teleglob ^[31]) | India | 6453 | 724 | 435,000 | 700,000 ^[32] | Peering ^[33] |
| Telecom Italia Sparkle (Seabone) ^[33] | Italy | 6762 | 482 | 347,987 | 560,000 | Peering ^[34] |
| Telia ^[35] (Subsidiary of Telefónica) ^[34] | Spain | 12956 | 304 | 40,000 | 65,000 ^[35] | Peering ^[36] |
| Telia Carrier ^[36] | Sweden | 1299 | 1,664 | 40,000 | 65,000 ^[37] | TeliaSon Internatc Rivai ih |

Figura 13: Tier 1 list

2.3 La rete della telefonia su IP

La deregolamentazione del mercato e la ricerca di nuovi modi per abbassare le tariffe di terminazione del traffico telefonico portarono molte nuove compagnie ad esplorare i vantaggi dell'infrastruttura internet. Molti operatori di transizione iniziarono a bypassare il sistema di *accounting rate regime* sfruttando la rete internet o transitando tramite più di un operatore di carrier per minimizzare i costi di transito (sfruttando le diverse tariffe da un paese verso un altro). Si stima che dal 1998 circa la metà del traffico internazionale globale invece di transitare tramite i circuiti di *switching* della PSTN transitava sulle *private leased lines* della rete internet, in special modo nella rotta tra Nord America e Europa.

La rete internet, tramite la telefonica *VOIP*, ha cambiato il mercato della comunicazione a lunga distanza. I costi inferiori della trasmissione via internet [2.2](#) rispetto alla trasmissione via *PSTN* ?? hanno favorito la creazione di nuove compagnie, dette *VOIP wholesale provider*. Anche i vecchi carrier internazionali IXC stanno gradualmente adottando l'infrastruttura internet, tanto che non sarebbe errato dire che la *core network* dell'infrastruttura telefonica è prevalentemente basata sulla rete internet.

Mentre il sistema telefonico tradizionale era (ed è tutt'oggi) un sistema chiuso, storicamente governato dagli stati e limitato a poche grandi compagnie telefoniche, il sistema della telefonica su IP è un sistema aperto in cui anche le piccole e medie aziende possono entrare senza difficoltà. Ricordiamo che lo scenario di riferimento di questa tesi è la

telefonia a lunga distanza e conseguentemente il mercato dei *VOIP wholesale provider* che operano tramite la rete internet ma non sono operatori "internet", sono operatori "telefonici". Questa distinzione è importantissima e vale la pena ripetere il concetto. Gli operatori "VOIP" che operano solo e unicamente tramite la rete internet non sono oggetto di studio di questa tesi. Invece gli operatori "VOIP" che operano tra la rete internet e la rete *PSTN* sono oggetto di studio di questa tesi. Questi operatori si interconnettono e competono direttamente con gli operatori telefonici locali e internazionali.

Lo scopo di questo capitolo è quello di capire come questi nuovi operatori telefonici si interconnettono tra e loro e con gli operatori telefonici tradizionali per definire lo scenario oggetto di questa tesi, appunto lo scenario del mercato *VOIP wholesale*.

2.3.1 Modello di interconnessione nel mercato della telefonia IP

Il mercato della telefonia su IP è un mercato popolato di migliaia di operatori *Tier3* e *VOIP reseller*. In questa tesi focalizzeremo l'attenzione sul mercato *VOIP Wholesale* ed in particolare sugli operatori di transito e terminazione. In questo settore gli operatori *VOIP* vengono pagati da altri operatori telefonici, sia *VOIP* che tradizionali, per instradare la chiamata verso una certa destinazione. Il caso considerato di riferimento è quello di una chiamata che origina da un operatore mobile e termina su un operatore mobile di un altro paese. Tra i due operatori mobile un numero non precisato di operatori di transito partecipa all'instradamento della chiamata. Tra questi operatori si nasconde il frodatore oggetto di questa tesi.

VoIP transit. L'operatore *VoIP* di transito guadagna attirando traffico a tariffe vantaggiose rispetto alla terminazione diretta e rispetto agli altri operatori di transito. Come funziona? Diciamo che una persona vuole chiamare dal paese A al paese B. Compone un numero e il segnale passa attraverso l'antenna GSM dell'operatore nel suo paese. L'operatore di origine addebita 1 euro al minuto (ad esempio) e converte il segnale in *VoIP*. Quindi cerca una compagnia di transito, il *carrier 1*, che accetti di fornire traffico al paese B per 50 centesimi. L'operatore di origine guadagna 50 centesimi da 1 euro pagato dal proprio abbonato.

L'operatore di transito *carrier 1*, a sua volta, cede ulteriormente il traffico, rivolgendosi alla prossima compagnia, il *carrier 2*. *Carrier 2* ottiene 30 cent al minuto e quindi ha due modi. Può inviare traffico a un operatore di telefonia mobile locale nel paese B per 20 centesimi (tramite *PSTN*), ovvero guadagnare solo 10 centesimi oppure può rivolgersi a un operatore di terminazione *VOIP* che termina la chiamata alla tariffa locale per un minimo di 10 centesimi. Quindi la compagnia di transito guadagnerà 2 volte di più. Di conseguenza, l'operatore locale non riceve nulla e l'abbonato nel paese B riceve con successo una chiamata internazionale.

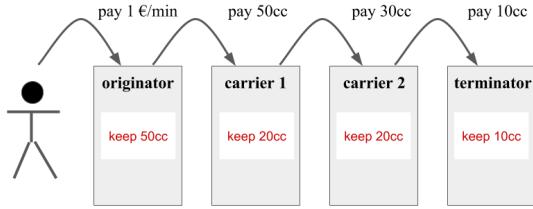


Figura 14: Pay chain

Voip terminator. Ad oggi, la terminazione del traffico vocale è una delle attività più redditizie (più del transito). Gli operatori di terminazione sono coloro che possiedono i *gateway GSM VoIP* (nel paese in cui operano). Grazie a questa apparecchiatura diventa possibile terminare le chiamate, cioè convertire il segnale VoIP ricevuto in formato GSM e pagare per la chiamata internazionale al costo della tariffa locale. Naturalmente per poter fare questa operazione in modo legale occorre che per legge sia permessa la conversione da VoIP a GSM. Inoltre può succedere che l'operatore *VoIP* non possegga direttamente le apparecchiature di rete ma le affitti da un terzo operatore che offre servizi di *SIP Trunking*.

Per semplicità consideriamo il terminatore come colui che possiede il gateway GSM. L'operatore che invia traffico al terminatore è chiamato operatore di transito. Gli operatori di transito guadagnano attirando quanto più traffico possibile da un paese all'altro (guadagnano una percentuale tipicamente del 5% sull'importo intero di una chiamata). Le rotte più profittevoli saranno molto richieste nel mercato e la concorrenza tra operatori porterà ad un continuo aggiustamento delle tariffe.

Questa pratica economica è chiamata arbitraggio. La compra-vendita dei minuti varia in funzione della domanda e dell'offerta. Gli operatori di terminazione che riescono ad avere la tariffa più bassa per una certa rotta (a parità di qualità di servizio) vinceranno sulla concorrenza attirando a se il tutto il traffico internazionale diretto verso quella rotta. I margini di profitto sulla singola chiamata variano da paese a paese e dipendono da molti fattori, ad esempio (1) il costo di terminazione sulla rete tradizionale, (2) la differenze fra tariffe locali e internazionali, (3) la differenze nelle tariffe in relazione al paese di origine, (4) il numero di chiamate dirette su quella rotta, (5) il numero di operatori concorrenti nel paese di destinazione e (6) perfino variazioni nello stato di congestione di chiamate su quelle rotte.

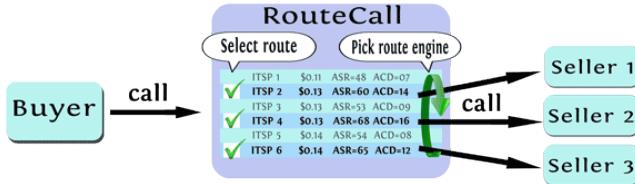


Figura 15: Sell and buy minutes *

* source: routecall.com

Tipicamente un operatore di terminazione, una volta definita la tariffa per terminare le chiamate tramite la propria rete, deve pubblicizzare la propria offerta affinché gli altri operatori invino il traffico internazionale tramite lui. In altre parole occorre un posto dove domanda e offerta si incontrino. Questo avviene o (1) tramite contrattazione privata via forum, pubblicando annunci di cerco/vendo [milioni di minuti] su rotte [+39,+485..] oppure (2) tramite piattaforme di scambio automatizzato come telecomsxchange.com, inaani.com..etc.

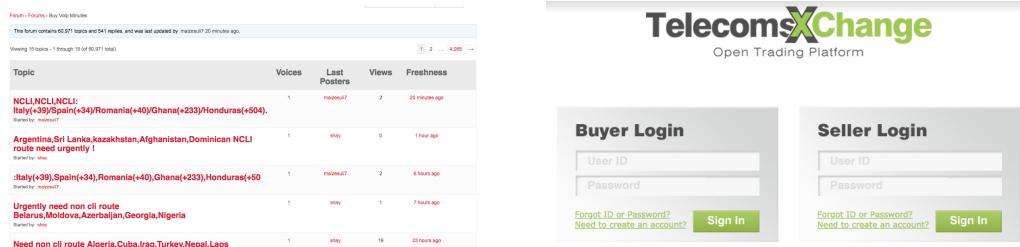


Figura 16: Demand offer mechanisms

Ai fini di questa tesi è importante sottolineare quanto segue. Nel primo caso, in virtù di un accordo tra operatori, la scelta dell'operatore successivo per una data rottura sarà piuttosto prevedibile (da 2 a 4 operatori partner per rottura). Diversamente nel secondo caso la scelta dell'operatore successivo avviene automaticamente tramite la piattaforma di scambio. L'operatore successivo verrà scelto tra le migliaia di operatori che nella piattaforma offrono una tariffa vantaggiosa per una data rottura. Inoltre i pagamenti in questo caso avvengono tutti in forma prepagata, chiamata per chiamata, e sono fatti dall'operatore verso la piattaforma e dalla piattaforma verso l'operatore successivo. Questo meccanismo favorisce l'interconnessione di operatori "volatili" ed è uno dei motivi per cui è difficile individuare gli operatori frondatori.

2.3.2 Least Cost Routing

Per quanto detto nel paragrafo precedente l'instradamento di una chiamata verso una certa destinazione è soggetto a cambiamenti frequenti nelle tariffe offerte dagli operatori di transito e di terminazione. Ad una prima impressione si potrebbe pensare che gli operatori di transito cerchino la tariffa più conveniente per massimizzare i profitti. Questo aspetto è certamente vero ma occorre sottolineare che la ricerca della tariffa più bassa (fissata la QoS) è una scelta obbligata per la maggior parte degli operatori di transito. Infatti il margine di profitto che un operatore di transito, chiamiamolo operatore X, ha

su una singola chiamata è estremamente basso. Se un operatore Y concorrente propone una tariffa più vantaggiosa per la stessa rotta, l'operatore X perde grossi volumi di traffico. La tariffa proposta da un operatore di transito è in funzione della tariffa che l'operatore deve pagare all'operatore successivo o all'operatore di terminazione più una commissione. E' evidente quindi che la ricerca del percorso a costo (economico) minore è una scelta obbligata per tutti gli operatori di transizione.

Prima che venissero implementati gli algoritmi *LCR* la ricerca delle tariffe più basse avveniva manualmente: i tecnici addetti controllavano le tariffe e modificavano manualmente le tabelle di routing. Oggigiorno tutto questo processo è automatizzato. In tabella 1 è raffigurato un esempio di tabella di routing. L'operatore di transito che deve decidere a quale operatore instradare la chiamata elenca in ordine di costo crescente gli operatori disponibili per la rotta richiesta (ovvero tutti gli operatori dichiarano di poter inoltrare la chiamata verso il codice operatore del numero chiamato). Supponiamo che gli operatori tra cui scegliere sia in ordine di priorità l'operatore A,B,C.

Tuttavia non è detto che la chiamata venga inoltrata all'operatore A. Infatti può succedere che l'operatore di transito abbia stipulato con l'operatore A un accordo per l'invio del traffico fino a 10 chiamate contemporaneamente. Se in quel momento ci sono già 10 chiamate attive verrà considerato l'operatore B. Anche in questo caso non è detto che l'operatore B sia pronto ad accettare la chiamata perché congestionato. L'operatore B notifica l'impossibilità a gestire la chiamata tramite un codice SIP 503. Infine l'operatore di transito sarà costretto ad inviare la chiamata tramite l'operatore più caro, ovvero l'operatore C.

Tabella 1: LCR roting table

| Destination code | Name | Cost Provider 1 | Cost Provider 2 | Cost Provider 3 |
|------------------|-----------------|-----------------|-----------------|-----------------|
| 1 | US / Canada | 5 | 5 | 6 |
| 1212 | US / New York | 6 | 5 | 6 |
| 44 | UK | 20 | 25 | N/A |
| 447 | UK mobile | 35 | N/A | 30 |
| 447766 | UK mobile Tele2 | 30 | N/A | 30 |

Questo esempio semplificato rende l'idea di come la selezione dell'operatore successivo nella catena di gestione della chiamata sia una operazione soggetto ad elevata variabilità. Ai fini di questo lavoro di tesi questo problema si traduce nella effettiva difficoltà a simulare una generazione di tracce telefoniche che riproduca esattamente quello che avviene nella realtà.

3 Problema

Esistono molte tipologie di frodi nel settore telefonico. Queste frodi si classificano in funzione di attori coinvolti, frodatori e vittime, servizi o tecnologie utilizzate, vulnerabilità o metodologie sfruttate per l'attacco. In questo lavoro di tesi le frodi studiate si classificano all'interno delle frodi nel settore *VOIP Wholesale*. Queste frodi riguardano le chiamate internazionali. Gli attori coinvolti in queste frodi sono lato frodatori gli operatori che agiscono da intermediari nell'instradamento della chiamata e lato vittime gli operatori di origine e di terminazione. Queste frodi sono spesso correlate al conteso normativo di paesi in cui la telefonia su IP è vietata per legge. Da un punto di vista tecnologico queste frodi sfruttano le vulnerabilità presenti nel protocollo di segnalazione *SIP*. In particolare i frodatori possono alterare le informazioni di segnalazione della chiamata, come la durata o il chiamante, per trarne beneficio economico. L'assenza di trasparenza tra operatori nella gestione della chiamata ostacola la rivelazione di queste anomalie e di conseguenza permette ai frodatori di perpetuare il meccanismo di frode per lunghi periodi.



Figura 17: Stima perdite annue per bypass

Per capire in modo più approfondito i meccanismi fraudolenti verranno analizzate le tre frodi più frequenti nel settore *VOIP Wholesale*:

1. frode FAS
2. frode LRN
3. call loop
4. frode Bypass

3.1 Le frodi

Frode FAS. La frode del *False Answer Supervision* è una frode che altera la durata di una chiamata per trarre beneficio dai profitti dati dalla tariffazione per la durata della chiamata. Esistono tre varianti di questa frode:

- Risposta anticipata: il frodatore aumenta la durata della chiamata in modo fraudolento rispondendo alla chiamata prima che il destinatario risponda. Per camuffare la connessione anticipata della chiamata viene registrato un suono che simula il tono di attesa normalmente gestito dai protocolli di segnalazione.

- Disconnessione ritardata: Il frodatore ritarda la trasmissione del messaggio di disconnessione dalla chiamata al chiamante facendo apparire all'operatore di origine una durata superiore a quella effettiva.
- Falsa risposta: il frodatore devia la chiamata ad un dispositivo di risposta automatica in cui è registrato un messaggio vocale del tipo "il cliente da lei chiamato non è al momento raggiungibile". Mentre nel caso normale il messaggio vocale viene gestito dai protocolli di segnalazione e non comporta fatturazione, nel caso del dispositivo di risposta automatica è come se la chiamata avesse ricevuto risposta. Questo ultimo caso è il caso più profittevole perché il guadagno del frodatore è definito dalla tariffa di terminazione che solitamente è molto più elevata rispetto alla tariffa di transito utilizzata nei primi casi.

frode LRN. La frode del LRN sfrutta la debolezza nelle ricerche sulla portabilità del numero chiamato. Infatti a seconda del numero chiamato le tariffe di terminazione possono variare. Ad esempio chiamare un numero di telefono in un area rurale costa molto di più che chiamare un numero in una area metropolitana. Questa differenza è particolarmente evidente negli Stati Uniti dove la differenza tra aree rurali e aree metropolitane è particolarmente accentuata. Normalmente quando un operatore di transito deve instradare una chiamata interroga il database LRN associato a quel numero per determinare il profilo tariffario corretto, che dipende come detto dall'area rurale o metropolitana associata a quel numero. Ogni volta che viene fatta una interrogazione l'operatore di transito paga un piccola tariffa al database. In alcuni casi l'operatore di transito riceve il numero LRN associato al numero nei campi opzionali della richiesta SIP. In questi casi per evitare costi aggiuntivi l'operatore di transito utilizza l'LRN specificato nel campo opzionale. Tuttavia se l'operatore di origine o uno degli operatori che lo hanno preceduto hanno inserito un LRN non corretto, associato ad un costo tariffario inferiore, l'operatore di transito cadrà vittima della frode LRN. Infatti l'operatore di transito fattura all'operatore che lo ha preceduto il costo associato al LRN non corretto, ma l'operatore di terminazione, che conosce l'LRN corretto, fatturerà all'operatore di transito il costo giusto (maggiorato) per quella area. Quindi l'operatore fraudolento spende meno e l'operatore di transizione perde la differenza tra il costo associato al LRN corretto e quello associato all'LRN fasullo.

Call loopin. I loop di chiamata possono verificarsi a causa di errori di routing o errori di configurazione, soprattutto quando gli algoritmi di routing utilizzano regole dinamiche per scegliere il percorso che ha una tariffa più bassa. Il frodatore può indurre la creazione di questi cicli per trarne profitto. Il frodatore, dicendo che la stessa chiamata è stata fatturata più volte, eviterà di pagare gli operatori successivi a lui ma non menzionerà il loop all'operatore da cui ha ricevuto la chiamata e per questo gli verrà pagata la tariffa di transito per un numero di volte uguali al numero di cicli effettuati.

Bypass fraud Una chiamata internazionale originata da una operatore di origine viaggia su più operatori di transito (intermediari) prima di raggiungere il paese di destinazione e l'operatore di terminazione. Ciascuno di questi operatori di transito riceve una percentuale sul prezzo pagato dall'operatore di origine. Alla fine della catena l'operatore locale nel paese di destinazione riceve una tariffa di terminazione per terminare questa chiamata internazionale sulla sua rete. Tipicamente l'instradamento della chiamata a livello internazionale avviene sulla rete internet, sia che la chiamata origini da un operatore *PSTN* sia che origini da un operatore *VOIP*.

La tariffa di terminazione di una chiamata internazionale può essere artificialmente alta, soprattutto nei paesi in cui a causa del monopolio le tariffe hanno prezzi elevati. Di fatto nei paesi in via di sviluppo i costi per la manutenzione di reti e infrastrutture sono piuttosto elevati e le tecnologie piuttosto antiquate. Per questo motivo le tariffe di terminazione delle chiamate cambiano da paese a paese. I paesi con le tariffe più elevate sono un obiettivo allettante per massimizzare i profitti nella frode di bypass. Inoltre questi paesi vietano l'uso della rete internet per terminare le chiamate, obbligando ogni operatore che deve terminare una chiamata in questi paesi a pagare le tariffe elevate all'operatore monopolista "di stato". La frode di bypass funziona instradando le chiamate tramite la rete internet, che come detto in precedenza, non è sempre possibile da un punto di vista giuridico in alcuni paesi di destinazione. La frode di bypass o di routing grigio, consiste nell'adottare dei gateway illeciti per evitare i gateway legittimi e le tariffe internazionali di terminazione. La frode di bypass porta a perdite finanziarie per l'operatore di destinazione e gli operatori di transito bypassati.

La frode di bypass può assumere molte forme a seconda del metodo utilizzato per raggiungere l'utente chiamato bypassando il gateway internazionale. Uno schema di bypass utilizzato sulle reti *GSM* è chiamato *SIM Boxing*. Questo tipo di frode consiste nel terminare la chiamata nel paese di destinazione utilizzando una *SIM* locale. Poiché la tariffa locale è molto più bassa di quella internazionale, il frodatore può raggiungere il chiamato facendo partire una chiamata dalla *SIM* locale. Un altro tipo di bypass relativamente recente viene eseguito terminando le normali chiamate internazionali nelle applicazioni OTT installate sugli smartphone dei destinatari.

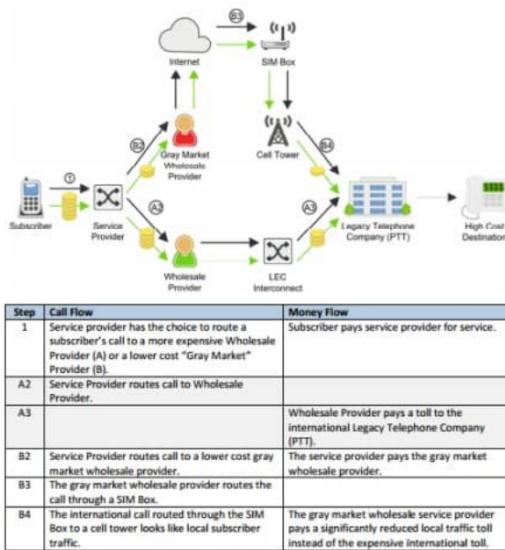


Figura 18: Parametri di valutazione del modello

3.2 I frodatori

Nel precedente paragrafo sono state illustrate alcune tecniche di frode. Tra le frode illustrate quella di bypass è certamente la più difficile da combattere. Un operatore fraudolento può attirare traffico di chiamate internazionali pubblicizzando basse tariffe

di terminazione delle chiamate e quindi terminare questo traffico su tali percorsi grigi. Nonostante la costante lotta contro il bypass di interconnessione, è ancora un problema irrisolto nelle reti telefoniche, con una perdita annuale stimata di \$ 5,97 miliardi [6]. Come mostreremo successivamente, il bypass OTT rende questo problema ancora più difficile.

Ci sono due attori principali coinvolti in questa attività: (1) i truffatori all'interno del Paese in via di estinzione; e (2) i vettori internazionali illegittimi da oltre confine. Il truffatore potrebbe essere un operatore di casella SIM, un operatore di rete locale o un titolare di una licenza di vettore nazionale. Il truffatore della casella SIM fondamentalmente imposta tutto: le caselle SIM, la connettività, la manodopera e le nuove forniture di SIM. Gli operatori di Local Loop, che introducono traffico illegale, possono utilizzare i loro switch al posto di una casella SIM. Ciò lo fa sembrare una chiamata locale che utilizza la propria serie di numerazione per interrompere il traffico sugli operatori mobili. I vettori nazionali possono immettere traffico illegale, modificare il numero "A" in modo da simulare il numero di loop locale per ogni chiamata e terminare lo stesso sugli operatori mobili sui propri tronchi nazionali anziché su tronchi internazionali. Le parti interessate sono gli operatori mobili, i vettori legali internazionali e il governo.



Figura 19: Beta todo

3.3 Rivelazione delle frodi

Gli approcci per l'identificazione delle frodi trovati in letteratura sono molteplici. In questo paragrafo illustrerò gli studi più significativi condotti nel settore.

Studi su CDR e reti neurali[6]. Lo studio è stato analizzato un CDR proveniente da una singola cella e contenente 234,324 chiamate. Nel record erano presenti 6415 abbonati, di cui il 33.14 % erano abbonati fraudolenti (SIMBoxes) e il restante 66.86% erano abbonati normali. Tramite l'uso di una rete neurale sono stati classificati gli abbonati fraudolenti e quelli normali. Il numero di falsi positivi è di 20 su 2126 ed il numero di falsi negativi è di 63 su 4289. I criteri usati per l'addestramento riguardano il numero di chiamate totali, la durata, la destinazione e la direzione.

Con un metodo di analisi simile è stato poi analizzato un dataset proveniente dal colosso AT&T in [7]. Il dataset conteneva 93,000 abbonati normali e 500 fraudolenti. I criteri usati per la classificazione, oltre a quelli elencati in precedenza, sono stati il numero elevato di IMSI per IMEI, la stazionarità della posizione GPS della SIM, il numero elevato di chiamate internazionali e la sproporzione tra chiamate in uscita ed in

entrata

• **Studi su analisi audio [8].** Altri studi si sono focalizzati sull’analisi in real-time dell’audio della chiamata. Questo studio permette l’individuazione di SIMBoxing nel 87% dei casi in 30 secondi di audio con zero falsi positivi. Il sistema di rivelazione usa tecniche di *fast signal processing* e deve essere posizionato sulle celle *BTS*.

Studi sul comportamento umano. [9] Altri studi focalizzano l’attenzione sul comportamento umano. In particolare usano un algoritmo di addestramento basato su un modello statistico, ad esempio le Bayesian network , per profilare l’utente e quindi rivelare gli utenti fraudolenti.

Studio su OTT in Europa[1]. Studio su frodi tramite *TCG* (test call generator) in paesi. In questo studio sono state analizzate più di 15,000 chiamate generate dal software di generazione delle chiamate di test. Su un periodo di 8 mesi. Al termine dello studio è emerso che il numero di chiamate soggette a *SIM Boxing* è di 17%. In questo articolo emerge anche un interessante osservazione sull’identificazione dell’operatore fraudolento che verrà analizzata in dettaglio in [3.4](#).

3.4 Rivelazione dei frodatori

L’unico approccio trovato in letteratura per l’identificazione dell’operatore frodatore è quello noto con il nome di *pinpointing bypassing operator*[1]. L’approccio è stato proposto nel contesto di una indagine sul fenomeno delle frodi di *OTT Bypass* (Over The Top Bypass) [4.3.1](#). In questa frode un operatore di transito internazionale fraudolento instrada la chiamata su una rotta non regolamentata, evitando così gli elevati costi di terminazione. Questo operatore è indicato dallo stesso articolo[1] come *bypassing operator*. L’approccio suggerito dagli autori per individuare il *bypassing operator* è quello di verificare la presenza della frode *OTT bypass* su una chiamata di prova generata tramite l’uso di un *TCG* (Test Call Generator) e di procedere per sospetti ed esclusione. Si considera come sospetto l’operatore successivo (*next-hop*) all’operatore che ha originato la chiamata. Si effettuano ulteriori n chiamate di prova che originano dall’operatore sospetto (sempre verso una stessa destinazione). Se viene rivelata la frode su una delle chiamate effettuate si scagiona l’operatore sospetto e si ripete l’analisi ripartendo dal primo operatore che ha inoltrato la chiamata fraudolenta. In questo modo si riesce a capire se il *bypassing operator* che ha causato la frode sulla chiamata in oggetto è il primo operatore verso cui è stata instradata la chiamata o se la colpa è di un operatore che si trova a valle nel percorso di instradamento.

Come sottolineato dagli autori dell’articolo il metodo proposto presenta troppe limitazioni. La corretta individuazione infatti richiederebbe che (1) le chiamate originate a partire dagli operatori ‘successivi’ vengano tutte instradate sulla stessa rotta, che (2) l’operatore si comporti ugualmente quando ha la funzione di operatore di transizione e origine, che (3) le rotte rimangano stabili nel tempo e che (4) sia sempre disponibile l’informazione sul *next-hop* . Oltre a questo la generazione delle numerose chiamate di prova necessarie al funzionamento di questo metodo sarebbe troppo elevata.

considerazioni. Dall’approccio proposto sopra emergono alcune osservazioni:

- **Conoscere i sospettati.** Per individuare il *bypassing operator* colpevole di aver attuato la frode del *OTT Bypass* occorre prima conoscerlo. L’unico modo per individuare gli operatori sospettati della frode è quello di ricostruire la catena

degli operatori *hop-by-hop*. La ricostruzione della catena degli operatori permette di avvicinarsi all’individuazione del *bypassing operator* ma non sempre è possibile a causa della mancanza dell’informazione sull’operatore successivo.

- **Confrontare le evidenze.(o gli operatori)** La colpevolezza del *bypassing operator* viene stabilita tramite il confronto con gli altri operatori sospettati. Non potrebbe essere diversamente, a meno di non avere la ’prova provata’ che lo dimostri (e.g un’indagine delle autorità preposte). Questo però porta a dover considerare un grado di incertezza sul risultato di colpevolezza.
- **Considerare l’analisi di sulla totalità delle chiamate.** La variabilità nell’instradamento lungo una stessa rotta scoraggia un approccio di identificazione basato sull’analisi di una sola chiamata (o di più chiamate sulla stessa rotta). Tanto maggiore è il campione dei dati analizzati tanto migliore sarà l’interpretazione degli stessi. Quindi un miglior approccio al problema dell’identificazione deve considerare la total

4 Soluzione

Nello scenario delle chiamate a lunga distanza descritto in 2 l’analisi a posteriori delle tracce consente agli operatori di capire se una chiamata è stata soggetta ad uno schema fraudolento oppure no. Individuare la frode è un passo significativo perché consente di analizzare il fenomeno ed implementare contromisure per combatterla. Tuttavia i meccanismi di mitigazione dell’attacco non sembrano essere sufficientemente efficaci nel combattere i frodatori. Infatti i frodatori, finché potranno continuare ad operare, troveranno sempre il modo di eludere le contromisure.

Individuare la frode non è abbastanza. Occorre individuare il frodatore.

Il sistema di telefonia globale è un sistema strutturato dove l’identità dell’operatore deve essere autorizzata da un sistema di autenticazione. Ciascun operatore possiede un codice univoco che gli viene rilasciato quando chiede la licenza ad operare. Per poter commettere una frode, il frodatore deve prima registrarsi nel sistema e ottenere tutte le autorizzazioni (offline e online). Naturalmente queste autorizzazioni possono essere rimosse o sospese se le autorità sospettano una attività fraudolenta da parte di un operatore.

Tuttavia l’attività di indagine da parte dell’autorità è una attività lenta e complessa. Inoltre, trattandosi di un panorama internazionale, è probabile che l’efficacia nell’individuazione del frodatore dipenda dal paese in cui l’autorità opera. Questo permette al frodatore di operare indisturbato per un tempo sufficientemente lungo a garantire un cospicuo guadagno.

Tutto questo evidenzia la necessità di ripensare un sistema di identificazione dei frodatori che possa meglio adattarsi al dinamismo e alla rapidità del sistema.

L’oggetto di questa tesi sarà quello di studiare un sistema di identificazione basato sulla analisi della reputazione degli operatori.

Si sottolinea che, in base alle ricerche effettuate alla data di pubblicazione di questa tesi, non sono state trovate pubblicazioni scientifiche riguardanti l’analisi della reputazione per l’individuazione dei frodatori nel mercato della telefonia internazionale.

4.1 Idea

Uno dei modi possibili per identificare comportamenti maligni in reti P2P è l’analisi della reputazione. La reputazione di ogni nodo viene costruita a partire dalle opinioni espresse dagli altri sulle base delle transazioni effettuate.

Nelle reti *P2P* di file sharing, come Torrent, i meccanismi di controllo della reputazione sono parte integrante della rete. La gestione della reputazione interessa sia gli utenti, detti *peers*, sia i contenuti inseriti dagli utenti, detti *file*. L’uso dei meccanismi di reputazione serve ad evitare la proliferazione incontrollata di *peers* e *file* maligni che porterebbe ad una rapida degenerazione del sistema rendendolo di fatto incompatibile con lo scopo per cui era stato creato.

Tuttavia gli stessi meccanismi di reputazione possono essere usati dai *peers* maligni per sovvertire il sistema, si parla di *malicious strategy*. La radice del problema sta nel fatto che non si può stabilire a priori la veridicità del feedback fornito da un *peer*. Ad esempio può succedere che un *peer* reputi maligno un altro *peer* che in realtà non lo è.

In questo caso occorre andare oltre la reputazione¹ e considerare la fiducia che i *peers* hanno sviluppato nel tempo. Lo studio di questi sistemi prende il nome di *Network of*

¹ Reputazione intesa come semplice media dei feedback espressi su un nodo

Trust.

Esistono molti modelli matematici che forniscono le formule per calcolare la fiducia di un *peer* sia dal punto di vista di un altro *peer* sia dal punto di vista dell'intera collettività dei *peers*. Il modello utilizzato in questa tesi verrà presentato in ??.

L'idea è considerare gli operatori di telecomunicazione in modo analogo ai *peers* di un rete P2P. La fiducia verrà costruita *call-by-call* valutando *hop-by-hop* tutte le transazioni avvenute verso gli operatori che hanno intermediato la chiamata, dall'operatore di origine fino all'operatore di terminazione lungo la direzione della chiamata. Ciascuna transazione è giudicata in modo negativo se la chiamata è stata soggetta a frode e positivo se la chiamata non è stata soggetta a frode. Le transazioni di tutte le chiamate di tutti gli operatori verranno memorizzate per un periodo di tempo al termine del quale verranno calcolate le reputazioni di tutti gli operatori.

Il concetto è esplicato in figura 20. I nodi raffigurati $\{A, B, C, D, E\}$ sono tutti operatori di transizione. In (a) sono evidenziate 4 chiamate che attraversano rispettivamente gli operatori $\{A, B, D\}, \{A, B, C, E\}, \{D, C, B, A\}$ ed $\{E, C, D\}$. Di queste sola la chiamata $\{E, C, D\}$ è stata sottoposta a frode. In (b) vengono mostrate le transazioni corrispondenti a quelle chiamate. Il dettaglio delle transazioni sarà chiarito più avanti. Si noti che le transazioni di uno stesso tipo sono cumulative, infatti sull'arco (A, B) in (b) sono presenti 2 transazioni positive. Inoltre si noti ciascun nodo avrà molteplici transazioni entranti ed uscenti frutto del fatto che le chiamate possono essere stabilite sia in una direzione che nella direzione opposta.

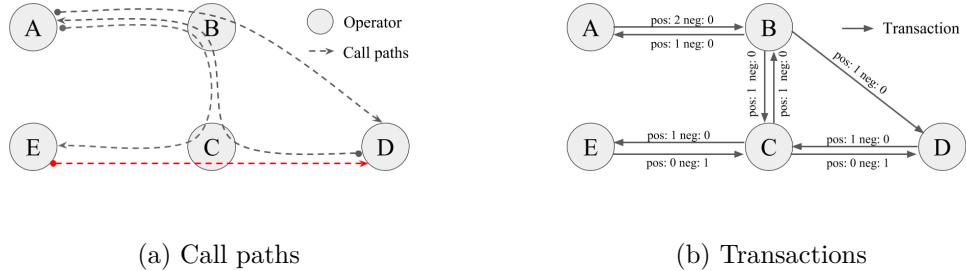
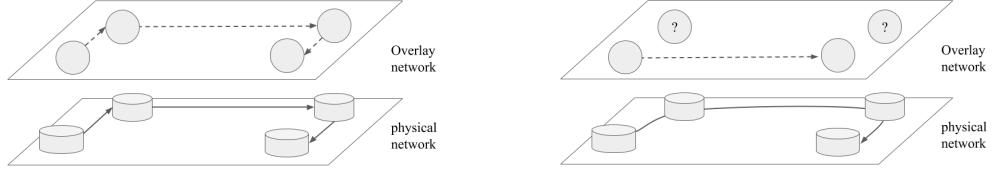


Figura 20: Transazioni generate dalla valutazione delle chiamate

Tuttavia esistono delle differenze tra la rete *P2P* e la rete degli operatori telefonici. In una rete *P2P* ciascuna transazione avviene direttamente tra due *peers*. I due *peers* possono esprimere un giudizio l'uno rispetto all'altro relativamente alla transazione che li ha coinvolti.

Nella rete telefonica ciascuna transazione (una chiamata) è intermediata da un numero non precisato di operatori che non è possibile conoscere a priori, come mostrato in figura 21.

La domanda che ci si pone è: *come posso attribuire un feedback ai nodi coinvolti nella transazione se non so neanche chi sono?*



(a) One-to-one correspondance in P2P model (b) Lack of knowledge in E2E model

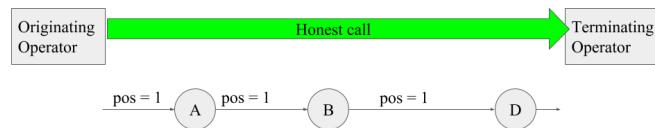
Figura 21: Mapping differences between physical and overlay transactions

La soluzione individuata è quella di far precedere l’analisi della reputazione ad una procedura di ricostruzione della traccia che identifichi tutti o parte degli operatori coinvolti in una chiamata.

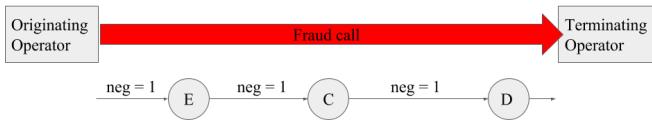
La procedura di ricostruzione della traccia ha il compito di capire quali operatori hanno partecipato a ciascuna chiamata e verrà presentata in 4.3. Non possiamo considerare di conoscere esattamente tutti gli operatori di una chiamata perché questo non sarebbe realistico, ma questo aspetto verrà trattato più avanti.

Una volta individuati gli operatori coinvolti si può ricondurre l’analisi della reputazione ad una analisi tramite *network of trust* in reti *P2P*. Infatti conoscendo i nodi coinvolti si potrebbe pensare di scomporre la transazione *E2E* in n transazioni. La prima transazione sarà quella dall’operatore di origine al primo operatore di transito. La seconda transazione sarà quella dal primo operatore di transito al secondo operatore di transito e così via fino all’operatore che precede la terminazione. L’esito della transazione, positivo o negativo, sarà stabilito sulla base della transazione *E2E* confrontando la traccia vista dall’operatore di origine con quella vista dall’operatore di terminazione.

Il risultato finale, in riferimento all’esempio precedente, è mostrato in figura 22. La chiamata fraudolenta $\{E, C, D\}$ genera le transazioni negative ($neg=1$) in (E, C) e (C, D) , mentre la chiamata onesta in $\{A, B, D\}$ genera le transazioni positive ($pos=1$) in (A, B) e (B, D) . A queste transazioni si è poi aggiunta la transazione generata dall’operatore di origine (non mostrato in figura) verso il primo operatore di transazione della catena, rispettivamente (O, E) ed (O, A) .



(a) One honest call feedback to n positive transactions



(b) One fraud call feedback to n negative transactions

Figura 22: Cascade effects of one call feedback to n transactions

In questo modo abbiamo considerato n transazioni dirette tra due *peers* ma abbia-

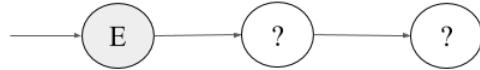
mo attribuito una transazione negativa anche ai nodi onesti che hanno partecipato a quella chiamata. Non si potrebbe fare diversamente. Non c'è modo di sapere a priori quale nodo è onesto e quale fraudolento, quindi non si può far altro che incolparli tutti.

Ricordando che l'obbiettivo finale è l'individuazione dell'operatore fraudolento, la domanda che ci si pone è: *come posso distinguere gli operatori onesti coinvolti indirettamente in una transazione fraudolenta dall'operatore frodatore (unico responsabile della frode)?*

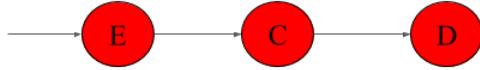
Considerando la totalità delle transizioni, sia positive sia negative, fatte su un arco di tempo sufficientemente lungo, il calcolo della reputazione dovrà marcare la differenza tra il comportamento degli operatori veramente fraudolenti e quello degli operatori onesti coinvolti in chiamate frodate. In teoria il fatto di considerare sia le transazioni positive che quelle negative dovrebbe mettere in evidenza il buon comportamento di un operatore onesto dal cattivo comportamento di un operatore maligno. In pratica solo la simulazione, che verrà presentata in ?? potrà evidenziare da un punto di vista statistico l'esito finale.

Tornando sull'esempio fatta in precedenza, l'immagine 23 ripercorre i passi salienti della procedura di identificazione. In particolare si nota che su C e D ho un numero maggiore di transazioni positive dovute alle chiamate oneste che li hanno coinvolti. Il diverso numero di chiamate positive incide provoca, a parità di $neg=1$, sul risultato della reputazione. Infatti E risulta avere una reputazione inferiore a 0.5 e per questo può essere indicato come il frodatore che ha causato la chiamata fraudolenta $\{E,C,D\}$.

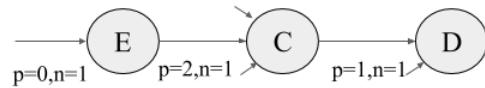
(0) No suspects for the fraud call. $\{E,C,D\}$



(1) All nodes are suspects of fraud. $\{E,C,D\}$.



(2) Monitoring global transactions.



(3) Identify the fraudster

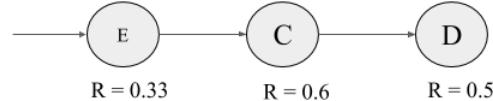


Figura 23: Passi verso l'identificazione dell'operatore frodatore

Riassumendo. L'idea consiste nel proporre una sistema di identificazione dei fraudolenti basato sulla cooperazione di operatori telefonici in assenza di fiducia (e.g accordi tra operatori). Il sistema si ispira ai modelli di *network of trust* in cui il nodo maligno è identificabile grazie alla sua cattiva reputazione. La reputazione è costruita in modo

dinamico dai nodi della rete sulla base delle interazioni scambiate e varia nel tempo in modo da riflettere l'effettivo comportamento del nodo.

Il sistema proposto si articola in tre distinte fasi. Nella prima fase occorre stabilire se la chiamata è stata soggetta a frode oppure no ed individuare gli operatori di transizione che hanno partecipato tramite un scambio di dati tra operatori a cui ci riferiamo come la **(1) procedura di ricostruzione della traccia**. Nella seconda fase l'elaborazione delle tracce ricostruite porta a definire le transizioni, positive o negative, tra i nodi della rete. L'aggregazione di tutte le transazioni porta a costruire la **(2) matrice globale delle transazioni**.

Nella terza e ultima fase viene calcolata la reputazione a partire dalle transazioni registrate nella matrice. In particolare se la chiamata non ha subito frodi, tutti i partecipanti verranno recensiti positivamente, se invece c'è frode allora verranno recensiti negativamente. Il valore di fiducia di un nodo verrà calcolato tramite il **(3) calcolo della reputazione**.

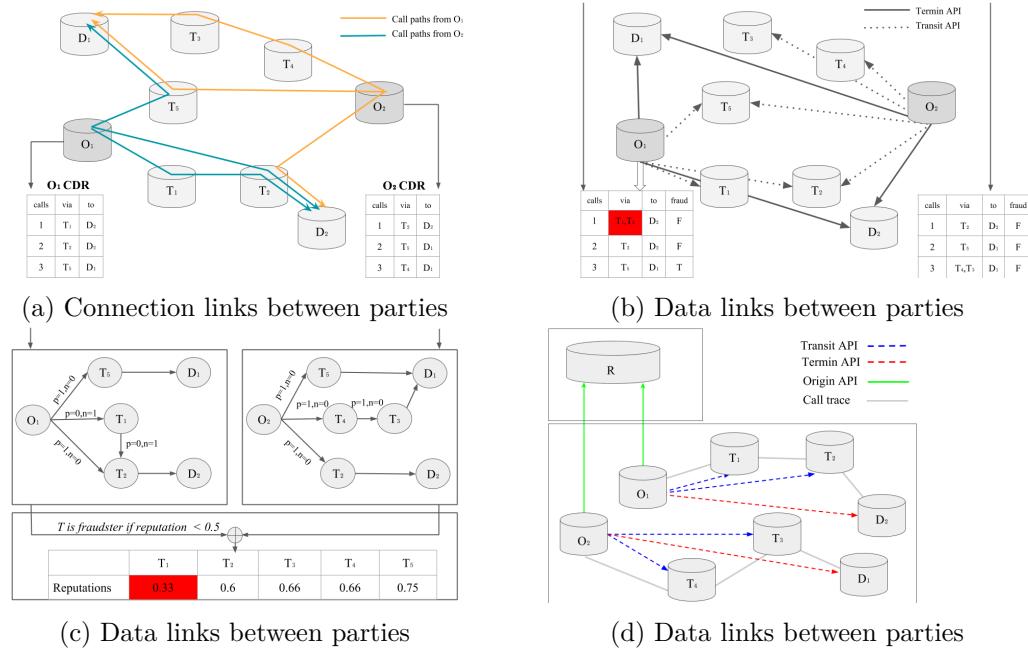


Figura 24: Paradigmi di comunicazione

4.2 Architettura

In questa sezione verrà descritta l'architettura del sistema di reputazione. L'architettura definisce l'infrastruttura fisica e logica che permette la memorizzazione, la trasmissione e l'elaborazione dei dati scambiati nel sistema.

L'architettura proposta è stratificata in (1) una rete *P2P* composta dai *data center* degli operatori e da (2) un server centralizzato per il calcolo della reputazione., vedi immagine 25.

Nel *layer P2P* gli operatori collaborano per la ricostruzione della traccia. La ricostruzione della traccia è la procedura con cui l'operatore di origine interroga gli altri per verificare se ha subito una frode su quella chiamata e per poter ricostruire la catena degli operatori che hanno instradato quella chiamata. La procedura di ricostruzione della traccia verrà discussa in fase di soluzione.

Al termine della procedura ciascun operatore che ha originato una chiamata ha una propria mappa delle transizioni. Questa pratica è molto usata nel conteso delle *Network of Trust*, tra cui è famoso l'algoritmo *Eigen Trust*. L'aggiornamento locale delle transazioni verrà discusso in seguito.

Infine i dati sulle transizioni locali, raccolti dai singoli operatori, vengono convogliati verso un server centrale per il calcolo della reputazione globale di tutti i nodi.

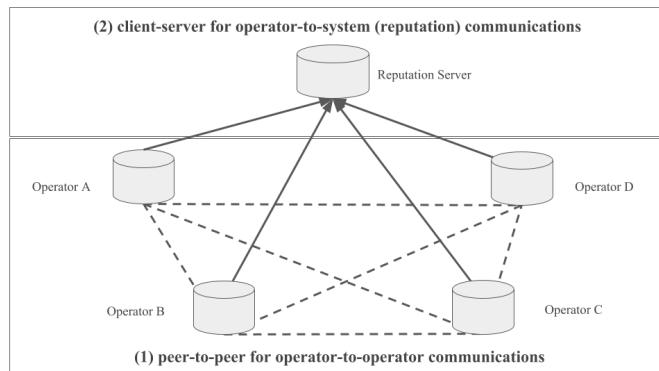


Figura 25: Topologia fisica del sistema

Questo approccio rende l'idea della topologia del sistema da un punto di vista hardware. Naturalmente i programmi ed i dati che si trovano presso l'operatore sono sotto il controllo dell'operatore stesso. Per garantire interoperabilità e differenziare l'informazione sono state definite tre API. Una unificazione e regolamentazione dei dati che gli operatori possono chiedere o cedere ad altri operatori. ..

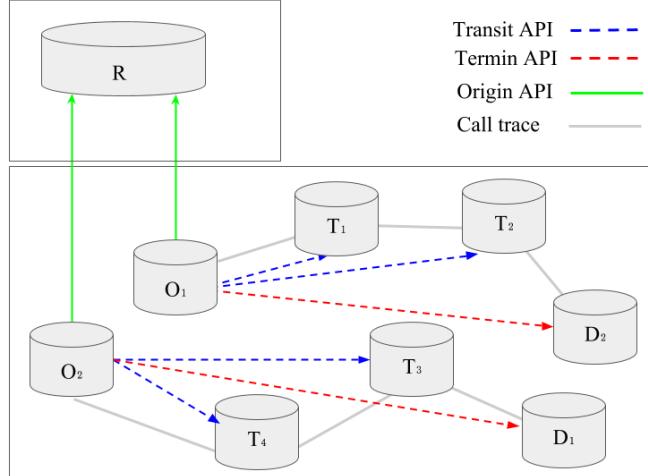


Figura 26: Data flows

Il pilastro portante del sistema di reputazione è da un lato la cooperazione tra operatori e dall'altro l'interazione tra operatori e sistema di calcolo della reputazione. La cooperazione implica uno scambio di dati.

Lo scambio di dati può avvenire in molti modi diversi a seconda della topologia e dello standard utilizzato.

Lo scambio dei dati deve soddisfare tre requisiti:

1. Non deve richiedere risorse di memoria e banda eccessivamente onerose.
2. Deve garantire l'interoperabilità tra operatori
3. Non deve poter essere usato dagli operatori per recuperare informazioni che danneggerebbero la concorrenza.

Il vincolo sulle risorse suggerisce di implementare una soluzione che minimizzi la ridondanza ed il trasferimento dei dati. I dati delle tracce telefoniche sono memorizzati da ciascun operatore nel proprio *data center*. Parallelamente occorre minimizzare l'impiego della banda. La soluzione adottata permette di implementare il servizio di scambio dati direttamente *on-site* presso il data center di ciascun operatore. **Il sistema sarà costituito da molti operatori interconnessi peer-to-peer e da un server centralizzato per la gestione della reputazione.**

L'interconnessione *peer-to-peer* degli operatori è una soluzione molto più vantaggiosa rispetto ad un modello completamente centralizzato i cui ciascun operatore comunica solo e unicamente con un server centrale. Infatti gli enormi volumi di dati che dovrebbero essere trasferiti dagli operatori verso un nodo centrale renderebbero il sistema costoso e non scalabile.

L'unico svantaggio dell'implementazione *peer-to-peer* è che non esiste un unico indirizzo IP a cui rivolgersi ma ne esistono molteplici, almeno uno per ogni operatore che partecipa al sistema. Per far comunicare un operatore A con un operatore B occorre un modo per dire ad A qual'è l'indirizzo IP a cui inoltrare la richiesta.

La soluzione più semplice consiste nell'associare il nome dell'operatore B al suo indirizzo IP. In particolare il nome di B dovrà essere lo stesso nome che A vede dalle tracce delle sue chiamate. Tipicamente un operatore si identifica o tramite il codice univoco

assegnatogli o tramite un account tipo *net1@operator_name.com*.

il vincolo sull'interoperabilità suggerisce di definire una interfaccia comune per l'accesso ai dati, anche detta *API*. Infatti nella parte *P2P* un operatore ha piena autonomia nell'implementazione del codice che estrapola i dati dal *CDR* e li rende accessibile a terzi operatori.

Inoltre, poiché uno stesso operatore può operare in ruoli diversi ed esporre dati di natura diversa, è necessaria una convenzione che formalizzi la scambio di dati tra operatori. La figura 26 mostra il flusso di dati tra diversi operatori che accedono alle API.

Il vincolo sulla riservatezza delle informazioni suggerisce di limitare la visibilità dei dati di una chiamata ai soli operatori che l'hanno instradata. Difatti gli operatori che hanno partecipato alla gestione di quella chiamata hanno già i dati che gli servirebbero a fini di concorrenza, ovvero la durata, i numeri chiamanti e alto. Un modo possibile che un operatore ha di esporre ad un altro operatore solo i dati delle chiamate che sono state gestite da quest'ultimo è l'uso di una *hash table* e di una interfaccia di *API*. La tabella di hash, un database del tipo *key-value pair*, deve essere costruita in modo tale che ogni operatore possa derivare le *keys* dai parametri noti della chiamata. Ad esempio eseguendo la funzione di hash su numero chiamata e timestamp già si avrebbe una un valore che tutti gli operatori che hanno gestito la chiamata possono recuperare velocemente. Gli operatori che non hanno questa informazione difficilmente riuscirebbero ad indovinarla per cui si può assumer un buon livello di riservatezza dei dati.

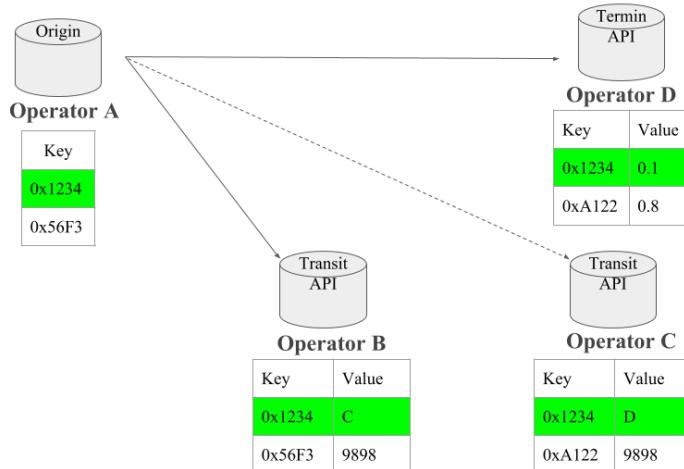


Figura 27: Architettura

4.3 Ricostruzione della traccia

L'assenza di trasparenza nell'instradamento della chiamata rende difficile, se non impossibile, individuare l'operatore fraudolento senza cooperazione tra operatori. Per quanto detto sull'opacità della catena nessun nodo da solo è in grado di vedere tutti gli operatori che hanno partecipato.

L'individuazione dell'operatore fraudolento richiede necessariamente di ricostruire la catena di operatori che dal primo all'ultimo hanno instradato la chiamata.

Negli accordi di servizio tra operatori è spesso definita la possibilità di condividere i CDR per indagare l'origine di una frode su una rotta, qualora quella rotta sia stata segnalata dai sistemi di rivelazione delle frodi.

Infatti l'unico modo per individuare il frosatore è far collaborare gli operatori ricostruendo a posteriori la catena di operatori che ha gestito la chiamata.

Lo stesso metodo di condivisioni delle informazioni può essere usato per confermare o smentire la presenza di frode in una chiamata internazionale.

Si consideri il CDR raffigurato in ???. I campi del CDR del operatore di origine indicati con il punto interrogativo rappresentano i dati su cui l'operatore di origine non ha la certezza che siano veritieri oppure che sono del tutto sconosciuti all'operatore di origine. Diciamo che un dato è veritiero per un operatore se la natura del dato dipende interamente dall'operatore stesso. Contrariamente diremo che un dato è sospetto se l'operatore ha ricevuto quel dato da una terza parte di cui non si fida.

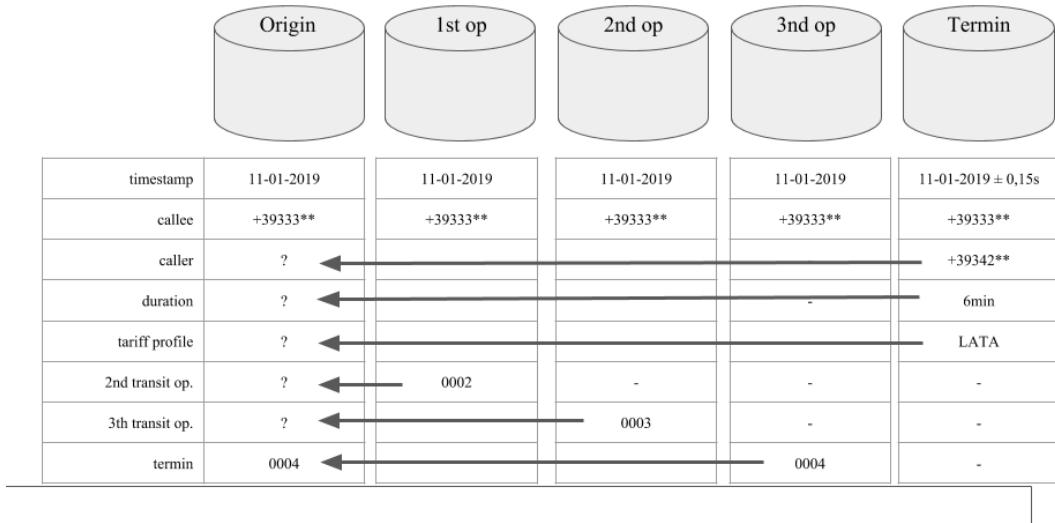


Figura 28: Parametri di valutazione del modello

La figura 28 mostra i campi del CDR che contengono dati sospetti, tra questi si evidenzia:

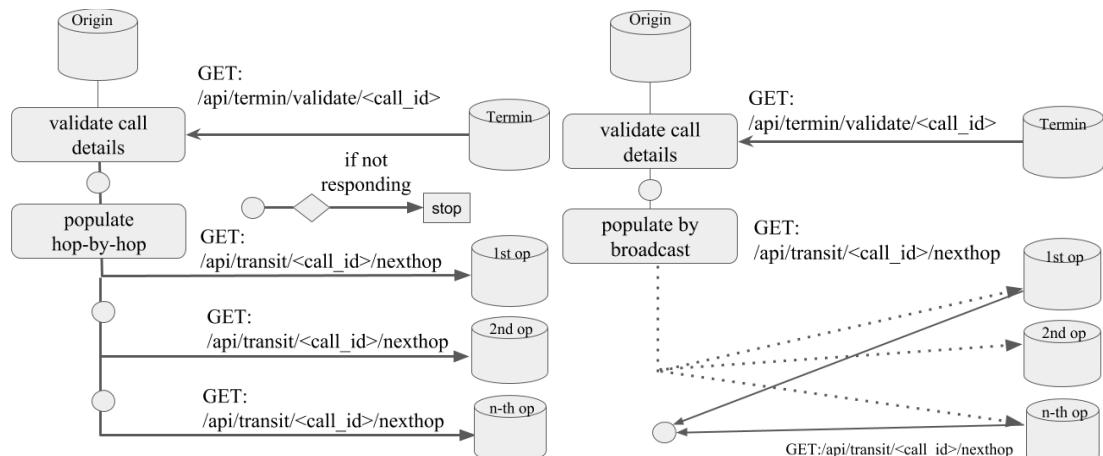
- La lista di tutti gli operatori di transizione , tranne il primo.
- La tariffa utilizzata per terminare la chiamata.
- L'effettiva durata della chiamata.
- Il numero chiamante visto dal chiamato.

Le frecce mostrano i messaggi scambiati tra operatore di origine, operatore di terminazione e operatori di transito necessari per ricostruire la traccia delle informazioni sospette o mancanti nel CDR dell'operatore di origine.

La traccia ricostruita al termine della procedura è mostrata in tabella 2.

Tabella 2: Traccia ricostruita

| | field | value | obtained from |
|----|------------------------|---------------------------|------------------------------------|
| 1 | Trace ID | 0x1234 | Hash(2, 4) |
| 2 | Timestamp | 20183112235901 | CDR |
| 3 | Caller | +1-415*** | CDR |
| 4 | Callee | +44-20*** | CDR |
| 5 | Duration | 7 min | IP(7)/api/termin/id/1234/duration |
| 6 | Tariff | 0.3 €/min - interLATA | IP(7)/api/termin/id/1234/tariff |
| 7 | Termin op. (MCCMNO) | 235-03 (UK Broadband Ltd) | CDR |
| 8 | 1st-Transit op. (CICs) | 0772 (AT&T) | CDR |
| 9 | 2nd-Transit op. (CICs) | 0002 (Worldcom Inc.) | IP(8)/api/transit/id/1234/nexthop |
| 10 | 3th-Transit op. (CICs) | 0714 (Phonelink, Inc) | IP(9)/api/transit/id/1234/nexthop |
| 11 | 4th-Transit op. (CICs) | 5431 (Telus) | IP(10)/api/transit/id/1234/nexthop |



(a) Ricostruzione della traccia via polling hop-by-hop
(b) Ricostruzione della traccia tramite broadcast

Figura 29: Possibili modi che l'operatore di origine ha per ricostruire la traccia utilizzando le appropriate API

4.3.1 Individuazione della frode

Il confronto fra parametri (come la durata, la tariffa, il chiamante ed il campo LRN) visti dall'operatore di origine *A* e gli stessi parametri visti dall'operatore di terminazione *B* permette di rivelare almeno tre tipi di frode.

Lemma 1: individuazione di frodi LRN

Se il LRN visto da *A* è diverso dal LRN visto da *B* per uno stesso *CID*, allora *A* ha subito la frode del LRN.

Lemma 2: individuazione di frodi FAS

Se la durata della chiamata vista da A è maggiore della durata vista da B per uno stesso CID , allora A ha subito la frode del FAS.

Lemma 3: individuazione di frodi Sim Boxing

Se il CID visto da A è diverso dal CID visto da B per un chiamata che ha stessa chiamato e stessa ora, allora A ha subito la frode del Toll bypass (anche detta SIM Boxing).

Di queste tre tipologie di frodi, solo l'ultima costituisce un vero problema perché non esistono metodi efficaci di combatterla. Per questo motivo questo lavoro di tesi si focalizzerà maggiormente su questa tipologia. Tuttavia il lavoro svolto non è vincolato a nessuna frode in particolare e per questo motivo parlerò di frode in senso generico.

4.3.2 Ricostruzione parziale della traccia

Gli operatori che maggiormente risentono del problema sono l'operatore di origine, di terminazione e di transito (onesti) nell'ordine in cui sono stati elencati.

L'operatore di origine è il primo truffato perché paga per un servizio che non corrisponde al servizio atteso. L'operatore di terminazione può essere truffato a causa dei mancati guadagni sulle tariffe di terminazione. Infine l'operatore di transito può essere truffato per mancati guadagni dati dalle percentuali sulla gestione della chiamata.

La cooperazione tra operatori è naturalmente incentivata dalle perdite economiche causate dagli operatori fraudolenti. L'operatore di origine è considerato la vittima perché paga per un servizio che non è conforme. Anche l'operatore di terminazione e gli operatori di transizione onesti sono vittime della frode. L'operatore di terminazione perde il surplus di valore che avrebbe dovuto ricevere dall'operatore d'origine. Gli operatori di transizione onesti a monte dell'operatore fraudolento che opera la frode del bypass perdono gli introiti che avrebbero ricevuto se la chiamata non fosse deviata. Le basse tariffe offerte dagli operatori di transizione fraudolenti sono una concorrenza "sleale" per quelli onesti.

A fronte di un problema comune la logica vuole che tutti gli operatori onesti cooperino per individuare i frodatori, estirpendo il problema alla radice. Questo non accade. La collaborazione tra operatori è "clusterizzata" a seconda degli accordi e dei contratti stipulati.

L'ostacolo maggiore ad una collaborazione estesa a tutti gli operatori del settore è la concorrenza. Ciascun operatore custodisce gelosamente i propri dati (minuti fatturati, numero di clienti, rotte più redditizie..).

La procedura di ricostruzione completa della traccia è un'utopia. La ricostruzione della traccia deve tenere presente due aspetti (1) non tutti gli operatori onesti partecipano al sistema di reputazione e (2) gli operatori fraudolenti possono mentire quando gli viene chiesto chi sia l'operatore successivo.

Le domande che si pone sono:

1. *I quali casi la traccia non contiene tra i sospettati il vero frodatore?*

2. *I quali casi una ricostruzione parziale della traccia porta ad individuare il frodatore sbagliato?*
3. *In che modo le false informazioni fornite dal frodatore possono depistare la sua corretta individuazione?*

Per rispondere a queste domande occorre una quadro completo della situazione: E' necessario analizzare caso per caso tutti gli scenari possibili che si possono presentare, con particolare riferimento alla percentuale di partecipazione dei nodi onesti nel sistema di reputazione.

1. caso in cui tutti partecipano (operatori di terminazione 100% e operatori di transito 100%).
2. caso in cui partecipano tutti gli operatori di transito ma non tutti quelli di terminazione (operatori di terminazione < 100% e operatori di transito 100%).
3. caso in cui partecipano tutti gli operatori di terminazione ma non tutti quelli di transito

In relazione ai casi sopra è interessante vedere come si popola il vettore dei sospettati.

4.3.2.1 caso di depistaggio del frodatore: variazione della probabilità di identificazione Si considera la catena dei nodi di transizione $T = \{T_1..T_{N-1}\}$ che hanno gestito una chiamata. Si ipotizza che la chiamata sia stata soggetta a frode e che si abbia il 100% degli operatori di terminazione e il 100% degli operatori di transito. Si considera l'insieme $S \subseteq T$ dei sospetti generati per una certa chiamata a partire dalle dichiarazioni successive dei nodi partendo da $A \rightarrow T_1$.

$$\begin{aligned} &\text{while}(T_{i+1} \neq B \text{ || } T_{i+1} \text{ non risponde}): \\ &\quad T_i \text{ aggiunge } T_{i+1} \text{ a } S \end{aligned} \tag{1}$$

Se i nodi sono tutti onesti tranne uno allora l'insieme dei sospettati S trovati tramite 1 conterrà sicuramente il frodatore. **Quindi se la partecipazione degli operatori onesti è al 100% l'operatore fraudolento sarà sempre presente in S perché aggiunto dal nodo onesto che lo precedeva.** In ultimo si sottolinea che l'operatore fraudolento può non dire la verità sull'operatore che lo ha succeduto e questo comportamento influisce sul vettore dei sospetti S . In particolare:

1. il frodatore non risponde. $S = \{T_1..T_i, F\}$
2. il frodatore mente sul nodo successivo. $S = \{T_1..T_i, F, T_x\}$
3. il frodatore dice la verità sul nodo successivo. $S = \{T_1..F..T_{N-1}\}$

4.3.2.2 caso di frode non rivelata: nessun sospettato Nell'immagine a sinistra si mostra il caso in cui l'operatore di terminazione (blu) non rispondendo non valuta la presenza o meno della frode. Questo implica che quella frode non verrà considerata dal sistema.

Nell'immagine a destra si mostra il caso in cui lo stesso frodatore instrada la chiamata verso un operatore di terminazione che partecipa al sistema e di conseguenza lo rivela. Si osserva che tra le tecniche di elusione dei meccanismi di fraud detection per la frode di bypass discussi in precedenza vi è quello di cambiare spesso operatore. Infatti il

frodatore che cambia spesso operatore di terminazione da meno nell'occhio perché i singoli operatori agiscono separatamente.

Questo evidenzia che la possibilità che il frodatore vari terminazione è concreta e quindi l'identificazione tramite una parte degli operatori può ancora essere valida.

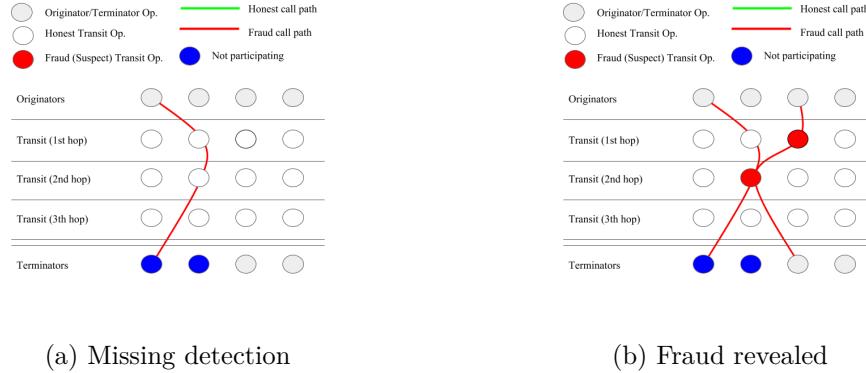


Figura 30: Fraud detection cases

4.3.2.3 caso di frodatore non inserito tra i sospetti: possibilità di erronea identificazione del frodatore L'immagine di sinistra mostra il caso in cui il nodo che non partecipa al sistema è proprio quello che precede il nodo frodatore indicato con una 'F'. Il vettore dei sospetti verrà comunque popolato con i nodi trovati perché il sistema non può sapere a priori che quei nodi sono innocenti.

L'immagine di destra mostra che i nodi innocenti incolpati ingiustamente sono discolpati di essere i frodatori perché nel bilancio tra transazioni positive e negative prevale il comportamento buono.

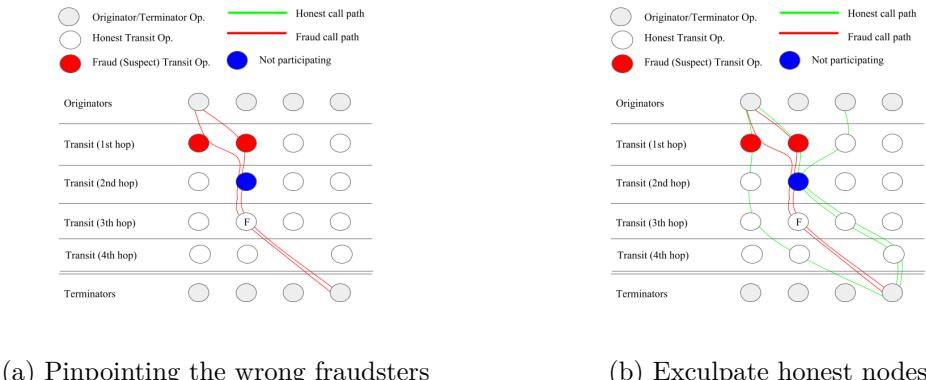


Figura 31: Fraudster identification errors

4.4 Gestione dei feedback

La ricostruzione della traccia, analizzata precedentemente, è necessaria per identificare i nodi coinvolti nelle chiamate e valutarne la presenza o meno di frodi.

In questa sezione verrà affrontato il passaggio dal dominio specifico della telefonia, in cui gli operatori si scambiano i dati relativi alle tracce delle chiamate, al dominio della analisi della reputazione, in cui gli operatori sono rappresentati come nodi e le transazioni tra nodi come archi di un grafo.

Un metodo usato nelle *Network of trust* è quello di gestire le transazioni aggregando feedback localmente e globalmente. Per un certo periodo di tempo un nodo registra l'esito delle transazioni che lo hanno coinvolto, positivo se la transazione è avvenuta con successo, negativo altrimenti. Successivamente, ad intervalli di tempo regolari, ciascun nodo comunica l'esito delle proprie transazioni ad un sistema globale centralizzato. Questo approccio in due tempi riduce la complessità di interazione tra il sistema di centrale ed i nodi della comunità.

I feedback localmente collezionati da un nodo verranno poi inseriti in un'unica matrice, detta matrice dei feedback. I valori presenti nella matrice dei feedback verranno poi usati in un secondo momento per calcolare la reputazione dei nodi.

Il sistema di reputazione oggetto di questa tesi utilizza lo stesso principio di *transazioni locali e transazioni globali*.

In questa sezione verranno illustrati i passaggi che, a partire dai feedback espressi dai nodi in certo momento, porteranno alla creazione di della matrice dei feedback che verrà poi utilizzata per il calcolo della reputazione.

1. **Generazione delle feedback.** Acquisizione delle tracce dagli operatori di origine e generazione dei feedback per i nodi coinvolti in ciascuna chiamata.
2. **Creazione della matrice dei feedback corrente.** Memorizzazione dei feedback relativi all'ultimo intervallo temporale in forma matriciale.
3. **Elaborazione a posteriori della matrice dei feedback.** Riduzione dei feedback negativi in transazioni per le quali è ragionevole pensare (sulla base di considerazioni legate allo scenario) che l'accusato sia un nodo onesto.
4. **Aggiornamento della matrice dei feedback.** Somma pesata dei feedback ottenuti in precedenza ai feedback ottenuti nell'ultimo periodo.

4.4.1 Generazione dei feedback

La rappresentazione *locale* delle transazioni viene generata nel sistema di reputazione a partire dai dati contenuti nelle tracce ricostruite dagli operatori di origine (*tramite origin API*) . Il concetto è illustrato in figura 32.

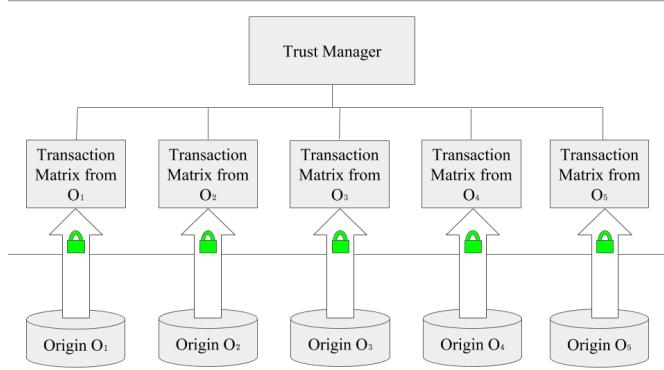


Figura 32: Topologia

Il senso dell’immagine è quello di mostrare il flusso di informazione. Dall’operatore di origine i dati delle tracce ricostruite sono trasferiti al sistema di reputazione. Dalla traccia vengono estrapolati i dati rivelanti al fine di valutare le transazioni, ovvero:

1. Il valore booleano che esprime l’esito, fraudolento o non fraudolento, di ogni traccia relativa ad una chiamata.
2. La lista ordinata dei codici degli operatori di transizione che hanno instradato la chiamata a cui la traccia fa riferimento.

Si sottolinea che i dati irrilevanti non devono essere trasmessi e che l’uso di *Origin API* permette al sistema di reputazione di ottenere solo i dati necessari.

I dati ottenuti permettono al sistema di reputazione di generare una rappresentazione *locale* delle transazioni.

Una rappresentazione tramite grafo orientato delle transazioni locali è illustrata in figura 33. Si sottolinea che la rappresentazione delle transazioni locali può avvenire tramite molteplici strutture dati. Questo aspetto non è stato studiato perché strettamente dipendente dalla tecnologia utilizzata e dal linguaggio di programmazione scelto.

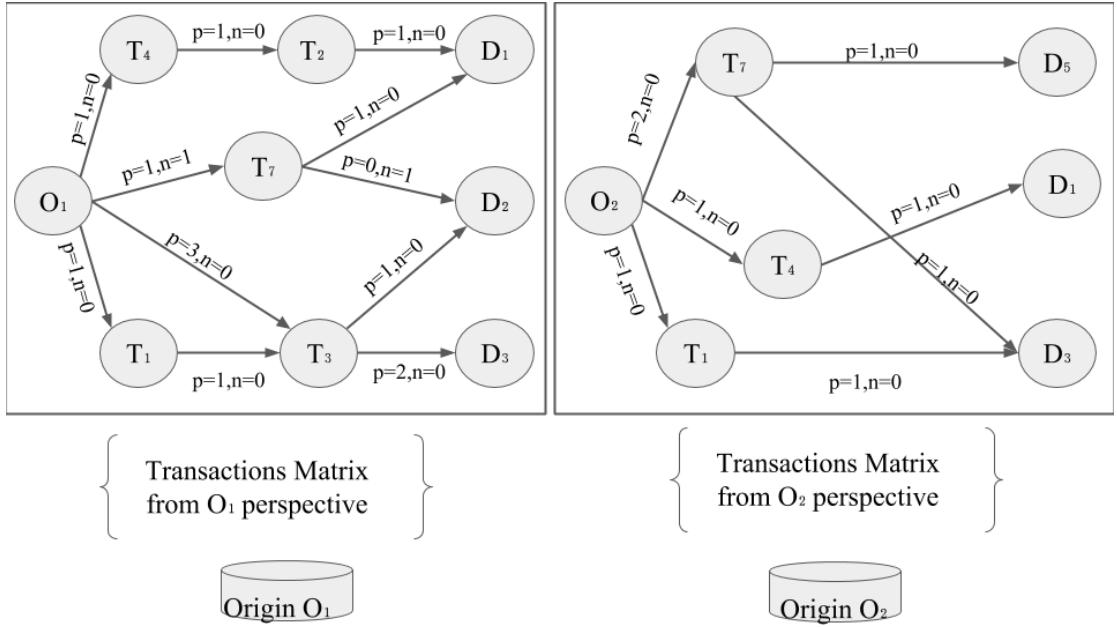


Figura 33: Grafo delle transazioni viste dall’operatore di origine in seguito alla procedura di ricostruzione della chiamata

4.4.2 Creazione della matrice dei feedback

Le transazioni tra i nodi del sistema sono memorizzate nella matrice M di dimensione $N \times N$, con N il numero di nodi del sistema. Per semplicità consideriamo che ogni nodo ha id uguale alla posizione sul vettore di dimensione N .

Quando un nodo i giudica positivamente una transazione verso un nodo j si incrementa il contatore $M[i][j] \rightarrow pos$. Invece se la transazione è giudicata negativamente si incrementa il contatore $M[i][j] \rightarrow neg$.

In questo modo scorrendo sulla riga i – esima si potrà vedere tutte le accuse che il nodo i ha rivolto ai nodi del sistema, mentre scorrendo sulla colonna j – esima si potrà vedere tutte le accuse che i nodi del sistema hanno rivolto contro il nodo j . Stesso discorso per le transazioni positive. La matrice M avrà la forma definita in 2.

$$M = \begin{bmatrix} n.a & \begin{pmatrix} pos = 10 \\ neg = 5 \end{pmatrix} & \dots & \begin{pmatrix} pos = 15 \\ neg = 7 \end{pmatrix} \\ \begin{pmatrix} pos = 3 \\ neg = 1 \end{pmatrix} & n.a & \dots & \begin{pmatrix} pos = 21 \\ neg = 8 \end{pmatrix} \\ \vdots & & \ddots & \vdots \\ \begin{pmatrix} pos = 10 \\ neg = 2 \end{pmatrix} & \begin{pmatrix} pos = 8 \\ neg = 6 \end{pmatrix} & \dots & n.a \end{bmatrix} \quad (2)$$

Più avanti in 4.4.3 verrà descritta una tecnica utilizzata per scontare le accuse in funzione di alcune considerazioni che riguardano la natura del sistema.

4.4.3 Elaborazione a posteriori della matrice dei feedback

La Procedura di sconto delle accuse nella rappresentazione locale delle transizioni viene eseguita tenendo in considerazione caratteristiche e comportamenti descritti che rimar-

cano le differenze nel comportamento dei nodi onesti rispetto ai nodi frodatori. In particolare si farà uso dei concetti relativi a:

- **Accordi esistenti tra operatori:** se due nodi hanno stipulato un accordo è improbabile che uno dei due si comporti in modo fraudolento verso l'altro. Questo aspetto verrà valutato nel calcolo della fiducia a priori in ??
- **Simmetria di accuse:** se due nodi si accusano a vicenda è improbabile che i due siano entrambi frodatori perché non né trarrebbero alcun beneficio economico.

Sconto per simmetria In riferimento allo scenario ?? si osserva che un nodo onesto tende a bilanciare le transazione da e verso un altro nodo al fine di minimizzare i pagamenti verso quest'ultimo. Al contrario un nodo fraudolento tende ad accentuare l'asimmetria delle transizioni da e verso il nodo 'pagante'. Il suo intento è quello di massimizzare i guadagni generati da un nodo nelle transizioni in entrata e di minimizzare i pagamenti dovuti a quel nodo nelle transizioni in uscita.

Lo sconto delle accuse reciproche tra due nodi qualsiasi avrà effetti significativi solo sui nodi onesti (a causa della simmetria tra le transazioni da/verso nodi onesti) mentre lascierà pressoché invariate le accuse verso i nodi fraudolenti. Il risultato è una diminuzione delle accuse rivolte ai nodi onesti. Lo sconto per simmetria, calcolato sulla matrice globale M , sarà:

$$\begin{aligned} d_{i,j} &= \min\{M[i][j] \rightarrow neg, M[j][i] \rightarrow neg\} \quad \forall j = 0..N-1, i = 0..j+1 \\ M[i][j] \rightarrow neg &= M[i][j] \rightarrow neg - d_{i,j} \\ M[j][i] \rightarrow neg &= M[j][i] \rightarrow neg - d_{i,j} \end{aligned} \tag{3}$$

4.4.4 Aggiornamento della matrice dei feedback con i valori precedentemente ottenuti

Ogni sistema di reputazione considera la reputazione passata. L'inclusione dei feedback precenti permette di rafforzare la reputazione dei nodi onesti e di individuare meglio i comportamenti maligni. Tuttavia il comportamento di un nodo può variare nel tempo. Per poter cogliere più rapidamente queste variazioni si introduce un valore scale, detto *forgetting factor*. Il *forgetting factor* è un fattore moltiplicativo che pesa meno le transazioni passate di quelle recenti. Il *forgetting factor* viene calcolato diversamente per feedback negativi e positivi. Questo meccanismo ricalca il principio per cui nelle relazioni di fiducia tra esseri umani una azione negativa richiede molte più azioni positive per essere dimenticata. Inoltre considerare lo storico delle transazioni ha il vantaggio di poter considerare in fase di computazione un numero di feedback superiore a quello acquisito nell'ultimo periodo (si tratta di una somma pesata). Per questo motivo il fattore di peso è stato scelto lineare. Fissato n , il numero massimo di feedback storici che vengono considerato nel calcolo, si ha che il feedback al tempo t_k è dato dalla somma

del feedback

$$\begin{aligned}
pos_{t_k} &= \sum_{p=0..n} post_{k-p} \lambda_p \alpha_{pos} \\
neg_{t_k} &= \sum_{p=0..n} neg_{k-p} \lambda_p \alpha_{neg} \\
&\text{con} \\
\lambda_p &= \frac{n-p}{n} \\
\lambda_p &= \frac{n-p}{n} \\
&\text{e} \\
\alpha_{pos} &= 0.5 \text{ se } p \neq 0, 1.0 \text{ se } p = 0 \\
\alpha_{neg} &= 1.0
\end{aligned} \tag{4}$$

4.5 Calcolo della reputazione

La matrice dei feedback esprime tutte le transazioni effettuate da un nodo verso un altro nodo in certo periodo di tempo. Idealmente se un nodo conosce per esperienza diretta tutti gli altri nodi non ci sarebbe bisogno di utilizzare un rete di fiducia. In pratica questo non accade perché (1) i nodi totali che fanno parte del sistema sono superiori di qualche ordine di grandezza ai nodi con cui mediamente interagisce un operatore e (2) a causa dell'algoritmo di *Least cost routing LCR* gli archi che rappresentano le transazioni instaurate tra nodi variano velocemente nel tempo. In sostanza occorre un meccanismo che permetta ad un nodo di derivare la fiducia in altro nodo che non conosce in modo diretto. Per questo scopo si usano gli algoritmi di *trust*. Esistono moltissimi algoritmi di trust ma la maggior parte sono delle varianti di due algoritmi: *EigneTrust* e *TNASL*. L'algoritmo che verrà usato in questa tesi è l'algoritmo *TNASL* acronimo di *Trust Network Analysis with Subjective Logic*.

4.5.1 TNASL

Subjective logic. La logica soggettiva è una base teorica usata nelle situazioni in cui non si ha certezza circa la veridicità di una affermazione. La metrica con cui si valuta una affermazione è chiamata opinione, *opinion*. Una opinione è una quaterna di valori, rispettivamente *belief*, *disbelief*, *uncertainty* e *base rate*.

$$\begin{aligned}
\omega_x^A &= (b, d, u, a) \\
&\text{con} \\
b, d, u, a &\in [0, 1] \\
b + d + u &= 1
\end{aligned} \tag{5}$$

L'opinione ω_x^A esprime la fiducia di A circa la veridicità della affermazione x . Il parametro b rappresenta la probabilità che l'affermazione x sia vera. Il parametro d rappresenta la probabilità che l'affermazione x sia falsa. La funzione di densità di probabilità con cui si modella b e d è chiamata *Beta density function* e verrà affrontata più avanti.

Il parametro u rappresenta la massa di probabilità che non è disponibile dalla prime due, misurando l'inabilità ad esprimere la veridicità dell'affermazione x . Sul piano della

probabilità $u = 1$ è un funzione densità di probabilità uniforme con larghezza infinita e altezza unitaria: Il valore $u = 1$ rappresenta la massima incertezza circa x , ovvero la totale inabilità ad esprimere un giudizio sulla veridicità della affermazione. Invece il valore $u = 0$ è un funzione densità di probabilità uniforme con larghezza infinitesima e altezza tendente all'infinito: il valore $u = 0$ può essere interpretato come il caso in cui l'opinione è costruita su un insieme infinito di evidenze e per questo diviene un "dogma".

Infine il parametro a esprime in che misura l'incertezza u debba essere considerata nel calcolo del valore atteso della probabilità circa l'affermazione x , ovvero $E(\omega_x^A) = b + ua$. Nel contesto delle *network of trust* il parametro a rappresenta la probabilità *a priori* che un membro della comunità ha nei confronti di un altro membro indipendentemente dall'evoluzione dinamica del sistema.

Beta probability function. La probabilità a posteriori di eventi binari può essere analizzata con la funzione di probabilità Beta. Questa funzione prende in ingresso due parametri α e β . La funzione Beta è esprimibile tramite la funzione di probabilità Γ .

$$f(p \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$0 \leq p \leq 1, \quad \alpha > 0, \beta > 0$$

con valore atteso:

$$E(p) = \frac{\alpha}{\alpha + \beta}$$
(6)

Si considera un processo con due possibili eventi $\{x, \bar{x}\}$ di cui si è osservato l'evento x per r volte e l'evento \bar{x} per s volte allora si può esprimere la funzione densità di probabilità in funzione delle osservazioni passate considerando:

$$\alpha = r + 1, \quad \beta = s + 1$$

Ad esempio con $r = 8$ e $s = 1$ si ha la funzione Beta illustrata in precedenza.

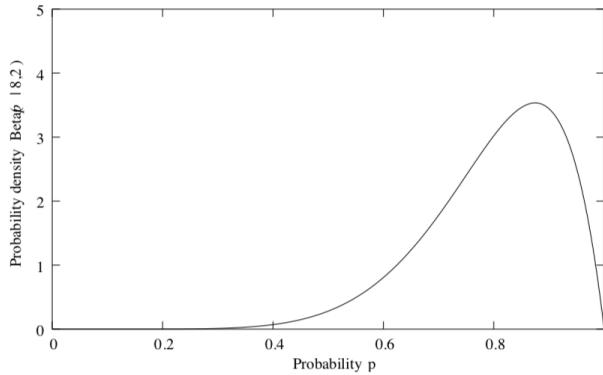


Figura 34: Beta PDF(8,2)

Nel caso della teoria *TNASL* i valori α e β vengono considerati anche in funzione del base rate a .

$$\alpha = r + 2a, \quad \beta = s + 2(1 - a)$$

Infine è possibile ricavare per via analitica la corrispondenza tra i valori dati dai feedback positivi e negativi, r e s , ed i parametri della corrispondente opinione (b, d, u, a) secondo:

$$\left\{ \begin{array}{l} b = \frac{r}{r+s+2} \\ d = \frac{s}{r+s+2} \\ u = \frac{2}{r+s+2} \\ a = \text{base rate of } x \end{array} \right. \quad (7)$$

Reputation score. Una volta nota l'opinione di A su x al tempo t_k si può calcolare il valore di reputazione di x visto da A in un valore scalare $Rep \in [0, 1]$ come:

$$Rep_{t_k} = b + au$$

Le reputazioni calcolate in tramite *TNASL* sono normalizzate tra zero e uno. In particolare le reputazioni ottenute al temine dell'analisi di un certo numero di transazioni permettono di classificare il comportamento di un nodo in onesto o fraudolento.

Lemma 4

Un valore di reputazione $r \in [0, 0.5)$ indica una cattiva reputazione e sarà associato con buona probabilità ad un nodo fraudolento. Un valore di reputazione $r \in (0.8, 1.0)$ indica una buona reputazione e sarà associato con buona probabilità ad un nodo onesto. Per valori di reputazione $r \in (0.5, 0.8)$ non si ha certezza sul comportamento onesto del nodo. Infine per un valore di reputazione di $r = 0.5$ non si può esprimere nessun giudizio.

La teoria di *TNASL* fornisce gli strumenti matematici per derivare la fiducia di un nodo A in un nodo B in modo indiretto sfruttando i rapporti di fiducia che A e B hanno con gli altri membri della comunità. Da un punto di vista matematico la teoria definisce due operandi chiamati *discounting* e *consensus*.

Discounting. L'operatore *discounting* permette di calcolare la fiducia transitiva. L'operatore è rappresentato col simbolo \otimes . Consideriamo il caso in cui A si fida di B e B si fida di C . Ci domandiamo qual'è la fiducia di A in C . Consideriamo l'opinione di A in B , $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ e l'opinione di B in C , $\omega_C^B = (b_C^B, d_C^B, u_C^B, a_C^B)$ e calcoliamo l'opinione di A in C , $\omega_C^{A:B}$

$$\left\{ \begin{array}{l} b_C^{A:B} = b_B^A b_C^B \\ d_C^{A:B} = b_B^A d_C^B \\ u_C^{A:B} = d_B^A + u_B^A + b_B^A u_C^B \\ a_C^{A:B} = a_C^B \end{array} \right. \quad (8)$$

Consensus. L'operatore *consensus* permette di fondere due opinioni, anche se in conflitto tra loro. L'operatore è rappresentato col simbolo \oplus . Consideriamo il caso in cui

A ha una opinione su C e B ha una opinione su C e calcoliamo l'opinione che A e B hanno di C , $\omega_C^{A \circ B}$:

$$\begin{aligned}
denom &= u_C^A + u_C^B + u_C^A u_C^B \\
&\text{if } denom \neq 0 \\
&\left\{ \begin{array}{l} b_C^{A \circ B} = b_C^A u_C^B + b_C^B u_C^A / denom \\ d_C^{A \circ B} = b_C^A d_C^B / denom \\ u_C^{A \circ B} = d_C^A + u_C^A + b_C^A u_C^B / denom \\ a_C^{A \circ B} = a_C^B / denom \end{array} \right. \\
&\text{if } denom = 0 \\
&\left\{ \begin{array}{l} b_C^{A \circ B} = b_C^A + b_C^B / 2 \\ d_C^{A \circ B} = 1 - b_C^{A \circ B} \\ u_C^{A \circ B} = 0 \\ a_C^{A \circ B} = a_C^A \end{array} \right.
\end{aligned} \tag{9}$$

4.5.1.1 Procedura per il calcolo della reputazione L'algoritmo *TNASL* offre gli strumenti matematici per elaborare la fiducia in una comunità. L'obiettivo finale è calcolare la reputazione di un nodo *target* vista da un nodo *source* al tempo t . Il calcolo di questa reputazione può essere effettuato in molti modi a seconda dei percorsi che vengono scelti. Questo problema prende il nome di directed series-parallel graphs (DSPG). Un DSPG è un grafo costruito da composizioni serie e parallelo. Il *path* ottimo lo si potrebbe trovare calcolando tutti i possibili percorsi da *source* a *target*. Poiché questo calcolo sarebbe troppo oneroso in questo lavoro di tesi ci siamo limitati a considerare i percorsi ad un solo *hop* che hanno verso il primo nodo una fiducia superiore a 0.8.

Naturalmente se il nodo *source* ha un'opinione diretta di *target* non c'è bisogno di utilizzare nessun algoritmo di trust.

Algorithm 1 Direct trust computation

```

1: if  $pos_A^B > 10 \parallel neg_A^B > 10$  then
2:   return  $\omega_A^B$ 
3: end if

```

Algorithm 2 Indirect trust computation

```

1: found  $\leftarrow F$ 
2: for  $i < N$  do
3:   if  $\omega_A^i > 0.8$  then
4:     found  $\leftarrow T$ 
5:      $\omega_A^T = \omega_A^T \oplus \omega_A^i$ 
6:      $\omega_T^B = \omega_T^B \oplus \omega_i^B$ 
7:   end if
8: end for
9: if found then
10:    $\omega_A^{T:B} = \omega_A^T \otimes \omega_T^B$ 
11:   STORE  $\omega_A^{T:B}$ 
12:   return  $\omega_A^{T:B}$ 
13: else
14:   return 0.5
15: end if

```

L'algoritmo 2 ricerca i nodi di fiducia di A e se non ne trova nessuno ritorna il valore di incertezza di 0.5. Per i problemi specifici dello scenario telefonico questo comportamento è preferibile rispetto al *DSPG* perché non favorisce i casi di falsi positivi. Infatti nel *DSPG* la computazione della reputazione può essere fortemente alterata dal numero di operazioni transitive necessarie a raggiungere quel nodo anche nei casi in cui le opinioni dei nodi coinvolti siano tutte positive. Questo fenomeno si ripercuote sul calcolo della fiducia in un nodo che potrebbe risultare arbitrariamente piccola anche nei casi di nodi onesti.

Si nota inoltre che dopo aver computato il valore finale che esprime l'opinione indiretta di A verso B si può memorizzare quella transazione. Col passare del tempo le opinioni che verranno memorizzate per via indiretta (consensus e discounting) si aggiungeranno a quelle inserite per via diretta causa feedback. Se ad esempio un nodo D volesse sapere la reputazione di un nodo B ed avesse come referenza solo A, A potrebbe aiutarlo perché conosce B.

```

1:  $N \leftarrow$  All nodes
2: if A has direct opinion over B then
3:   return  $opinion_A^B$ 
4: else
5:   for  $t < N$  do
6:     if A trusts t and t knows B then
7:        $opinion_A^T = opinion_A^T \oplus opinion_A^t$ 
8:        $opinion_T^B = opinion_T^B \oplus opinion_t^B$ 
9:     end if
10:   end for
11:    $opinion_A^B = opinion_A^T \otimes opinion_T^B$ 
12:   return  $opinion_A^B$ 
13: end if

```

4.5.1.2 Fiducia a priori per accordi esistenti tra operatori Gli accordi tra gli operatori nel settore delle telecomunicazioni sono un dato di fatto.

L'idea è quella ridurre considerare la fiducia a priori per modellare la situazione in cui un nodo T_i ed un nodo T_j hanno stipulato un accordo. Possiamo considerare la matrice $preTrust$ di dimensione $N \times N$ in cui il valore 1.0 in $preTrust[i][j]$ indica che i si fida a priori di j . Naturalmente un accordo è valido se ciascuna delle parti ne afferma l'esistenza, ovvero se $preTrust[i][j] = preTrust[j][i]$, ma la trattazione di questo aspetto verrà fatta in fase di implementazione. la matrice $preTrust$ avrà forma:

$$PreTrust = \begin{bmatrix} n.a & 0.5 & 1.0 & \dots & 1.0 \\ 0.5 & n.a & 0.5 & \dots & 1 \\ 1.0 & 0.5 & n.a & 1.0 & 0.5 \\ \vdots & & & \ddots & \vdots \\ 1.0 & 0.5 & 1.0 & \dots & n.a \end{bmatrix}$$

Il valore di $preTrust$ indicato nella matrice verrà utilizzato nel calcolo dell'opinione nel parametro a (base rate) discusso in precedenza.

4.6 Mitigazione della frode

Nel capitolo 3.1 è stato evidenziato che il fattore che maggiormente incide sull'impossibilità di trovare i *carrier* fraudolenti è l'uso del routing *LCR*. Lo studio del routing *LCR* evidenzia come sia possibile implementare automatismi che influenzano le scelte di routing. Inoltre l'aggiornamento delle tariffe tra operatori avviene in modo cooperativo. Ciascun operatore pubblica le proprie tariffe in formato *machine readable* in modo che chiunque possa importare quelle tariffe in modo dinamico nei software di gestione delle tabelle di routing.

In relazione a queste considerazioni e visti i problemi che derivano dalla mancata rivelazione dei frodatori si potrebbe pensare di valutare la reputazione dei nodi in modo analogo a come avviene la valutazione delle tariffe. La figura 35 esplicita il concetto.

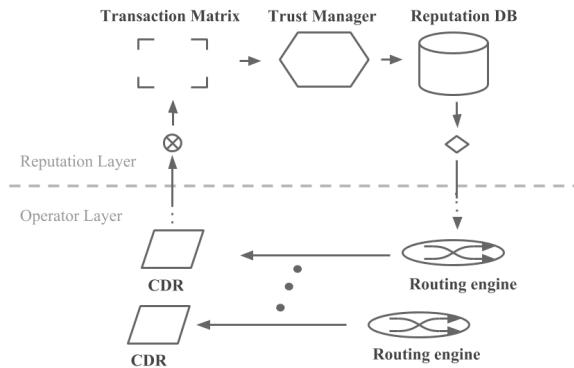


Figura 35: Best Reputation Routing

Il funzionamento di un meccanismo di routing basato sulla reputazione potrebbe essere implementato tramite *temporary blacklist*. Se la reputazione del nodo selezionato come *hop* successivo è inferiore a 0.5 il nodo viene inserito per 24h in *temporary blacklist*. Le valutazioni sull'uso di questo metodo verranno affrontate nel capitolo 6.2.4.

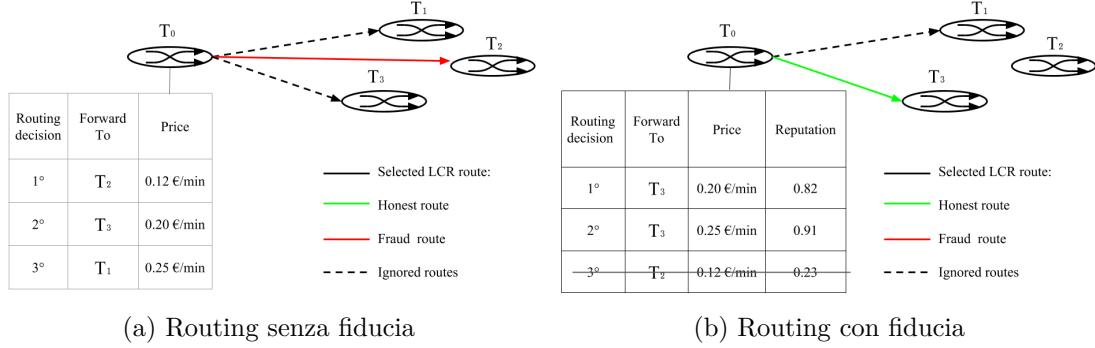


Figura 36: Routing secondo fiducia

4.7 Considerazioni

4.7.1 Contrasto alle strategie maligne dei frodatori

Una rete di trust è resa vulnerabile dal comportamento maligno di uno o più frodatori coalizzati. Le strategie maligne sono state raggruppate nelle categorie:

1. **Sybil attack**
2. **Malicious gateway**
3. **Disguised malicious**

Poiché gli attacchi sono strettamente correlati al problema dell'identità reale di un nodo, si considera una ulteriore classificazione in base al modo con cui un nodo fraudolento può essere inserito nel sistema:

- **Diretto.** Il nodo è registrato dall'operatore che lo controlla. L'identità reale dell'operatore verrà verificata offline e gestita attraverso un sistema di chiavi crittografiche che verrà approfondito in fase di implementazione. Il nodo potrà essere registrato o col ruolo di operatore di origineterminazione o col ruolo di operatore di transizione. Il nodo potrà essere registrato all'interno di un gruppo.
- **Indiretto.** Il nodo viene inserito in fase di popolamento del vettore dei sospetti 4.3.2 dal nodo onesto che lo precedeva nella catena di comunicazione. L'identità reale di questo nodo non è verificabile dal sistema. Inoltre il nodo non potrà essere classificato come operatore di origineterminazione, quindi non potrà accusare nessun altro nodo. Infine non potrà avere un gruppo di appartenenza.

Sybil - Identità indiretta8da modificare) Nelle reti di trust P2P la strategia indicata col nome di *sybil attack* sfrutta il fatto che la reputazione neutra di un nuovo nodo che entra a far parte della rete possa essere maggiore della cattiva reputazione che si era creato. In questo modo il nodo frodatore continua a far parte della rete sotto nuove identità e non viene mai eliminato definitivamente.

Questo problema non risulta particolarmente gravoso per lo scenario indicato in 2 perché l'identità dei nodi viene gestita dal sistema di autenticazione della rete telefonica. Il tempo ed il costo di attivazione di un nuovo nodo sono piuttosto onerosi.

Malicious gateway La strategia di *Malicious gateway* fa parte delle strategie di frodi collettive. Un operatore fraudolento usa un gruppo di nodi onesti sotto il suo controllo per attribuire feedback positivi ai nodi fraudolenti, anch'essi sotto il suo controllo.

Diversamente dal caso P2P, nello scenario E2E il ffordatore deve prendere il controllo anche dell'originator e del terminator che sono i nodi che controllano l'esito del feedback. Infatti nel caso E2E la generazione del feedback avviene in conseguenza di una chiamata. Poiché ogni chiamata ha un costo, l'unico modo che il ffordatore ha di sfruttare questo attacco è quello di inserire nel sistema una chiamata fittizia.

Disguised malicious Si tratta di una strategia che consiste nel fare in modo che il comportamento benevolo sia leggermente superiore a quello malevolo, ottenendo una reputazione bassa ma sufficiente a essere considerato un nodo buono. Per combattere questo attacco si può pensare di mantenere memoria della reputazione pregressa ed usarla per calcolare la reputazione corrente.

4.7.2 Onestà dei providers telefonici

Onestà dell'operatore di terminazione L'individuazione della frode tramite il confronto dei parametri fra l'operatore d'origine e quello di terminazione richiede che l'operatore di terminazione sia assolutamente attendibile. L'attendibilità sull'operatore di terminazione implica l'attendibilità sull'esistenza della frode. Per questo motivo è importante sottolineare che l'operatore di terminazione non ha interesse nel mentire:

- Non può ffordare A perché non può trarre vantaggi economici dalle frodi descritte in 4.3.1.
- Svolge anche il ruolo di operatore d'origine, quindi possibile vittima di ffordatori che colpiscono A .
- Nelle frodi di bypass è indirettamente colpito dalle frodi rivolte ad A a causa di anomalie sul traffico interno.

Onestà dell'operatore di origine L'operatore d'origine non può essere sempre considerato una fonte attendibile: Nel caso della frode del LRN il provider d'origine risparmia soldi a discapito del rivenditore. In teoria un operatore d'origine fraudolento potrebbe sovvertire il sistema di fiducia dichiarando fraudolente anche le chiamate che non lo sono e facendo così incolpare i rivenditori onesti che hanno gestito quella chiamata. In pratica gli operatori di origine e terminazione considerati in questa tesi sono compagnie telefoniche serie e strutturate che hanno rapporti contrattuali con gli intermediari a cui affidano il traffico. Per questo considereremo attendibile anche l'operatore di origine.

4.7.3 Considerazioni su scelte progettuali

La scelta di individuare la frode analizzando chiamata per chiamata non è la scelta ideale, ma è la scelta necessaria. Infatti il sistema deve gestire globalmente quasi (molte) chiamate al minuto e dovendo condividere i dati di ogni chiamata si avrà una mole enorme di dati da gestire.

Una diversa strada poteva essere quella di considerare i dati aggregati. Ad esempio invece di considerare la differenza di durata di ogni singola chiamata da A verso B si poteva considerare la differenza della somma della durata delle chiamate da A verso B. Nel caso esista una differenza la tratta A-B ha subito la frode del *FAS*. Allo stesso modo si poteva considerare di raggruppare i LRN dichiarando il numero di volte che ciascuno ricorre nella tratta A-B. Se i dati del nodo d'origine sono diversi dai dati del nodo di terminazione, nella tratta A-B è stata fatta la frode del *LRN*. Infine contando le chiamate che A vede verso B e le chiamate che B vede verso A si ha una indicazione sulla

presenza della frode di *Bypass*. Questa strada avrebbe portato ad una drastica riduzione dei dati coinvolti ma avrebbe reso praticamente impossibile identificare i frodatori

- Le chiamate aggregate nella tratta A-B sono state instradate tramite catene di nodi che non sono sempre uguali. Tra questi instradamenti solo alcuni saranno soggetti a frodi ma a causa dell'aggregazione si è perso il dettaglio di questa informazione.
- Affinché i nodi intermedi onesti possano comunicare la partecipazione ad una chiamata ed indicare il nodo successivo è necessario che conoscano l'id della chiamata in modo da confrontarlo con i propri dati.
- affinché un nodo intermediario onesto possa provare di aver partecipato ad una chiamata, e dunque avere diritto ad inserire informazioni per quella chiamata, è necessario un identificativo di chiamata. Diversamente non si ha modo di sapere se un nodo che inserisce informazioni su delle chiamate aggregate abbia realmente partecipato a quelle chiamate.

5 Validazione

L'uso di un simulatore è necessario per valutare numericamente i modelli proposti in 4. Per questo motivo è stato realizzato un simulatore in Python dedicato a questo scopo. Il simulatore è disponibile a <https://github.com/FrancescoErmini/FraudDetectorSimulator>

Occorre sottolineare che il simulatore è stato realizzato per validare l'idea proposta in questa tesi e che per questo motivo non è stato realizzato per favorire l'evoluzione e la riusabilità del codice.

Prima di scegliere di implementare un simulatore dedicato sono stati valutati due simulatori esistenti

- Omnetpp: <https://omnetpp.org/>
- p2p_sim: <https://rtg.cis.upenn.edu/qtm/p2psim.php3>

Omnetpp è un simulatore di rete che supporta il protocollo VOIP. Tuttavia la creazione di una rete di trust on-top in Omnetpp avrebbe richiesto una complessità inutile al fine di valutare gli algoritmi di trust.

Per questo si è valutato l'uso del simulatore p2p_sim, un simulatore dedicato a valutare le reti di trust. Tuttavia gli algoritmi di trust implementati nel simulatore sono creati per essere applicati in reti P2P dove un ciascuna transazione va da un nodo verso un altro. In una rete E2E le transazioni non sono dirette da un nodo verso l'altro bensì intermediate da altri nodi.

5.1 Uso del simulatore

Il simulatore viene configurato ed eseguito da linea di comando, come illustrato.

```
python3 computeTrust.py --providers 200 --intermediaries 400 --calls
    100000 --hops 4 --fraudsters 1 --frauds 5 --pcoop 100 --icoop 50
    --cycles 5 --scenario "esempioABC"
```

Mentre altri parametri cambiati con minor frequenza sono espressi come parametri del sistema nella classe *config.py*.

```
1 class TrustConfig:
2     fraudsters_camo = True
3     simmetry_strategy = True
4     use_tmp_blacklist = False
5
6 class TNASLsettings:
7     trustee_score = 0.8
8     cycle_deep_max = 10
9     pos_forgetting_factor = 0.1
10    neg_forgetting_factor = 1.0
11    pretrust_agreements = 4
12    use_pretrust = False
```

```

MBPdiFrancesco:FraudDetectorSimulator francescoermini$ python3 computeTrust.py --help
usage: TRACES GENERATOR [-h] [--providers PROVIDERS]
                         [--requirements requirements.txt]
                         [--intermediaries INTERMEDIARIES] [--calls CALLS] [--coop COOP]
                         [--hops HOPS] [--fraudsters FRAUDSTERS] [--frogs FROGS]
                         [--frauds FRAUDS] [--pcoop PCOOP] [--icoop ICOOP]
                         [--scenario SCENARIO] [--cycles CYCLES]

optional arguments:
  -h, --help            show this help message and exit
  --providers PROVIDERS
                        Number of local telco providers
  --intermediaries INTERMEDIARIES
                        Number of intermediary providers
  --calls CALLS
                        Number of calls
  --hops HOPS
                        Numer of hops per call
  --fraudsters FRAUDSTERS
                        Percentage of fraudulent intermediaries
  --frogs FROGS
                        Percentage of fraud calls.
  --pcoop PCOOP
                        Percentage of cooperation provider
  --icoop ICOOP
                        Percentage of cooperation intermediaries.
  --scenario SCENARIO
                        Name of the simulation directory
  --cycles CYCLES
                        Number of traces to simulate

```

Figura 37: Console usage

Al termine della simulazione l'output a console viene memorizzato in un file testuale sotto la carella indicata per quella simulazione. Il simulatore permette anche la generazione di grafici. Questa sezione è tuttavia lasciata all'implementazione dell'utilizzatore che può scrivere il codice necessario a rappresentare i risultati nella classe *Plot.py*.

5.2 UML Concettuale

In figura 38 è rappresentato il diagramma concettuale delle classi. Ciascuna classe ha un responsabilità.

- **Scenario.** Definisce il contesto telefonico, includendo il sistema di reputazione. Conosce gli operatori telefonici ed i ruoli che svolgono. Conosce il rate di chiamate fatte nel periodo di riferimento e la percentuale delle chiamate soggette a frode. In altre parole la classe scenario è un "descrittore della realtà nel contesto telefonico".
- **TraceGenerator.** Si occupa di generare le tracce delle chiamate, selezionando gli operatori che partecipano ad una certa chiamata. Rispetta i vincoli sul numero di chiamate generate in modo fraudolento imposti da *Scenario*.
- **TrustManager.** Simula l'architettura del sistema di reputazione. Si occupa di leggere le tracce generate da *TraceGenerator*, di creare la matrice dei feedback relativa all'ultimo periodo e infine aggiornare la matrice dei feedback con i valori ottenuti nei cicli precedenti.
- **TNASL.** Implementa le funzioni primitive della teoria *TNA-SL*, ed specifica come viene calcolata la reputazione (problema noto come *belief propagation* lungo il grafo delle relazioni tra nodi).
- **Classifier.** Classifica il comportamento di un nodo in base al valore di reputazione di quel nodo e conteggia i casi di falsi positivi, falsi negativi, mancata rivelazione e sospetta identificazione.

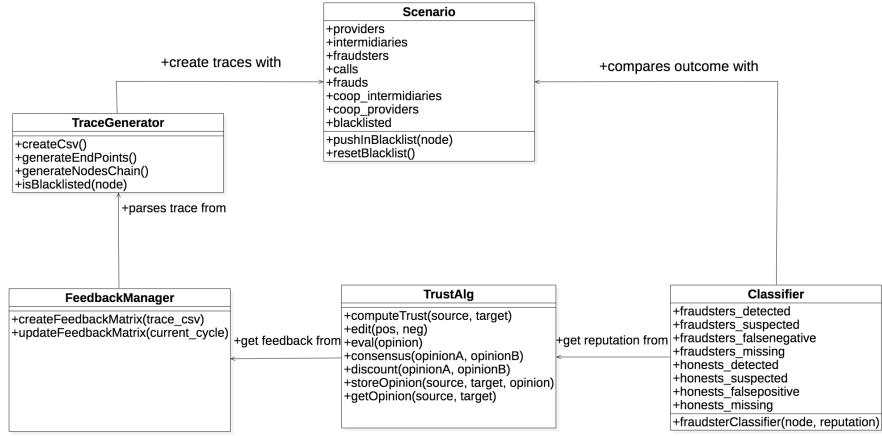


Figura 38: UML concettuale

5.3 Funzionamento

Il simulatore legge i parametri in input e crea uno *Scenario* caratterizzato da questi parametri. Ad ogni ciclo il simulatore crea una directory numerata e su questa crea ex-novo una base dati persistente accessibile tramite il formato binario hdf5.



Figura 39: Cycles

Il campo *Dataset* viene inizializzato tramie:

```

1  with h5py.File(self.dataset, "a") as f:
2      f.create_dataset("fback_matrix", shape=(N,N,2), dtype='uint16')
3      f.create_dataset("opinion_matrix", shape=(N,N,4), dtype='uint16')
4      f.create_dataset("trust_score", shape=(N,0))
5      f.create_dataset("fback_matrix_updated", shape=(N,N,2), dtype='float16')
  
```

Successivamente il sistema crea le tracce telefoniche csv relative a quel ciclo tramite:

```

1  traceGenerator = TraceGenerator(scenario=scenario)
2  traceGenerator.createCsv(file=trace_file)
  
```

Una volta ottenuto il file csv si procede a fare il parsing delle tracce. Si genera la matrice dei feedback corrispondente a quel periodo, e poi si aggiorna la matrice che tiene in considerazione i feedback del passato. Si osserva come le due matrici sono state mantenute disaccoppiate per necessità.

```

1  manager = TrustMan(scenario=scenario, dataset=dataset)
2  manager.createFeedbackMatrix(infile=trace_file)
3  manager.updateFeedbackMatrix(scenario_directory=scenario_directory, cycle=c)
  
```

Infine tramite *TNASL* viene calcolata la reputazione da tutti i nodi *source* verso tutti i nodi *target*.

```

1  trust = TNASL(scenario=scenario, dataset=dataset)
2  for j in range(len(sources)):
3      if scenario.isCoopProvider(sources[j]):
4          results[c][i] = trust.computeTrust2(sources[j], targets[i])
5          res = result.fraudsterClassifier2(targets[i], results[c][i])
6          if res and TrustConfig.use_tmp_blacklist: #is Fraudster
7              scenario.push_in_blacklist(targets[i])

```

5.3.1 Generazione dei nodi

I nodi sono identificati dalla posizione sul vettore dei nodi dal $[0, N_{\text{fraudster}}]$. Il ruolo che ogni nodo riveste dipende dalla suddivisione del vettore ed in particolare:

- Nodi di origine/terminazione: $[0, N_{\text{provider}}]$.
- Nodi di transito onesti: $[N_{\text{provider}}, N_{\text{provider}} + N_{\text{intermediaries}} - N_{\text{fraudsters}}]$.
- Nodi di transito fraudolenti: $[N_{\text{provider}} + N_{\text{intermediaries}} - N_{\text{fraudsters}}, N_{\text{provider}} + N_{\text{intermediaries}}]$

I nodi nelle varie categorie sono raggruppati in cluster di dimensione $N_{\text{clustersize}}$.

Tabella 3: Parametri di configurazione dei nodi

| | |
|------------------|----|
| n_provider | 4 |
| n_intermediaries | 32 |
| n_fraudster | 4 |
| n_cluster_size | 4 |

```

1  class Scenario:
2
3      def __init__(self, n_providers, n_intermediaries, n_calls, l_chain,
4                   fraudsters_percentage, frauds_percentage, provider_participation,
5                   intermediaries_participation, cycles):
6
7          def isIntermediary(self, index):
8              ..
9
10         def isFraudster(self, index):
11             ..
12
13         def isFraud(self, value):
14             ..
15
16         def isCoopProvider(self, index):
17             ..
18
19         def isCoopIntermediary(self, index):
20             ..

```

5.3.2 Generazione tracce

il file *TraceGenerator.py* si occupa di generare N_{calls} tracce, di cui N_{calls_fraud} soggette a frode, Ogni traccia setta l'id della chiamata al valore del contatore di chiamate. Le chiamate da $[0, N_{calls} - N_{call_fraud}]$ sono considerate oneste, mentre le successive malvagie.

La restante parte di generazione della traccia dipende dalla presenza o meno della frode:

1. senza frode: origin e terminator sono scelti su tutto l'insieme dei provider per favorire la simmetria. I l_{chain} intermediari sono scelti tra l'insieme degli intermediari onesti.
2. con frode: origin e terminator sono scelti su metà insieme rispettivamente per favorire l'asimmetria. Inoltre $l_{chain} - 1$ intermediari sono scelti tra gli onesti, mentre l'ultimo è scelto dall'insieme dei frodatori.

Ogni traccia è definita tramite formato *csv*.

Listing 1 Traccia di chiamata

```
1 id,fraud,origin,transit1,transit2,transit3,transit4,termin
2 0,0,107,145,217,563,503,593
3 1,0,73,16,239,365,310,243
4 2,0,199,39,440,423,288,586
5 3,0,51,13,473,250,325,469
6 ..
7 995,1,25,117,487,278,441,599
8 996,1,99,135,511,488,594,597
9 997,1,58,170,272,316,486,599
10 998,1,11,194,274,477,573,597
11 999,1,62,112,548,472,214,596
```

La generazione delle tracce segue le regole dette prima

5.3.3 Creazione della matrice dei feedback

La classe *TrustManager* si occupa di prendere in ingresso le tracce generate e di creare la matrice dei feedback. Il codice illustra il fatto che la mancata partecipazione di un nodo impedisce l'acquisizione del feedback, come descritto nel capitolo relativo alla soluzione.

```
1 for trace in traces:
2     if self.scenario.isCoopProvider(trace[Csv.TERMIN]) and not self.scenario.isFraud(
3         trace[Csv.FRAUD]):
4         for i in range(self.scenario.l_chain-1):
5             source = int(trace[Csv.TRANSIT+i])
6             target = int(trace[Csv.TRANSIT+i+1])
7             if self.scenario.isCoopIntermediary(source):
8                 matrix[source,target,POS] = matrix[source,target,POS] + 1
9             if self.scenario.isCoopProvider(trace[Csv.TERMIN]) and self.scenario.isFraud(
10                trace[Csv.FRAUD]):
```

```

10     for i in range(self.scenario.l_chain-1):
11         source = int(trace[(Csv.TRANSIT+i)])
12         target = int(trace[(Csv.TRANSIT+i+1)])
13
14     if self.scenario.isCoopIntermediary(source):
15         matrix[source,target,NEG] = matrix[source,target,NEG] + 1
16     if TrustConfig.simmmetry_strategy:
17         if matrix[target,source,NEG]>=1 and matrix[source,target,NEG]>=1:
18             matrix[source,target,NEG] = matrix[source,target,NEG] - 1
19             matrix[target,source,NEG] = matrix[target,source,NEG] - 1

```

5.3.4 Somma dei feedback precedentemente ottenuti

La classe *TrustManager* si occupa di creare ad ogni ciclo la matrice risultante dalla somma dei feedback ottenuti al ciclo precedente e dei feedback ottenuti ai cicli passati. Il codice illustra la scansione delle directory create per le simulazioni dei cicli precedenti, l'acquisizione dei feedback relativi a quel periodo ed infine la moltiplicazione dei valori trovati per il fattore di peso linearmente decrescente, diversificato nel caso di feedback positivi e negativi.

```

1 def updateFeedbackMatrix(self, scenario_directory, cycle):
2     curr_dataset_path = scenario_directory+'/'+str(cycle)+'/dataset.hdf5'
3     curr_dataset = h5py.File(curr_dataset_path, 'a')
4     '''inizializzo fback_matrix_updated con i valori acquisiti nell'ultimo
       periodo'''
5     curr_dataset['fback_matrix_updated'][:] = curr_dataset['fback_matrix'][:]
6     ''' accedo alle matrici dei feedback ottenute nei periodi precedenti fino ad
       un massimo di 10 cicli '''
7     for i in range(0,cycle_deep):
8
9         prev_dataset_path = scenario_directory+'/'+str(cycle-1-i)+'/dataset.hdf5'
10        prev_dataset = h5py.File(prev_dataset_path, 'a')
11
12        ''' scorro la matrice caricando in memoria una colonna alla volta '''
13        for j in range(self.scenario.N):
14
15            ''' calcolo fattore di peso per feedback positivi e negativi '''
16            pos_forgetting_factor = ((cycle_deep_max-i-1)/cycle_deep_max)*
17                TNASLsettings.pos_forgetting_factor
18            neg_forgetting_factor = ((cycle_deep_max-i-1)/cycle_deep_max)*
19                TNASLsettings.neg_forgetting_factor
20
21            ''' sommo i valori ottenuti nella matrice aggiornata dei feedback '''
22            curr_dataset['fback_matrix_updated'][:,j,0] += np.array(prev_dataset['
               fback_matrix'][:,j,0]) * pos_forgetting_factor
23            curr_dataset['fback_matrix_updated'][:,j,1] += np.array(prev_dataset['
               fback_matrix'][:,j,1]) * neg_forgetting_factor

```

5.3.5 Calcolo della reputazione

Infine il codice che segue illustra il calcolo della reputazione di A verso B. Si osserva che se A ha almeno 10 feedback verso B calcola direttamente la fiducia in B. Viceversa A cerca tra i nodi con cui ha interagito quelli di cui si fida (ovvero quelli che hanno una reputazione superiore alla soglia di onesta) e che hanno interagito con B. Le opinioni da A verso i nodi di cui si fida sono aggregate tramite l'operatore consenso definito nella *TNASL*. Allo stesso modo le opinioni dai nodi di cui A si fida verso B sono aggregate in una unica opinione. Le due opinioni risultanti sono valutate tramite l'operatore *discounting* (operatore di transitività). Alla fine l'opinione ottenuta è valutata per generare il valore di reputazione tra zero e uno risultante.

```

1  def computeTrust2(self, A, B):
2      pos = fback_from_A[A][B][TNSLA.POS]
3      neg = fback_from_A[A][B][TNSLA.NEG]
4
5      if pos > 10 or neg > 10:
6          #A has direct feedback over B, then compute directly A option over B
7          opinion_A_B = self.edit(pos, neg, self.hasPreTrust(A,B))
8          return self.eval(opinion_A_B)
9      else:
10         ..
11         for i in range(self.scenario.N):
12             pos_A_i = fback_from_A[A][i][TNSLA.POS]
13             neg_A_i = fback_from_A[A][i][TNSLA.NEG]
14             pos_i_B = fback_from_A[i][B][TNSLA.POS]
15             neg_i_B = fback_from_A[i][B][TNSLA.NEG]
16
17             if self.eval(self.edit(pos_A_i,neg_A_i, self.hasPreTrust(A,i))) >
18                 trustee_score and self.eval(self.edit(pos_i_B,neg_i_B, self.hasPreTrust(i,
19                 B)))!= 0.5:
20                 opinion_A_i = self.consensus(opinion_A_i, self.edit(pos_A_i,neg_A_i, self.
21                     hasPreTrust(A,i)))
22                 opinion_i_B = self.consensus(opinion_i_B, self.edit(pos_i_B,neg_i_B, self.
23                     hasPreTrust(i,B)))
24
25                 opinion_A_B = self.discount(opinion_A_i, opinion_i_B)
26
27             return self.eval(opinion_A_B)

```

6 Risultati

Nel precedente capitolo è stato documentato il codice del simulatore creato per valutare il modello proposto. In questo capitolo verranno discussi i risultati ottenuti dalla simulazione. Per valutare l'efficacia degli algoritmi di *trust* discussi rispetto al problema dell'identificazione dei frodatori ci interessa simulare il numero di transazioni che legano tra loro due operatori e tra queste quante transazioni sono positive e quante negative. In altre parole occorre simulare l'interconnessione di operatori telefonici per un certo numero di chiamate. La simulazione di uno scenario attendibile è la prima sfida da affrontare e verrà discussa in seguito.

L'obbietto della simulazione sarà quello di valutare l'impatto del sistema di reputazione nella individuazione dei frodatori. In particolare gli obbiettivi sono:

1. Valutare il modello creato analizzandone le proprietà "matematiche". Valutare gli errori di rivelazione. Valutare i risultati al variare del tempo di rivelazione. Valutare i risultati al variare del comportamento del frodatore.
2. Valutare i risultati del punto precedente ma considerando che solo una parte degli operatori coopera.
3. Valutare il rapporto fra probabilità di rivelazione dei frodatori e convenienza economica della frode.

6.1 Acquisizione e stima dei dati

La valutazione dei risultati di rivelazione dei frodatori tramite il sistema descritto in questa tesi avviene confrontando i valori ottenuti elaborando le tracce generate su uno scenario simulato ed i valori di riferimento su cui si è costruito lo scenario. Modelare lo scenario nel modo più verosimile possibile rispetto allo scenario della telefonia internazionale è una sfida necessaria per valutare correttamente i risultati del sistema proposto.

Numero di operatori locali. Gli operatori locali, anche chiamati LEC (Local exchange carrier), sono gli operatori che vendono che forniscono il servizio direttamente ai consumatori (vendita al dettaglio). Nello scenario di riferimento rappresentano sia l'operatore di origine che quello di transizione. Si può stimare che siano presenti circa 2000 operatori. Tra questi 800 sono gli operatori con licenza GSM ed i restanti sono i cosiddetti operatori virtuali.

I dati usati per la stima sono raccolti sotto. Si nota che è stata considerata una sottostima dovuta al fatto che alcune compagnie vengono riportate più volte a causa dei sotto-brand.² ³ ⁴

- 1025 MNOs (LG Telecom, H3G, Orange, Glo Mobile, Cambridge Telephone Company Inc.)
- 226 MCC codes, 225 countries
- 1686 Carrier + code combinations

² <http://www.quora.com/How-many-mobile-operators-are-there-worldwide>

³ <http://www.mcc-mnc.com/>

⁴ www.gsma.com/membership/who-are-our-gsma-members/full-membership/

Numero di carrier internazionali. (wholesale operators o IXC) Il numero complessivo di carrier è di 10,000 [17]. Tra questi:

- 100 carrier T1 (Telus Mobility, Bell Canada, AT&T, Verizon)
- 8800 carrier T2,T3 (compagnie private)

Ciascun carrier possiede nodi (detti Point of presence) in più paesi. Gli operatori più grandi, come Telstra ?? hanno più di 2000 nodi. Gli operatori più piccoli hanno dai 20 ai 50 nodi ??.

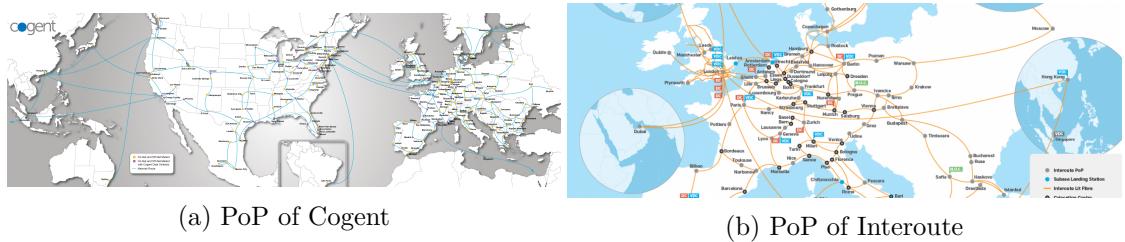


Figura 40: Carrrier PoPs

Numero di chiamate internazionali annue. Il numero di chiamate internazionali annue è misurato in milioni di minuti. La stima migliore ottenuta è di 30 miliardi di minuti annui [2] con una durata media di 6 minuti [3].

6.1.1 Providers, intermediari e minuti di chiamate

Lo scenario globale descritto sopra offre una fotografia d'insieme ma risulta difficile da modellare con precisione. Ricordo che l'obiettivo della tesi è quello di individuare i frodatori. Per poter ottenere dei risultati sensati è fondamentale modellare lo scenario in modo quanto più possibile realistico. Limitare lo scenario analizzando un sottoinsieme dello scenario globale consente di semplificare la complessità della modellazione migliorando la veridicità del modello. Infatti dei 220 paesi che sono mappati tramite i codici MCC una buona parte non contribuirà in modo significativo al traffico di chiamate soggette a frodi internazionali e potrà essere ignorata.

ITU nel 2017 ha pubblicato la lista dei 10 paesi da cui le chiamate fraudolente hanno origine e quella dei 10 paesi in cui terminano.

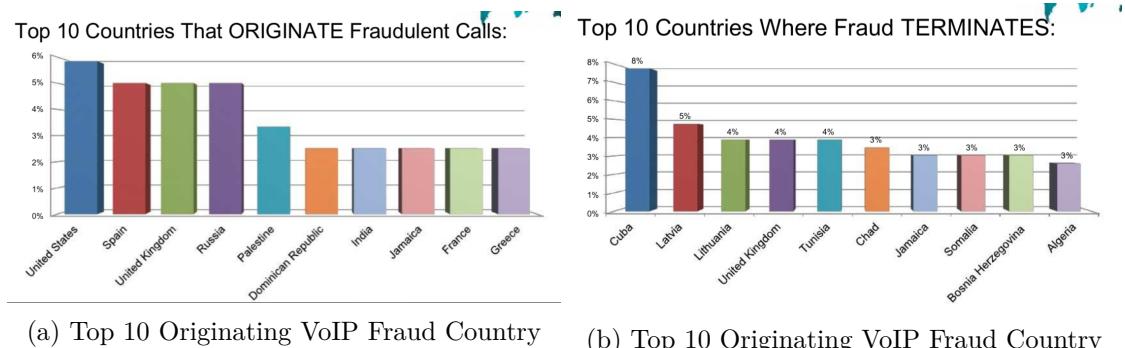


Figura 41: source: ITU 2017[4]

Dalla lista di tutti gli operatori locali è possibile conteggiare il numero di operatori mobile per ogni paese.

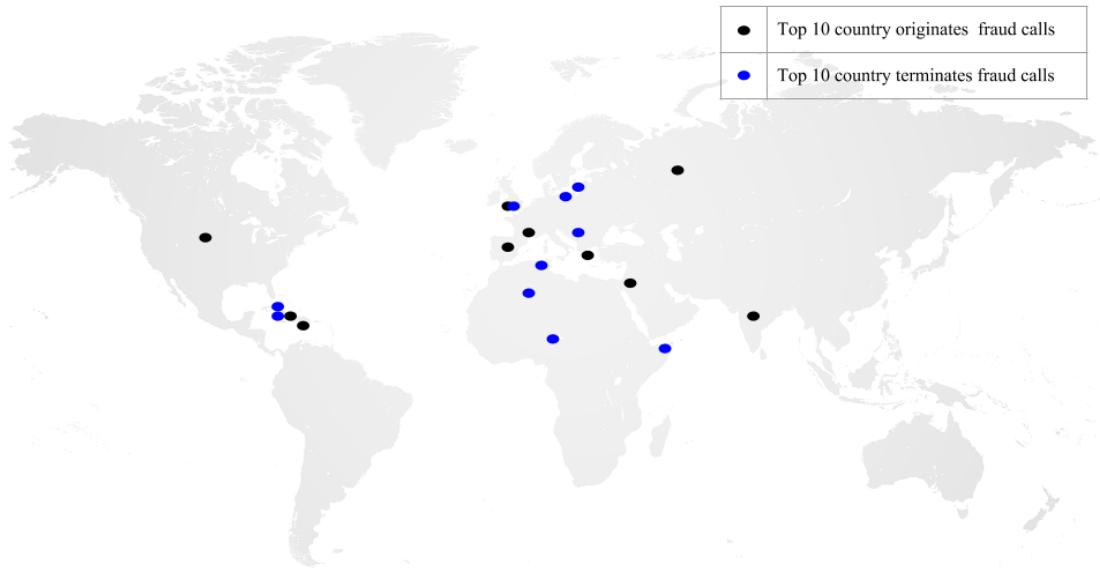


Figura 42: Caso di studio su 20 paesi

Ora consideriamo il numero di operatori di transito. Questi operatori non sono licenziati e non operano in un posto geografico. Tuttavia possiamo iniziare a stimare il numero di operatori VOIP presenti nei paesi di origine e terminazione. che offrono servizi di terminazione Conteggiando gli operatori di *VOIP Wholesale* dal sito voipproviderslists.com per i paesi indicati si ha:

| global | reduced |
|--------|---------|
| 2000 | 200 |
| 10,000 | 1000 |

Tabella 4: scenario

Numero di minuti terminati su un gateway internazionale. Infine occorre considerare il numero di chiamate che raggiungono i gateway internazionali di questi paesi. Il fatto di considerare solo il traffico che termina è giustificato dal fatto che la percentuale di frodi si misura sulla terminazione.

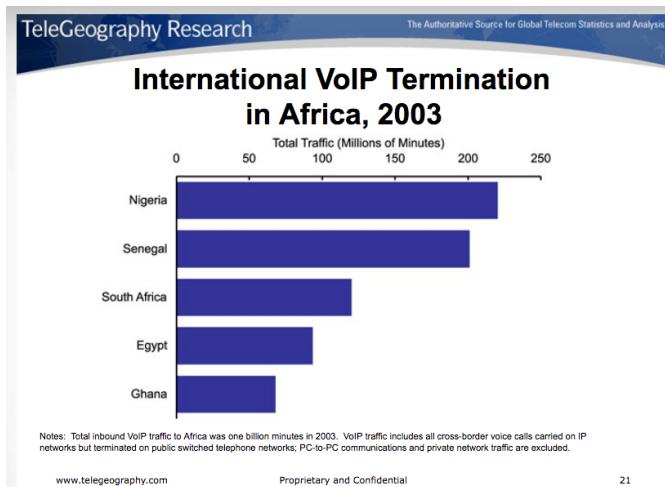


Figura 43: source: Telegeography research, 2003

6.1.2 Frodi e frodatori

Analisi dei costi dei frodatori. Un apparato SIMBox è raffigurato in figura 44. Tipicamente un apparato è composto da 60 SIM cards. Ogni SIMCard è in grado di gestire una chiamata proveniente dal traffico VOIP internazionale. La banda internet necessaria per gestire una chiamata è di 16 kbps. Quindi è necessaria una connessione internet di 960 Kbps, ovvero 1 Mbps con costo di 1000 € al mese. Il costo dell'infrastruttura di gateway e router è di 40k €, che ammortato mensilmente ha un costo di 3,333 €.

Occorre poi valutare il costo per l'uso delle SIM. Ogni SIM ha un abbonamento mensile di 15 € (che non comprende il traffico). In tutto le 60 SIM costano 900 € al mese. In tutto i costi fissi complessivi a carico del frodatore ammonta a 5223 € al mese.

Analisi dei ricavi dei frodatori. L'apparato SIMBox è in grado di gestire 60 SIM. Ciascuna SIM viene usata per 5 ore al giorno. In totale si ha una capacità massima di 18,000 minuti di chiamate gestite dal frodatore ogni giorno. Mensilmente il frodatore gestisce 540,000 minuti di chiamate. Considerando una media di 0,14 € di tariffa di terminazione, il frodatore guadagna mensilmente 75,600 € ($540,000 \times 0.14$). A questa cifra va sottratto il costo che il frodatore paga all'operatore locale del chiamato che assumiamo sia di 0.10 €. In questo caso il frodatore spende 54,000 €.

Al netto delle spese il frodatore guadagna 16,000 € in una settimana. La cifra può aumentare fino a 50,000 € se il frodatore riesce ad ottenere tariffe vantaggiose verso l'operatore locale. Si evidenzia come il limite di guadagno che il frodatore deve avere per non andare in perdita è di 0.03 €.



- 60 SIMs
- 1 Mbps internet connection
- 2 person
- 50k€ monthly

Figura 44: SIMBox device

6.2 Simulazione tramite scenari

Lo scenario della telefonia internazionale è tutt’altro che ben definito. Nel precedente capitolo si è stimato il numero di transazioni (numero di chiamate) ed il numero di nodi (numero di operatori), rispettivamente in 200 *providers* e 400 *intermediaries*. D’ora in avanti considereremo questi valori come valori fissi, non soggetti a variazioni nelle future simulazioni. Invece considereremo variabili tutti gli altri parametri, sia quelli che caratterizzano l’incertezza sullo scenario sia quelli che caratterizzano la configurazione del sistema di rivelazione.

E’ evidente che i risultati ottenuti considerando lo stesso scenario variano al variare dei parametri di configurazione del sistema di rivelazione. Viceversa una volta definiti i parametri del sistema di rivelazione occorrerà studiare come variano i risultati al variare dello scenario. Le variabili che incidono maggiormente sui risultati sono in ordine di importanza (1) la percentuale di *providers* e *intermediaries* cooperanti, (2) le strategie di camuffamento adottate dai frodatori, (3) i parametri usati nell’algoritmo di *TNASL*, (4) l’adozione o meno delle tecniche di sconto delle accuse descritte in precedenza (5) il numero di chiamate analizzate nel tempo che intercorre fra una computazione e quella successiva.

La variazione di questi parametri influenza la capacità di identificazione dei frodatori. Per studiare singolarmente i contributi introdotti da tutti questi fattori ho scelto di organizzare la trattazione in scenari. L’ordine con cui vengono trattati i fattori variabili sopra elencati è stabilito in base al fatto che i risultati ottenuti per via sperimentale in uno scenario possono essere utilizzati per fissare alcuni parametri nello scenario successivo. Gli scenari individuati sono:

1. ***Scenario ideale***. Nello *Scenario ideale* viene studiata la reputazione di alcuni nodi nella fase transitoria e a regime. In questo scenario si considera la partecipazione dei nodi al 100% e si fissa il comportamento fraudolento a circa 50%. I parametri soggetti a variazione sono i parametri di configurazione dell’algoritmo di trust (3), il tempo di ritardo tra computazioni successive (5) e l’adozione delle tecniche di sconto delle accuse (4). L’intento di questo scenario è quello di trovare i parametri che caratterizzano il sistema di rivelazione dei frodatori nel caso ottimo.

2. **Scenario variabile.** Nello *Scenario variabile* viene studiato (5) l'errore di rivelazione al variare del comportamento dei frodatori. In questo scenario considera la partecipazione di tutti i nodi e si fissa i parametri di configurazione del simulazione in base ai risultati precedentemente ottenuti. Variando la percentuale di frodi e frodatori nel campione si varia il comportamento fraudolento. L'intento di questa simulazione è quello di capire a che punto è possibile individuare un frodatore che commette poche transazioni negative frodi rispetto a quelle positive ed è quindi più difficile da individuare.
3. **Scenario applicativo.** Nello *Scenario applicativo* viene preso in considerazione il caso in cui solo una parte dei *providers* e degli *intermediaries* partecipi al sistema e le conseguenze che comporta.
4. **Scenario attuativo.** Nello *Scenario attuativo* viene preso in considerazione il caso in cui ogni volta che un nodo è identificato come frodatore viene inserito in una blacklist temporanea.

6.2.1 Scenario ideale

L'algoritmo di trust *TNASL* è stato descritto in precedenza. In questo capitolo verranno studiati i risultati in uscita all'algoritmo di trust al variare dei parametri di configurazione. Ricordo che le simulazioni effettuate in *Scenario ideale* vengono fatte con frodatori al 1% e frodi al 5% che risulta in un comportamento fraudolento intorno al 50%, ovvero ciascun frodatore effettua all'incirca tante chiamate positive quante negative. Ricordo infatti che la reputazione calcolata nel caso i feedback positivi siano uguali a quelli negativi è di 0.5, ovvero la massima incertezza. Naturalmente il numero di transazioni positive e negative effettuate da un nodo fraudolento oscilla a causa della randomicità con cui vengono generate le tracce. Inoltre nello *Scenario ideale* si considera il 100% dei partecipanti, 200 *providers* e 400 *intermediaries*. Ogni chiamata è gestita da 2 providers, uno di origine e uno di terminazione e da 4 intermediari.

Ciascuna simulazione ripete la generazione delle tracce telefoniche ed il calcolo della reputazione per un certo numero di volte simulando l'evoluzione temporale del sistema. L'algoritmo di trust genera in uscita un valore che contiene la reputazione espressa da un nodo *source* nei confronti di un nodo *target* a ciascun ciclo. Inoltre il sistema può essere configurato per considerare al ciclo *i*-esimo i valori dei feedback pesati dei precedenti *n* cicli. Al termine la simulazione mostra l'evoluzione nel tempo della reputazione di quattro nodi, due fraudolenti e due onesti.

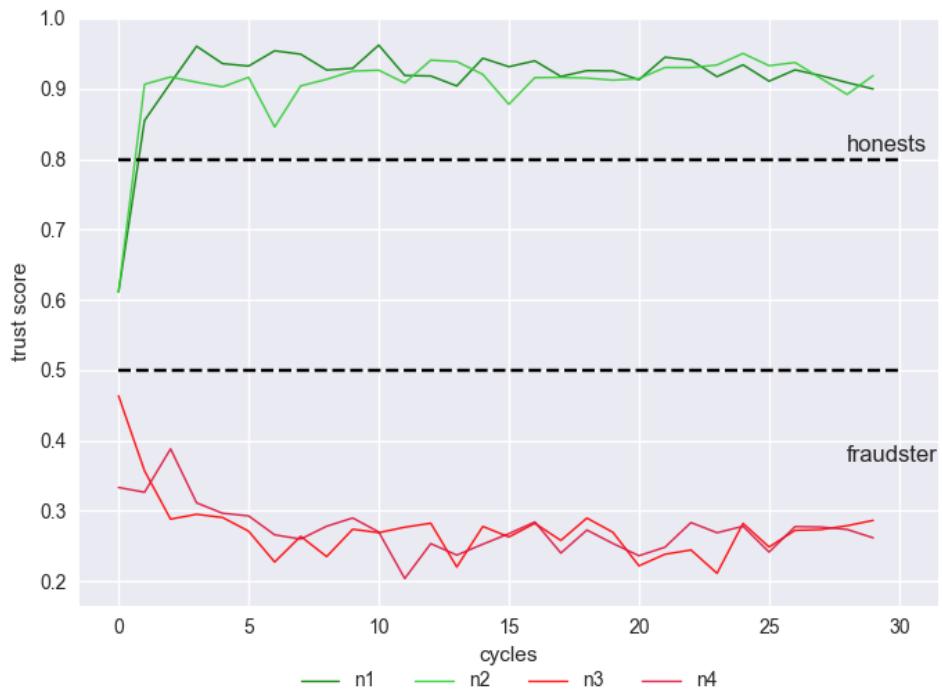


Figura 45: 50k chiamate per ciclo

Il diagramma nella figura sopra mostra che la reputazione dei due nodi onesti è superiore a 0.8 mentre la reputazione dei due nodi frodatori è inferiore a 0.5. Si nota inoltre che la reputazione dei nodi si assesta subito dopo un breve periodo transitorio ai valori indicati nel diagramma. I risultati mostrati in figura sono piuttosto positivi ma se si variano i parametri in ingresso al simulatore questi risultati possono cambiare in modo radicale. In particolare si consideri che:

- Al variare del numero di chiamate analizzate su un ciclo varia il numero di feedback positivi e negativi cumulati. Se *source* non ha transazioni dirette verso *target*, *source* deve individuare i nodi che per lui hanno una reputazione superiore a 0.8, chiamati *trustee*. Tuttavia se le chiamate analizzate sono poche anche i feedback positivi visti da *source* verso *trustee* sono pochi. Ad esempio nel caso analizzato sopra, solo un nodo su 400 ha accumulato 4 transazioni positive e zero negative ed è stato utilizzato per il calcolo della reputazione indiretta avendo ottenuto una reputazione di 0.833. Nelle transazioni successive alla prima il problema è meno accentuato perché i feedback correnti vengono sommati a quelli passati. Nel caso illustrato i feedback da *source* verso *trustee* a regime variano da 3 a 10 transazioni, che sono moltissime.
- Maggiore è numero di cicli precedenti che vengono utilizzati nel calcolo della reputazione ad un certo istante nel tempo maggiore è la robustezza del sistema a comportamenti di *disguised malicious*. D'altra parte però andrebbe simulato il tempo che viene impiegato a individuare il cambiamento da onesto a fraudolento di un nodo.

- A seconda dei valori numerici considerati, il fatto di valutare di più i feedback negativi rispetto quelli positivi cumulati nei cicli precedenti facilità la classificazione dei frodatori ma rischia di penalizzare i nodi onesti.

Per misura gli effetti delle considerazioni sopra esposte tramite sperimentazione numerica si valuterà:

- 1. numero di chiamate per ciclo:** Analizzare 10k, 20k o 50k chiamate per ogni ciclo.
- 2. numero (memoria) di cicli precedenti:** Analizzare lo storico dei feedback delle ultime 4, 10 o 20 passate elaborazioni (valutare smorzamento nel cambio di reputazione o velocità con cui cambia la reputazione)
- 3. coefficiente di peso dei feedback precedenti:** Utilizzare *forgetting_factor* diversi per feedback positivi e negativi

6.2.1.1 Numero di chiamate per ciclo La durata temporale di ciascun ciclo dipende dal numero di chiamate generate dal simulatore. Infatti noto il tasso medio di chiamate globali si può facilmente ricavare il ritardo temporale associato all’analisi di una traccia e da qui stimare il ritardo di rivelazione del frodatore. Questo aspetto verrà trattato nelle conclusioni. In questa sezione il ritardo temporale verrà espresso in numero di chiamate per ciclo. Nelle simulazioni effettuate si considera i casi di 10k, 30k e 100k chiamate per ogni ciclo. Il numero delle chiamate, ovvero la finestra temporale tra una elaborazione e l’elaborazione successiva, deve essere scelto in modo adeguato. Tanto maggiore è la finestra tanto più accurato è il calcolo della reputazione perché si hanno più feedback. Tuttavia tanto più corta è la finestra quanto prima si può individuare il frodatore. Nel caso in cui si considerano solo 10k chiamate si vede dalla figura che il sistema non è in grado di accumulare sufficienti feedback a valutare la reputazione. Aumentando il numero di chiamate a 20k il calcolo della reputazione dei frodatori migliora ma quella dei nodi onesti è inferiore alla soglia. Aumentando il numero di chiamate a 50k il calcolo della reputazione dei frodatori migliora decisamente. Possiamo quindi temporaneamente definire 50k chiamate per ciclo come valore di riferimento per le simulazioni successive.

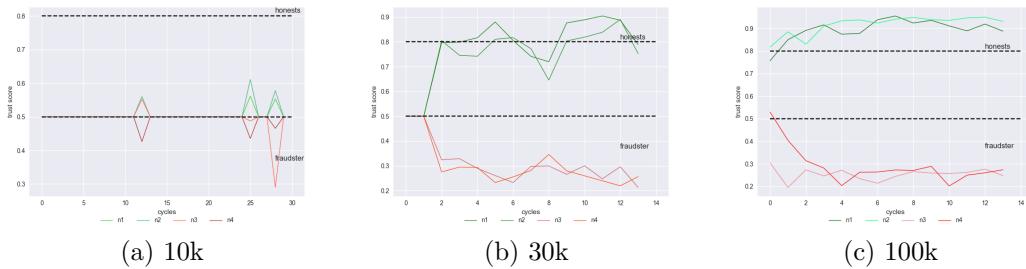


Figura 46: Variazione del numero di chiamate analizzate per ciclo

6.2.1.2 Numero (memoria) dei cicli precedenti Dalla precedente simulazione abbiamo capito che il numero di chiamate per ciclo non può scendere sotto le 50k chiamate. Con questa informazione si va a studiare come varia l’andamento della reputazione all’aumentare del numero di cicli precedenti considerati. Ricordo che la memoria del

sistema di reputazione viene considerata a livello dei feedback. I feedback relativi all'ultimo periodo sono sommati ai valori dei feedback precedentemente ottenuti, pesando maggiormente i feedback recenti rispetto a quelli meno recenti.

L'effetto di "somma" sui feedback passati contribuisce ad aumentare il numero dei feedback effettivamente analizzati per il calcolo della reputazione. In altre parole il numero minimo di chiamate per ciclo dipende anche dal numero di cicli precedenti considerati. Riprendendo l'esempio in figura 46c e ripetendo la simulazione del sistema con gli stessi valori ma memoria azzerata (numero di cicli precedenti uguale a zero) si ottiene l'andamento in figura 47.

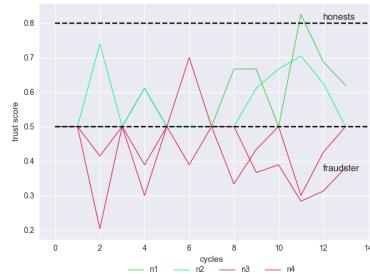


Figura 47: 50k, without memory

Per valutare l'effetto della memoria sul sistema di reputazione senza essere influenzati dal numero di chiamate per ciclo effettueremo due simulazioni con 100k chiamate a ciclo, una con e una senza memoria. In figura 48a si è simulato il caso di assenza di memoria (0 cicli con 100k chiamate per ciclo). In figura 48b si è simulato il caso di sistema con memoria (4 cicli con 100k chiamate per ciclo). Si osserva che il grafico con 4 cicli di memoria ha un andamento più "smussato" rispetto allo stesso grafico con zero memoria. Si osserva inoltre che l'introduzione di memoria nel sistema ha un diverso effetto sull'andamento della reputazione positiva e negativa. Le due reputazioni positive hanno una minore escursione nel tempo rispetto a quelle negative. Il fattore di memoria gioca un ruolo decisivo nel contrastare i casi di falsi positivi perché attenua eventuali "sbandate" nella reputazione di un nodo onesto grazie allo storico di valori positivi che nel tempo un nodo onesto ha accumulato.

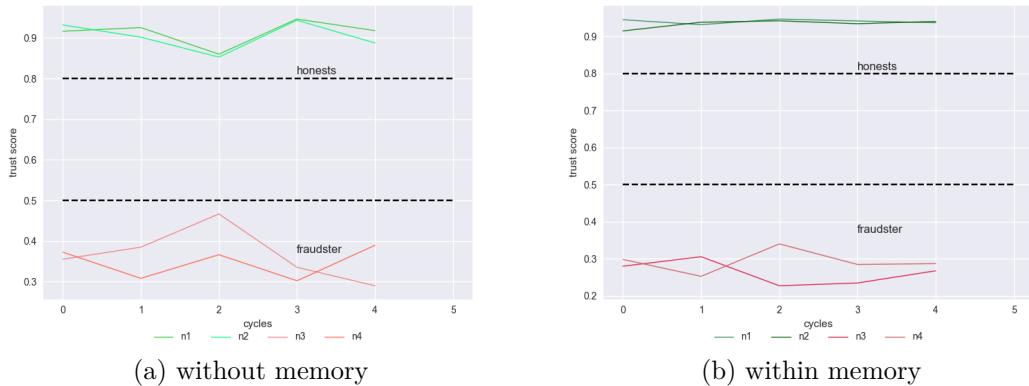


Figura 48: Variazione del numero di cicli precedenti (memoria del sistema)

6.2.2 Scenario variabile

Lo studio degli errori di rivelazione consiste nell'analizzare la percentuale di (1) falsi negativi, (2) falsi positivi e (3) rivelazioni mancate al variare delle circostanze. Dalle statistiche note in letteratura sulle frodi è possibile ricavare la percentuale di chiamate frodate sul totale. Si può assumere che le frodi di bypass si attestano al 5% a livello globale. Tuttavia questa percentuale varia di anno in anno e varia a seconda delle rotte analizzate fino ad arrivare al 17% del traffico sulle rotte più profittevoli. Per quanto riguarda il numero di frodatori non sono state trovate statistiche al riguardo. Infine non sappiamo se gli operatori fraudolenti instradano tutto il traffico in modo fraudolento (comportamento purely malicious) oppure se instradano solo una percentuale di traffico fraudolento e in questo quanto è la percentuale del traffico fraudolento rispetto al traffico onesto (disguised malicious).

Tutto questo porta ad una situazione di incertezza che occorre analizzare variando i parametri frodi e frodatori. In particolare ci domandiamo:

- Se il frodatore diminuisce il numero delle chiamate fraudolente (quindi aumenta il numero di chiamate gestite onestamente) il sistema di rivelazione riesce a classificare correttamente il nodo frodatore?
- Se il numero di chiamate fraudolente aumentano drasticamente rispetto al totale delle chiamate analizzate provocando un aumento delle accuse rivolte ai nodi onesti il sistema riesce a classificare correttamente i nodi onesti?

Per rispondere a queste domande si procederà, come fatto in precedenza, tramite sperimentazione numerica. In particolare si definiscono tre casistiche di studio "ragnanevoli":

1. **caso *neutral*.** Il caso *neutral* è quello che è stato utilizzato per valutare tutte le simulazioni viste in precedenza e rappresenta il caso neutro per cui il comportamento fraudolento risulta essere del 57%. Questo caso è ottenuto quando si ha il 5% delle frodi e l'1% dei frodatori.
2. **caso *disguised*.** Il caso *disguised* simula la situazione in cui ciascun frodatore per non essere trovato diminuisce il numero di chiamate fraudolente rispetto alle chiamate oneste. In questo caso il comportamento fraudolento valutato sarà del 20%. Questo comportamento è ottenuto considerando il 5% delle frodi e il 5% dei frodatori.
3. **caso *malicious*.** Il caso *malicious* simula il caso di un aumento repentino dei casi di frode e porta il comportamento fraudolento ad un aumento del 83%. Questo comportamento è ottenuto mantenendo i frodatori al 1% ma valutando un aumento delle frodi pari al 17%.

Il fraudster behaviour esprime quante transazioni positive e quante transazioni negative sono state attribuite ad un nodo da tutti gli altri nodi. Il parametro fraud behaviour è ottenuto analizzando il numero di transazioni oneste a cui i frodatori hanno partecipato con il numero di transazioni fraudolente rivelate. Poiché le tracce sono generate in modo randomico non impostare con precisione quante transazioni positive e quante negative deve effettuare ciascun frodatore ma possiamo farlo in modo indiretto. Infatti se a parità di frodi aumentano i frodatori si avrà che ciascun frodatore sarà associato

ad un numero minore di frodi. Poiché l'inclusione di un frodatore in una traccia onesta non dipende dai parametri legati alle frodi, si avrà che il comportamento fraudolento in questo caso diminuirà.

I valori dei comportamenti fraudolenti sono riassunti in tabella.

| | caso <i>neutral</i> | caso <i>disguised</i> | caso <i>malicious</i> |
|------------------|---------------------|-----------------------|-----------------------|
| fraudsters | 1% | 5% | 17% |
| frauds | 5% | 5% | 1% |
| fraud. behaviour | 57% | 20.7% | 83% |

Tabella 5: study false negative with 5% frauds

Lo studio dei risultati in questo scenario e nel successivo sarà effettuato calcolando la reputazione da tutti i nodi *source* che partecipano al sistema, N_s , verso tutti i nodi *target* N_t . Quindi per ogni nodo *source* vengono calcolate N_t reputazioni. Per evitare rallentamenti nella simulazione i nodi *source* sono scelti dall'insieme degli intermediari con passo 10, ottenendo un sottoinsieme più piccolo (20 nodi al posto di 200). In totale si calcolano 8,000 reputazioni in ogni ciclo. Ognuna di queste reputazioni viene inviata al classificatore che a seconda del livello di reputazione e del nodo *target* corrispondente (onesto o frodatore) seleziona la casistica opportuna, come mostrato in tabella. Ogni volta che viene selezionata una casistica viene incrementato il relativo contatore. In questo modo al termine di un ciclo si ha una idea di quanti casi di falsi positivi, falsi negativi, sospetti e mancate rivelazioni si sono verificati.

| trust score | if fraudster | if honest |
|-------------|----------------|----------------|
| [0,0.5) | detected | false positive |
| 0.5 | missed | missed |
| (0.5-0.8] | suspected | suspected |
| (0.8, 1.0] | false negative | detected |

Tabella 6: Classificatore

Lo studio che segue nei prossimi paragrafi si focalizzerà su:

- La riduzione dei casi di falsi negativi e di reputazione sospetta al variare del fattore di peso dei feedback passati descritto per il caso *disguised*.
- La riduzione dei casi di falsi positivi con l'adozione delle tecniche di sconto delle accuse proposte per il caso di *malicious*.

6.2.2.1 Riduzione dei casi sospetti e falsi negativi nel caso *disguised* Il fattore di peso dei feedback passati è il valore per cui si moltiplicano i valori di feedback ottenuti ai cicli precedenti. I feedback più recenti contano di più rispetto a quelli meno recenti. Questo fattore modella la rapidità con cui il sistema di reputazione si accorge del cambiamento nel comportamento di un nodo. In questa tesi ho scelto di utilizzare un fattore peso lineare che attribuisce un valore unitario all'ultimo feedback ricevuto e scala i valori passati in modo proporzionale rispetto al numero massimo di cicli che si considerano come memoria del sistema. Il valore ottenuto tramite questo modello viene

moltiplicato per un fattore moltiplicativo che differenzia il comportamento dei feedback positivi da quelli negativi. Nel contesto umano una azione negativa nel passato influenza maggiormente la valutazione di fiducia rispetto ad una azione negativa; si dice che per dimenticare un azione negativa non basta una sola azione positiva. Nel simulatore questa differenza è evidenziata tramite due parametri, il *negative forgetting factor* che per brevità chiameremo NFF ed il *positive forgetting factor*, che per brevità chiameremo PFF , con $NFF, PFF \in (0, 1.0]$. Se si considera il caso in cui per dimenticare una azione negativa ne servono due positive si avrà $NFF = 1, PFF = 0.5$, mentre se si enfatizza questa differenza fino a dieci volte di più si avrà $NFF = 1, PFF = 0.1$.

I risultati sulle statistiche di rivelazione di onesti e frodatori effettuate con $NFF = 1, PFF = 0.5$ sono usati in precedenza. Nel caso *neutral* l'uso di questi parametri permette una perfetta classificazione di onesti e frodatori. Ricordo che per essere onesti è necessaria una reputazione superiore a 0.8 e per essere frodatori una reputazione inferiore a 0.5. Invece come si può osservare nel caso *disguised* indicato in precedenza una variazione nel comportamento fraudolento da 57% a circa 21% compromette la capacità di corretta classificazione dei nodi frodatori. Maggiore è il numero di chiamate gestite in modo fraudolento rispetto a quelle gestite in modo onesto, più difficile diviene identificare i frodatori. Per capire meglio questo fenomeno occorre analizzare la reputazione dei frodatori nel tempo, dalla fase transitoria fino alla fase di regime.

L'andamento della reputazione media di nodi frodatori (barre rosse) e nodi onesti (barre verdi) nel caso *disguised* con $NFF = 1, PFF = 0.5$ è illustrato in 49a. L'effetto di riduzione dei feedback negativi è dovuto al rapporto 2:1 con cui si pesano diversamente i feedback precedenti negativi da quelli positivi. Questo provoca a regime un effetto di amplificazione dei feedback negativi rispetto a quelli considerati all'inizio del transitorio. Tuttavia quello che si osserva è che, nonostante a regime il livello di fiducia di onesti e frodatori siano ben distinguibili (onesti circa 0.9, fraudolenti circa 0.6), il livello di reputazione dei frodatori non scende sotto la soglia di 0.5 e per questo vengono classificati come sospetti e non come frodatori.

Verifichiamo se una diversa configurazione dei parametri, in particolare $NFF = 1.0, PFF = 0.1$, possa portare ad una maggiore robustezza del sistema rispetto agli attacchi di *disguised malicious* (diminuzione del comportamento fraudolento fino al limite utile per non essere rivelati). In figura 49b sono mostrati i risultati. Come si vede il livello medio di reputazione dei frodatori scende sotto la soglia di 0.5, migliorando quindi le capacità di rivelazione del sistema. Si osserva che la reputazione media con cui partono i frodatori è di 0.78 (simile al caso precedente) e che questa scende fino al valore di 0.49, con una differenza tra la reputazione iniziale e finale di circa 0.3. Invece la reputazione dei nodi onesti scende da 0.95 a 0.86, ovvero con una differenza poco inferiore a 0.1. Da questi valori si deduce che una differenza di 10:1 nella valutazione dei feedback negativi rispetto a quelli positivi consente di individuare i frodatori nel caso il comportamento fraudolento scenda fino al 20%. Sotto a questo valore non è più possibile classificare correttamente un nodo fraudolento come fraudolento.

Si sottolinea che in questa situazione il livello di reputazione dei nodi onesti può scendere avvicinandosi alla soglia di 0.8 e compromettere la capacità di individuare il comportamento onesto di un nodo onesto. Nel prossimo paragrafo vedremo come si può fronteggiato il problema dei falsi positivi tramite alcuni accorgimenti volti a migliorare

la reputazione dei nodi onesti.

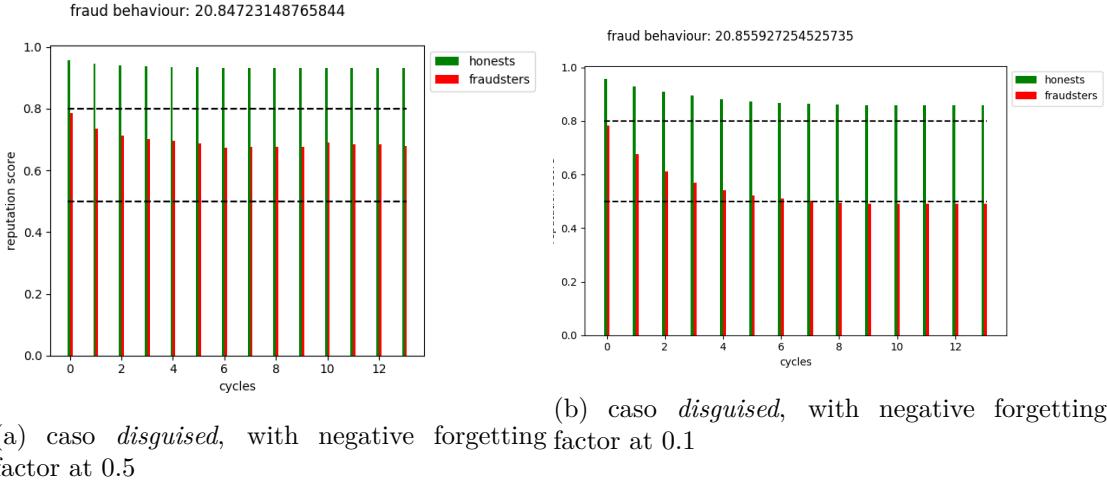


Figura 49: Variazioni nel comportamento fraudolento

6.2.2.2 Riduzione dei falsi positivi nel caso malicious Tutte le simulazioni fatte in precedenza sono state fatte considerando parimenti colpevoli tutti i nodi che hanno partecipato ad una chiamata fraudolenta. L'effetto di ciò è l'attribuzione di una transazione negativa a tutti i nodi onesti che precedono quello fraudolento nella chiamata con frode. E' evidente quindi che all'aumentare delle frodi aumenteranno anche le transazioni negative associate ai nodi onesti. Come verrà sottolineato più volte nel corso della tesi i casi di falsi positivi incideranno in modo radicale sulla valutazione del risultato finale. Come si vede in figura l'aumento delle frodi porta la reputazione dei nodi fraudolenti vicina allo zero e quella dei nodi onesti poco sotto la soglia di 0.5.

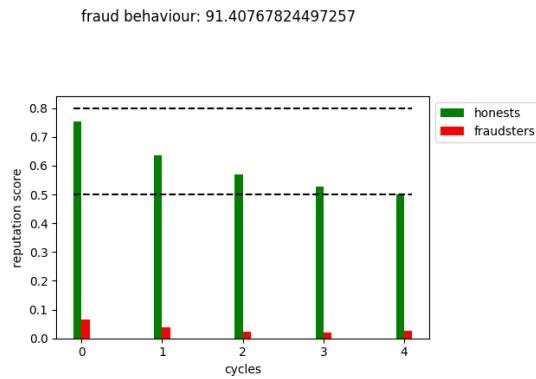


Figura 50: 1% fraudsters, 30% frauds

Nel caso *neutral*, dove ricordiamo che il comportamento fraudolento è poco più del 50% non si hanno casi di falsi positivi. Variando la percentuale di frodi da 1% a 17% il comportamento fraudolento sale quasi fino a 84%. In questo caso, che precedentemente abbiamo chiamato caso *malicious*, la corretta classificazione dei nodi onesti viene quasi del tutto compromessa, portando a classificare parte dei nodi onesti come frodatori.

Classificare un nodo onesto come sospetto non ha nessun effetto sul risultato finale. Viceversa è molto importante verificare che la reputazione di un nodo onesto non scenda sotto la soglia di 0.5. In questo caso il nodo onesto sarebbe classificato come frodatore. Per verificare qual'è la soglia di percentuale di chiamate fraudolente oltre la quale il sistema di reputazione classifica i nodi onesti come frodatori si sono effettuate ulteriori simulazioni impostando il comportamento fraudolento al 100% ed aumentando il numero di frodi al 17%. Questi valori simulano il caso in cui su una specifica rotta le tariffe vantaggiose offerte da pochi frodatori (1% ovvero 4 su 400) possono portare ad avere un parte significativa del traffico su quella rotta soggetto a frode (17% del traffico totale).

Lo studio sulla riduzione dei casi di falsi positivi sarà eseguito tramite:

1. l'uso della fiducia a priori (pre-trust)
2. lo sconto delle accuse reciproche.

Sconto delle accuse reciproche In particolare si è considerato il fatto che i nodi frodatori hanno nei confronti degli altri operatori un traffico unidirezionale (isolato in ingresso) mentre gli operatori onesti hanno un traffico bidirezionale (sia in ingresso che in uscita) vero gli operatori con cui interagiscono. L'attribuzione dei feedback avviene nella direzione della transazione quindi rispecchia questa tipologia di interazione. In altre parole i casi di accuse reciproche verranno registrati molto più dai nodi onesti che non da quelli fraudolenti. In virtù di questa considerazione un azzeramento delle accuse reciproche tra due nodi qualsiasi avrà un effetto benefico sui nodi onesti lasciando le accuse ai nodi disonesti pressoché inalterate.

L'immagine 51b mostra la medesima situazione considerando lo sconto delle accuse per simmetria. Sebbene la differenza tra i due casi non sia molto accentuata, si vede chiaramente che l'uso della tecnica di simmetria porta a regime ad una reputazione media dei nodi onesti superiore alla soglia di onestà.

La quantificazione della riduzione dei casi di falsi positivi con e senza lo sconto delle accuse reciproche è mostrata in figura 51b. Si osserva che dal 2% i casi di falsi positivi diminuiscono al 1.4%. Probabilmente questa piccola differenza nel caso con e senza l'uso della simmetria è giustificata dalla natura randomica con cui si sono costruite le tracce. Infatti la scelta randomica dell'operatore successivo non modella il caso di traffico bilanciato da e verso due operatori.

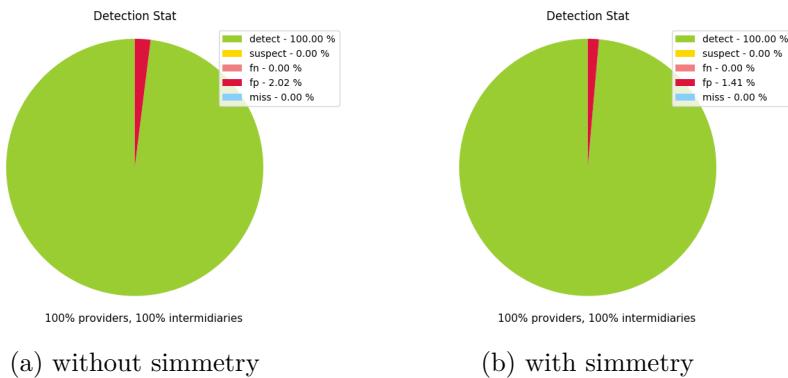


Figura 51: False positive reduction with simmetry

6.2.2.3 Pre-trust La teoria *TNASL* prevede l'utilizzo della fiducia a priori. In una situazione reale un operatore onesto stipula accordi commerciali con altri operatori. Possiamo considerare questi operatori come operatori di fiducia. Tipicamente un operatore ha dai 2 ai 4 operatori per rotta. Questa situazione è simulata dal codice del simulatore che per ogni nodi *source* definisce 4 nodi *target* di fiducia tra gli *intermediaries* e 2 tra i *providers*.

L'effetto della reputazione a priori non porta come ci si aspetterebbe ad un risultato migliore sulla riduzione dei falsi positivi. Infatti se da un lato la fiducia a priori verso alcuni operatori diminuisce dall'altro la valutazione di questi nodi di fiducia nei confronti dei falsi positivi né peggiora ulteriormente la reputazione. In virtù di questa considerazione, per non alterare i risultati ottenuti, ho scelto di non utilizzare la fiducia a priori.

6.2.3 Scenario applicativo

Lo *Scenario applicativo* è lo scenario più importante da studiare. In questo scenario si considererà gli effetti dovuti alla partecipazione parziale di *providers* e *intermediaries*. Una partecipazione parziale dei *providers* limita la capacità di valutare se una traccia è soggetta a frode oppure no. Infatti nel caso in cui un *provider* non partecipi al sistema di reputazione, tutte le tracce in cui quell'operatore è l'operatore di terminazione vengono scartate. Una partecipazione parziale degli *intermediaries* aumenta le probabilità di falsi negativi, falsi positivi e casi di mancata rivelazione. Infatti se un intermediario non partecipa al sistema, la ricostruzione della traccia sarà incompleta mancando il giudizio che l'intermediario assente avrebbe espresso verso l'operatore che lo succedeva nella chiamata. La ridotta partecipazione altera lo stato delle transazioni, sia positive che negative, influenzando il calcolo della reputazione e quindi la corretta classificazione dei nodi.

La simulazione viene effettuata per via sperimentale, fissando quattro casi di indagine:

- **full cooperation.** Cooperazione dei providers al 100% e degli intermediari al 100%.
- **halved cooperation.** Cooperazione dei providers al 50% e degli intermediari al 25%.
- **decimated cooperation.** Cooperazione dei providers al 20% e degli intermediari al 10%.
- **realistic cooperation.** Cooperazione dei providers al 10% e degli intermediari al 5%.

Questi quattro casi verranno studiati nei tre casi descritti in precedenza, il caso *neutral*, *disguised* e *malicious*. I risultati sono mostrati in figura.

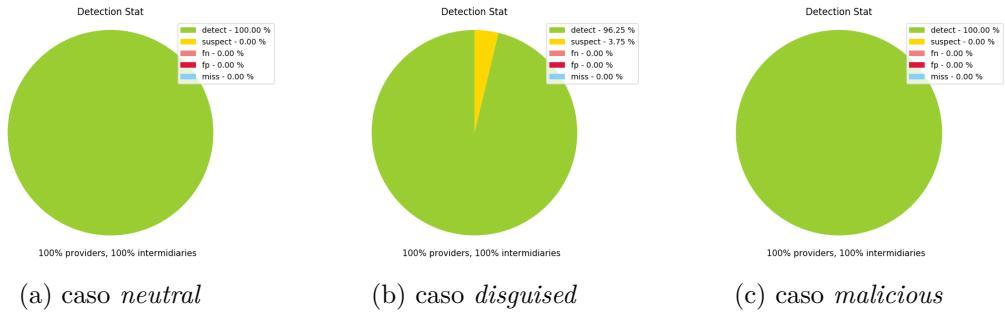


Figura 52: Full cooperation

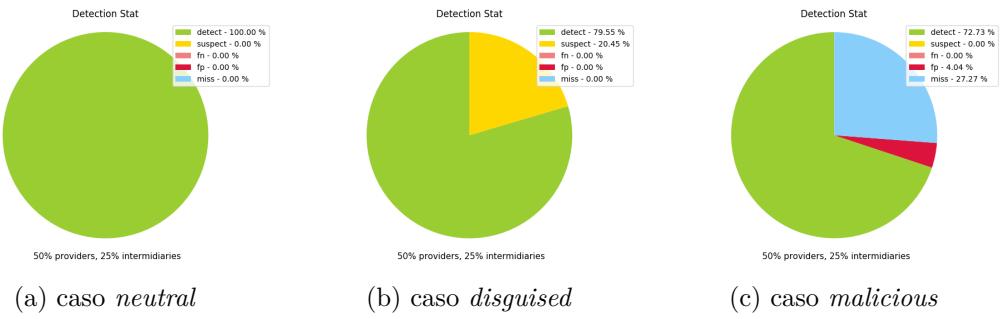


Figura 53: Halved cooperation

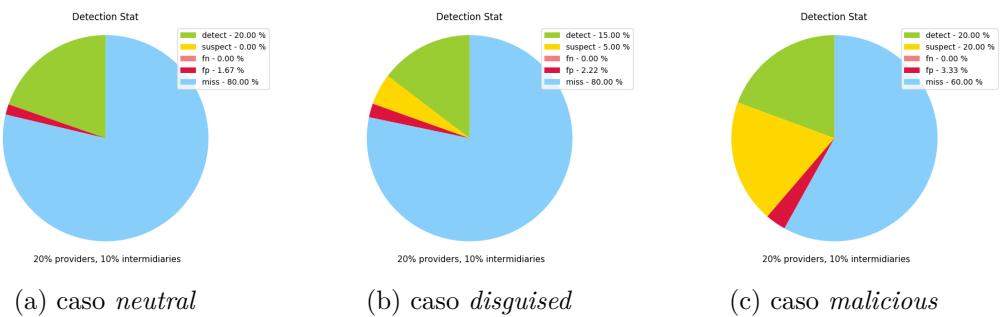


Figura 54: Decimated cooperation

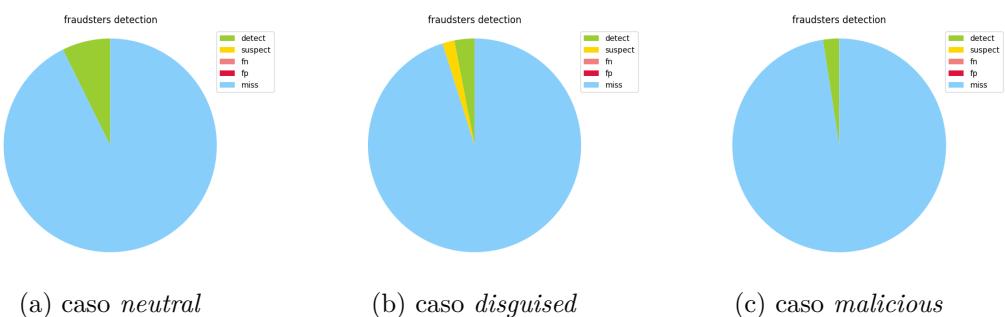


Figura 55: realistic cooperation

I risultati ottenuti verranno commentati come segue:

- Nel caso *full* il sistema di reputazione è in grado di classificare correttamente frodatori ed onesti. Solo nel caso *disguised* una piccola percentuale di frodatori viene classificata come sospetta.
- Nel caso *halved* si osserva che per un comportamento *malicious* il numero di falsi positivi aumenta al 4.4%, ben oltre un livello di tolleranza accettabile.
- Sia nel caso *decimated* che nel caso *realistic* i casi di mancata rivelazione aumentano a causa delle minori tracce analizzate e delle mancate accuse dei nodi assenti. Tuttavia una piccola percentuale di frodatori è identificata.

6.2.4 Scenario attuativo

Dai risultati ottenuti nello scenario precedente emerge chiaramente il problema dei falsi positivi nei casi *malicious*. Idealmente il sistema di reputazione dovrebbe essere in grado di classificare correttamente frodatori ed onesti in ogni situazione. In pratica la situazione di un aumento incontrollato dei casi di frode non sarebbe veritiera perché dopo aver individuato un frodatore, come sarebbe logico aspettarsi, quest'ultimo andrebbe eliminato dal sistema.

Nello scenario attuativo verrà studiato come variano i risultati se si considera la rimozione temporanea dei nodi che ad un certo punto vengono classificati come frodatori.

Per simulare questa dinamica verrà implementata una blacklist temporanea in cui vengono aggiunti al termine del ciclo tutti i nodi che hanno ottenuto una reputazione inferiore a 0.5. In questo modo la generazione della traccia successiva verrà eseguita escludendo i nodi inseriti in balcklist. Ad ogni ciclo dispari la blacklist si svuota ed il procedimento si ripete. In questo modo si simula l'adozione del sistema di reputazione nelle tabelle di routing: se un nodo non è fidato, si evita di mandargli il traffico.

L'adozione delle blacklist temporanee ha un effetto sulla riduzione dei casi di falsi positivi. Infatti nei cicli in cui gli operatori fraudolenti vengono esclusi dalle tracce non si hanno accuse ingiuste rivolte agli onesti. Questo provoca un aumento dei feedback positivi rivolti ai nodi onesti senza avere aumenti dei feedback negativi. D'altra parte i nodi fraudolenti, anche se bloccati per un ciclo, non perdono la reputazione negativa ottenuta al ciclo precedente. Nel complesso questo meccanismo fa sì che più avanza il tempo e più si riduce il numero di falsi positivi.

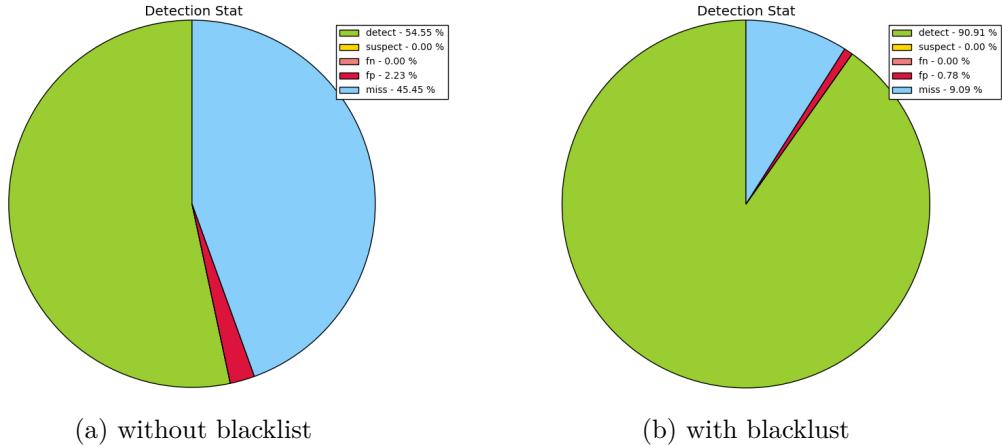


Figura 56: Blacklist effect on false positive reduction

La simulazione mostrata è stata eseguita nel caso *halved cooperation, malicious* su 14 cicli con 240,000 chiamate a ciclo ed il 100% di comportamento fraudolento. (il caso peggiore in cui considerare i falsi positivi). Si precisa che l'immagine mostra lo stato al 14-esimo ciclo e che analizzando i dati grezzi si osserva una differenza tra i falsi positivi di uno e dell'altro generalmente superiore al 1.5%.

Riassunto. La variabilità sul comportamento dei frodatori, del numero di frodi in nella traccia analizzata e del numero di frodatori nello scenario considerato influenzano la corretta classificazione di nodi frodatori e nodi onesti. Per questo motivo sono state studiate tre situazioni possibili, indicate come A,B e C. I parametri inizialmente usati per la simulazione permettevano di soddisfare il requisito sulla corretta classificazione di onesti e frodatori solo nel caso *neutral*. Sperimentando configurazioni diverse da quelle iniziale si è giunti alla conclusione che con i parametri illustrati in tabella si riesce a soddisfare il requisito sulla corretta classificazione in tutti e tre i casi. In particolare il sistema di reputazione individuato ha:

1. Capacità di classificare correttamente i frodatori con comportamento fraudolento uguale o superiore al 20%.
2. Capacità di classificare correttamente gli onesti fino al 17% di frodi nel campione.

I valori dei comportamenti fraudolenti sono riassunti in tabella.

| param | value |
|---------------------------|----------|
| number of calls per cycle | > 240k |
| number of prev. cycles | 10 |
| trustee score | 0.8 |
| pos. forgetting factor | 1.0 |
| neg. forgetting factor | 0.1 |
| acc. discount simmetry | True |
| use blacklist | True |
| fraudsters camouflage | True |
| fraudsters | 1% - 5% |
| frauds | 1% - 17% |

Tabella 7: final params

6.3 Analisi costi benefici

Nel capitolo precedente sono stati studiati i parametri di configurazione del sistema di reputazione in funzione della capacità di rivelazione dei frodatori e degli errori di identificazione. Si è scoperto che il numero minimo di chiamate per ciclo (rispetto allo scenario considerato, 200 providers e 400 intermediari) è di 100,000 chiamate.

Successivamente si è valutato gli errori di identificazione al variare del comportamento fraudolento (anche noto come attacco di *disguised malicious*). Si è scoperto che sotto la soglia del 20% di comportamento maligno il sistema non è in grado di identificare il frodatore, mentre per comportamenti spiccatamente fraudolenti, fra 80% e 100%, i frodatori sono sempre identificati.

In questa sezione si valuterà se la soglia del 20%, valore di comportamento fraudolento sotto il quale non il frodatore non viene individuato, è sufficiente a disincentivare i frodatori. Infatti l'uso di un sistema di identificazione costringerebbe i frodatori a mantenere bassa la soglia di traffico frodato per non essere individuati. Ci si potrebbe domandare se questa soglia è sufficiente a disincentivare i frodatori. Per farlo occorre valutare l'aspetto economico della frode e quantificare il tempo necessario al frodatore per avere un ritorno di investimento positivo.

Grazie ai dati ricavati da [5] relativamente alla frode di SIMBoxing è stato possibile analizzare nel dettaglio l'impatto economico della frode. Ciascun frodatore effettua fino a 18,000 minuti di chiamate fraudolente al giorno, ovvero circa 3,000 chiamate al giorno. Valutiamo in che modo possiamo considerare questi dati nello scenario utilizzato per la simulazione. Se consideriamo che i frodatori sono l'1% di 400, ovvero 4, si hanno in totale 12,000 chiamate al giorno fraudolente. Sappiamo inoltre che il traffico soggetto a frode è il 5% per cui si ricava che il numero di chiamate totali da analizzare è di 240,000 chiamate al giorno. Tuttavia nel caso la partecipazione dei *providers* si riduca ad essere il 10% degli operatori, le chiamate utili alla generazione dei feedback si riducono di un fattore 10. Ricordo infatti che se l'operatore di terminazione non partecipa al sistema non si ha modo di stabilire se una traccia sia o no fraudolenta e conseguentemente quella traccia non verrà considerata. Per questo motivo dobbiamo aumentare la finestra di tempo necessaria ad acquisire le transazioni a 2,400,000 chiamate, che per le considerazioni fatte prima equivale ad un arco temporale di 10 giorni.

Dalla analisi degli investimenti fatti da un frodatore per acquistare apparecchiature e servizi necessari alla realizzazione della frode sappiamo che il ritorno di investimento considerando i dati sopra (12,000 chiamate fraudolente al giorno) è di circa due settimane nel caso migliore (per il frodatore). Per cui se il frodatore operasse al 100% verrebbe individuato prima di aver compensato le perdite di investimento. In questo caso l'adozione del sistema di identificazione del frodatore avrebbe esito positivo.

Occorre però valutare il caso in cui il frodatore per non essere individuato riduce al 20% il numero di chiamate fraudolente inviate al dispositivo *SIM Boxing* e gestendo la restante parte in modo onesto. In questo caso il frodatore commette 2,400 chiamate fraudolente al giorno e per avere un ritorno di investimento impiegherà cinque volte tanto il tempo precedente, ovvero circa due mesi e mezzo. In questo caso il simulatore mostra che numero di cicli necessari per avere una rivelazione di tutti i frodatori è pari

a 3, per cui circa 30 giorni. Anche in questo caso risulta evidente che la frode non è più economicamente vantaggiosa.

7 Conclusioni

L'intento di questa tesi di laurea è quello di valutare l'efficacia e l'applicabilità di un approccio al problema dell'identificazione dei frodatori mai studiato prima.

Lo studio dell'evoluzione storica dell'infrastruttura telefonica e la valutazione sulle differenze nelle leggi e nelle tariffe in alcuni paesi del mondo ha messo in luce la complessità dello scenario telefonico internazionale. In un quadro così eterogeneo le dinamiche di mercato, in particolare la ricerca della tariffa più economica, hanno evidenziato la natura *untrusted* (non di fiducia) delle interconnessioni fra operatori telefonici.

Dalle ricerche effettuate nella letteratura scientifica relativa al contesto telefonico non sono emersi studi inerenti l'individuazione dei frodatori (solo delle frodi). Tuttavia in altri contesti, ad esempio nel *P2P File Sharing*, esiste un'ampia letteratura scientifica legata al problema di individuare ed eliminare i nodi maligni. Questo ambito di ricerca prende il nome di *Network of Trust*. Gli algoritmi di *Trust* vengono studiati in contesti totalmente *untrusted* dove nessun nodo si fida di nessun altro e non c'è una autorità centrale che stabilisce quale nodo è onesto e quale nodo è fraudolento. L'attribuzione di fiducia ad un nodo avviene sulla base dell'esperienza diretta, valutata tramite i *feedback* espressi dagli stessi partecipanti al sistema. Tramite simulazione è stato valutato l'uso dell'algoritmo di *TNASL* nel contesto sopra citato.

Per validare l'approccio descritto sopra è stato realizzato un simulatore in *Python*. Tramite simulazione è stato riprodotto uno scenario semplificato che emula l'interconnessione di più operatori nella gestione di una chiamata. Selezionando in modo randomico l'operatore di origine, quello di terminazione ed un certo numero di operatori intermediari si emula la generazione di una traccia telefonica completa di cui si sa chi sono gli operatori e se la chiamata generata è stata soggetta a frode oppure no.

Sulle tracce così generate è stato applicato l'approccio proposto in questa tesi per l'individuazione dei frodatori. In particolare è stato utilizzato l'approccio noto come *Trust Network Analysis with Subjective Logic*. I valori di reputazione ottenuti in uscita dal sistema sono stati utilizzati per classificare il comportamento dei nodi in onesti o fraudolenti. Confrontando i risultati ottenuti con i dati noti a priori si è valutato efficacia e problematiche del approccio proposto.

La simulazione è stata condotta sulla base dei dati acquisiti dalle statistiche pubblicate da *ITU* e come risultato di moltissime ricerche in rete, tutte accuratamente documentate. La difficoltà nel riprodurre il contesto reale, ovvero nel considerare le differenze fra tipologie di operatori (Mobili, fissi, internet, retail, wholesale, reseller...etc) e diversi volumi di traffico telefonico gestito, ha portato ad una semplificazione del contesto. In particolare il contesto studiato è stato limitato al caso in cui la chiamata origina o termina da un operatore locale e viene gestita da 4 operatori di transizione di tipo *VOIP wholesale*. Si è quindi assunto che né l'operatore di origine né quello di terminazione sono frodatori. Il frodatore si nasconde tra gli operatori *VOIP wholesale*.

Infine per valutare se l'approccio proposto abbia una utilità nello scenario "reale" della telefonia internazionale si è valutata la situazione in cui la variabilità nel comportamento dei frodatori porta al caso peggiore per la rivelazione dei frodatori, ovvero il caso *disguised*) e in cui la partecipazione dei frodatori è la più bassa tra i casi studiati,

ovvero il caso di *realistic cooperation*. infine i dati ottenuti sono state confrontati con l'analisi dei costi e dei benefici per i frodatori nel caso di frodi *SIM Boxing*.

In questo contesto è emerso che l'approccio proposto può essere considerato un approccio valido per risolvere il problema dell'identificazione dei frodatori che si nasconde lungo la "catena" di operatori che gestisce una chiamata fraudolenta.

Tuttavia le simulazioni hanno messo in luce il fatto che sul corretto funzionamento del sistema incidono (1) la variabilità nel comportamento dei frodatori e (2) la percentuale di operatori che partecipano (indirettamente) alla generazione dei feedback.

Nonostante le ipotesi fatte siano tutte supportate dal ragionamento logico, si conclude che per stabilire se l'approccio proposto sia effettivamente efficace ed accurato nel rivelare i frodatori è necessario modellare lo scenario (relazioni tra operatori e generazione delle tracce) in modo che rispecchi meglio la situazione reale. Tuttavia questa strada potrebbe non essere percorribile perché per essere attuata richiederebbe la condivisione da parte di molti operatori telefonici delle proprie tracce telefoniche e, non ultimo, la possibilità di confrontare il risultato ottenuto con il dato veritiero (il frodatore individuato è realmente un frodatore?).

8 Appendice A

In questa appendice sono riportati gli operatori che al 3 Aprile del 2019 sono stati identificati come operatori di telefonia mobile e VoIP wholesale providers e che sono stati conteggiati in questa tesi per configurare lo scenario di riferimento. Ho scelto di riportare la lista completa degli operatori come prova di veridicità circa le considerazioni fatte.

| MCCMNC | MCC | MNC | Country and Operator |
|--------|-----|-----|--|
| 24701 | 247 | 1 | Latvia,Latvian Mobile Phone |
| 24702 | 247 | 2 | Latvia,Tele2 |
| 24703 | 247 | 3 | Latvia,TRIATEL/Telekom Baltija |
| 24704 | 247 | 4 | Latvia,Beta Telecom |
| 24705 | 247 | 5 | Latvia,Bite Latvija |
| 24706 | 247 | 6 | Latvia,SIA Rigatta |
| 24707 | 247 | 7 | Latvia,SIA Master Telecom / Bite / MTS |
| 24708 | 247 | 8 | Latvia,SIA IZZI |
| 24709 | 247 | 9 | Latvia,SIA Camel Mobile |
| 24601 | 246 | 1 | Lithuania,Omnitel |
| 24602 | 246 | 2 | Lithuania,Bite |
| 24603 | 246 | 3 | Lithuania,Tele2 |
| 60502 | 605 | 2 | Tunisia,Tunisie Telecom |
| 60503 | 605 | 3 | Tunisia,Tunisiana / Orascom Telecom |
| 62201 | 622 | 1 | Chad,Zain/Airtel/Celtel |
| 62202 | 622 | 2 | Chad,Tchad Mobile |
| 62203 | 622 | 3 | Chad,Tigo/Milicom/Tchad Mobile |
| 62204 | 622 | 4 | Chad,Salam/Sotel |
| 338020 | 338 | 20 | Jamaica,LIME / Cable & Wireless |
| 338050 | 338 | 50 | Jamaica,DIGICEL/Mossel |
| 338070 | 338 | 70 | Jamaica,Claro / Oceanic Digital Jamaica Limited |
| 338180 | 338 | 180 | Jamaica,LIME / Cable & Wireless |
| 21803 | 218 | 3 | Bosnia & Herzegov.,Eronet Mobile / Public Enterprise Creation Telecom |
| 21805 | 218 | 5 | Bosnia & Herzegov.,M-Tel / RS Telecommunications JSC Banja Luka |
| 21890 | 218 | 90 | Bosnia & Herzegov.,BH Mobile |
| 60301 | 603 | 1 | Algeria,ATM Mobils |
| 60302 | 603 | 2 | Algeria,Orascom / DJEZZY |
| 60303 | 603 | 3 | Algeria,Wataniya / Nedjma |
| 21401 | 214 | 1 | Spain,Vodafone |
| 21403 | 214 | 3 | Spain,Orange |
| 21404 | 214 | 4 | Spain,Yoigo |
| 21405 | 214 | 5 | Spain,Movistar |
| 21408 | 214 | 8 | Spain,Euskaltel SA |
| 21412 | 214 | 12 | Spain,Contacta Servicios Avanzados de Telecomunicaciones SL |
| 21413 | 214 | 13 | Spain,Incotel Ingenieria y Consultoria SL |
| 21415 | 214 | 15 | Spain,BT Espana Compania de Servicios Globales de Telecomunicaciones SAU |
| 42505 | 425 | 5 | Palestinian Territory,Jawwal |
| 42506 | 425 | 6 | Palestinian Territory,Wataniya Mobile |
| 366110 | 366 | 110 | Dominica,C & W |
| 37002 | 370 | 2 | Dominican Republic,Claro (Compania Dominicana de Telefonos) |
| 37003 | 370 | 3 | Dominican Republic,TRIcon |
| 37004 | 370 | 4 | Dominican Republic,Trilogy Dominicana S. A. |
| 40468 | 404 | 68 | India, MTNL |
| 40470 | 404 | 70 | India, Hexacom |
| 40471 | 404 | 71 | India, BSN |
| 40509 | 405 | 9 | India,RELIANCE TELECOM |
| 40528 | 405 | 28 | India,Tata Teleservices Ltd |
| 40573 | 405 | 73 | India,Essar Spacetel |
| 405756 | 405 | 756 | India,Vodafone |
| 405810 | 405 | 810 | India,Aircel |
| 405824 | 405 | 824 | India,Videocon |
| 405845 | 405 | 845 | India,Idea |
| 405873 | 405 | 873 | India,Loop/Jio |
| 405875 | 405 | 875 | India,Uninor Assam |
| 405881 | 405 | 881 | India,S Tel |
| 405907 | 405 | 907 | India,MTS West |
| 405908 | 405 | 908 | India,Spice |
| 405912 | 405 | 912 | India,Etisalat DB |
| 20801 | 208 | 1 | France,Orange |
| 20803 | 208 | 3 | France,MobiQuThings |
| 20804 | 208 | 4 | France,SISTEER |
| 20807 | 208 | 7 | France,GlobalStar |
| 20809 | 208 | 9 | France,S.F.R. |
| 20814 | 208 | 14 | France,Lliad |
| 20820 | 208 | 20 | France,Bouygues Telecom |
| 20822 | 208 | 22 | France,Transatel SA |
| 20825 | 208 | 25 | France,Lycamobile SARL |
| 20826 | 208 | 26 | France,NRJ |
| 20827 | 208 | 27 | France,AFONE SA |
| 20828 | 208 | 28 | France,Astrium |
| 20831 | 208 | 31 | France,Mundio Mobile (France) Ltd |
| 20888 | 208 | 88 | France,Bouygues Telecom |
| 20889 | 208 | 89 | France,Omer/Virgin Mobile |
| 20892 | 208 | 92 | France,Association Plate-forme Telecom |
| 34011 | 340 | 11 | French Guiana,TelCell GSM |

| | | | |
|-------|-----|----|--|
| 54720 | 547 | 20 | French Polynesia |
| 20201 | 202 | 1 | Greece,Cosmote |
| 20203 | 202 | 3 | Greece,OTE Hellenic Telecommunications Organization SA |
| 20204 | 202 | 4 | Greece,Organismos Sidirodromon Ellados (OSE) |
| 20205 | 202 | 5 | Greece,Vodafone |
| 20207 | 202 | 7 | Greece,AMD Telecom SA |
| 20209 | 202 | 9 | Greece,Tim (Telecom Italia) /Wind |

Tabella 8: Mobile Operators in selected countries

| Country | VoIP Wholesale provider |
|-----------|-------------------------|
| Latvia | TELECOM VOIP |
| Latvia | Sotus SIA |
| Latvia | Rigatta |
| Latvia | Nulltel |
| Latvia | Xotel SIA |
| Latvia | MKTC |
| Latvia | Master Telecom SIA |
| Latvia | Sigis |
| Latvia | Telekomunikaciju Gru |
| Latvia | Phoneserve Telecom S |
| Latvia | Telekom Baltija |
| Latvia | IT Group |
| Lithuania | Voip shop.lt |
| Lithuania | Alderada |
| Lithuania | JSC "Medium Group |
| Lithuania | Medium Group |
| Lithuania | UAB Solocomas |
| Lithuania | Upnet Taide Balti |
| Tunisia | VoipBestChoice |
| Tunisia | Global Centrez |
| Tunisia | Skulls Networking C |
| Tunisia | NGN Concept |
| Tunisia | Maghreb telecom |
| Jamaica | Digital Domain Jamai |
| Jamaica | DOW Networks Jamaic |
| Jamaica | Avoxi |
| Jamaica | People's Telecom Ja |
| Jamaica | Knutsford Telecoms |
| Bosnia | Aneks |
| Bosnia | EPI |
| Algeria | samttelecom |
| Algeria | Zater ISP |
| Algeria | Techni Communicatio |
| Algeria | TilifouNET |
| Algeria | Maghreb telecom |
| Spain | Siptize |
| Spain | Masquevoz |
| Spain | Voipllama |
| Spain | Telefonicaweb Nuovovo |
| Spain | SipCel Telecom |
| Spain | NuovoVoip |
| Spain | Tus Comunicaciones |
| Spain | Telemo |
| Spain | TelefonicaWeb |
| Spain | TELEVOIP CANARIAS S.L |
| Spain | GRUPALIA INTERNET S.A |
| Spain | VoiceTrunk |
| Spain | Teleformacion sl |
| Spain | Global Red Soluciones |
| Spain | Team To Win |
| Spain | Cuba Telecom |
| Spain | Nonotel ahorro de lla |
| Spain | Balear Networks |
| Spain | ChipTelecomunicacione |
| Spain | CLEPAR TELECOMUNICACI |
| Spain | Cecil Projects Ltd. |
| Spain | taritelecom espa?±a |
| Spain | VozTelecom |
| Spain | DialTrix Telecom |
| Spain | invoco. |
| Spain | Vozelia |
| Spain | Waymoble |
| Spain | ASP |
| Spain | tantelcom |
| Spain | Amitelo Wireless |
| Spain | Ivoice Communications |
| Spain | Internet Telecomunica |
| Spain | Glovip Winona Tecno |
| Spain | Eurocomm Group |
| Spain | Elmer Sinaza, sp. z o |
| Spain | Tecnica Instrumentos d |
| Spain | www.andescall.com |
| Spain | Meissen Investment |
| Spain | Voiptel, S.L. |
| Spain | Holaphone |
| Spain | Can Talk |
| Spain | Telemintutos |
| Spain | Sagitel Telecom |
| Spain | Asistelc VoIP |
| Spain | VozIpGlobal 1 |
| Spain | VozIpGlobal.com |
| Spain | C4 Telecom, S.L. |

| Country | VoIP Wholesale provider |
|--------------------|-------------------------|
| Spain | TerraSip S.A. |
| Spain | Alturaphone |
| Spain | Vocalpad |
| Spain | Universal Telecom |
| Spain | Alta Tecnologia en Co |
| Spain | VoIPCall SL |
| Spain | Quantum Sistemas, SA |
| Spain | Dile4G |
| Spain | Bankoi |
| Spain | Amitelo Communication |
| Spain | Telcom Business Solut |
| Spain | Nodo53 |
| Spain | KEM Informatica y Tel |
| Spain | Isivos, S.L. |
| Spain | Ingetel Ing. Telefoni |
| Spain | ENCOM Telecom |
| Palestine | SuperLink Communi |
| Palestine | SuperLink Communi |
| Palestine | Jinan Telecommuni |
| Palestine | Jinan Telecommuni |
| Palestine | Madar Communicati |
| Palestine | Fusion Co. |
| Palestine | Fusion Co. |
| Palestine | GlobalCom |
| Palestine | Masar Communicati |
| Palestine | VQTel |
| Dominican Republic | VoIP PLA |
| Dominican Republic | Wind Tel |
| Dominican Republic | Corsami |
| Dominican Republic | DgTec, S |
| India | Adore Infotech Pvt. L |
| India | IVON NETWORKING SOLUT |
| India | STCVOIP SERVICES |
| India | Marc Resources |
| India | Mabrook International |
| India | Vindcomm Infosoft |
| India | ISD Networks |
| India | IDEA VOIP |
| India | IDEA VOIP |
| India | Mabrook Internatinal |
| India | Mabrook Media Technol |
| India | Mabrook Media Technol |
| India | Mabrook Media Technol |
| India | Talk2Free Telecomm. |
| India | Whitelene Communicati |
| India | Laczone Technologies |
| India | Laczone Technologies |
| India | Easytel Communication |
| India | Mabrook Media Pvt Ltd |
| India | Laczone Technologies |
| India | Laczone Technologies |
| India | Mabrook VoIP |
| India | Divox FZ LLC |
| India | Spectranet |
| India | Divox Communicatio P |
| India | Aeon Media Technologi |
| India | K.R.SOLUTIONS (pvt) L |
| India | Opto Network Pvt. Ltd |
| India | Altius Info Solutions |
| India | KRVOIP SOLUTIONS (PVT |
| India | Dail2Buddy Telecomm (|
| India | IT Voipcall |
| India | zainvoip communicatio |
| India | Call2Baby.com |
| India | ITPL INNOVMOX |
| India | laczone communication |
| India | OXY Communications |
| India | DONA VOIP SERVICE IND |
| India | Laczone Technologies |
| India | Mabrook Media technol |
| India | Aeon Media Technologi |
| India | Matrix Shell |
| India | voxvalley technologie |
| India | VIVA COMMUNICATIONS P |
| India | Viva Communications |
| India | Voice Calls Solutions |
| India | Pulse Telesystems Pvt |
| India | PRAB Communications |
| India | BSNL VOIP TELECOMMUNI |

| Country | VoIP Wholesale provider |
|---------|-------------------------|
| India | hms4call |
| India | Astral Communications |
| India | viva communications P |
| India | MOON2WORLD TELECOMM |
| India | Reliance Communicatio |
| India | Sonizon Telecom |
| India | Beacon Infotech |
| India | ExcellentComputerServ |
| India | FRONTLINE |
| India | Micro Village Communi |
| India | VIVA COMMUNICATION PV |
| India | Askvoip |
| India | Askvoip |
| India | IPTEL TELECOMMUNITION |
| India | IPTEL TELECOMMUNITION |
| India | 1A NMR Communication |
| India | 1A NMR Communication |
| India | KGVOIZE |
| India | KGVOIZE |
| India | SRS TECHNOLOGIC |
| India | Cordia LT |
| India | Cordia LT |
| India | Gsm support |
| India | Aircel Limited |
| India | Openvoips Solution |
| India | sip2save |
| India | sip2save |
| India | Gaytes Information Sy |
| India | Gaytes Information Sy |
| India | CHINTAN NETWORKS PRIV |
| India | CHINTAN NETWORKS PRIV |
| India | Gaz Consultancy Servi |
| India | N Core BPO Solutions |
| India | CheetChat |
| India | CheetChat |
| India | Xoom Technologies |
| India | ANN Telecom |
| India | 1 WORLD VoIP |
| India | 1 WORLD VoIP |
| India | viva communication Pv |
| India | Computech VoIP |
| India | VIVA Communications |
| India | Easytel |
| India | Open Coders Pvt Ltd |
| India | endtoenditsolutions |
| India | endtoenditsolutions.c |
| India | smahs |
| India | smahs |
| India | Sabbal Kompany |
| India | Myown Infotech Pvt Lt |
| India | SPM Information Solut |
| India | Worldphone Internet S |
| India | WORLDPHONE Inernet Se |
| India | Bankai International |
| India | xenn networks |
| India | xenn networks |
| India | sai communication |
| India | sai communication |
| India | Voipwala India |
| India | Voipwala India |
| India | Gulffoon |
| India | Tata Communications L |
| India | Tata Communications L |
| India | A1 Calling solutions, |
| India | VoIP Hardware Pvt Ltd |
| India | VoIP Hardware Pvt Ltd |
| India | Liya telecom Pvt Ltd |
| India | Liya Telecome Pvt Ltd |
| India | sip2save Communicatio |
| India | sip2save |
| India | K Cube Communication |
| India | Phonology It Solution |
| India | YOU Telecom India Pvt |
| India | flixbay |
| India | Vomart International |
| India | Reliance Communicatio |
| India | Mobitel Communication |
| India | Best VOIP Service Pro |
| India | 3xTechnologies Privat |
| India | 3xTechnologies Privat |
| India | Vb Plustech |
| India | Panamaxil |

| Country | VoIP Wholesale provider |
|---------|-------------------------|
| India | deDSL Internet Pvt. |
| India | DeDSL Internet Pvt. |
| India | Intezar |
| India | Net2phoneworld.com |
| India | SwissFone India Ventu |
| India | Parshwa Communication |
| India | Pacific Internet Indi |
| India | Sachitel Communicatio |
| India | IPage Telecom |
| India | VcallWorld |
| India | Voizbay |
| India | Pulavarty technologie |
| India | CYBASE DOMAIN INC. |
| India | Cybase Domain Inc |
| India | Vivel Inc |
| India | IPvaani Inc |
| India | Enterux Solutions |
| India | Net 4 India Limited |
| India | Infeetalk |
| India | World Tele Net (India |
| India | Novanet Technologies |
| India | Syndrome Technologies |
| India | Net4India |
| India | Bicnet Infoservices |
| India | Earlybirds |
| India | IP Communications |
| India | Techway Communication |
| India | Maksh IT Solutions & |
| India | Voipcarrier Services |
| India | Chaaaya Communication |
| India | Anyuser Telecom (Indi |
| India | Fass Tel VoIP Solutio |
| India | Decibels Voip & IT S |
| India | P.K. Creations |
| India | Pulse Telesystems PV |
| India | Sanver E Solutions |
| India | IRC |
| India | Amtel |
| India | Shree Tele Network Pv |
| India | Iservicesonline |
| India | Voip Communication |
| India | Karunus Telecom |
| India | Pioneer Online Pvt. L |
| India | Vebtel Obconic Intern |
| India | Infoneia Systems |
| France | Altern telecom / Clo |
| France | ConnectPhone Ltd |
| France | CallsUp! |
| France | Siplab |
| France | Oufitel |
| France | Aianna Corp |
| France | VALIDYNE SYSTEM |
| France | VALIDYNE |
| France | NET1C |
| France | bisatel |
| France | Longphone |
| France | Longphone |
| France | CHEZSIP |
| France | TRADE OFF |
| France | Multiline Sarl |
| France | Multiline Sarl |
| France | Andrexen |
| France | Andrexen |
| France | ClearCall |
| France | Activatel |
| France | Activatel |
| France | OpenIP |
| France | Provence Media |
| France | Provence Media |
| France | Dimension Telecom |
| France | Aianna Corp FR |
| France | Aianna Corp |
| France | Ipnotic Telecom |
| France | Vivaction |
| France | Gccom |
| France | Telemedia Communicat |

| Country | VoIP Wholesale provider |
|---------|-------------------------|
| France | ITD Paris |
| France | International Teleco |
| France | Distriion |
| France | Toubatel |
| France | Veco |
| France | Direct Centrex |
| France | Tectip |
| France | VoxIP Telecom |
| France | Plug to Tel |
| France | Eone Telecom |
| France | MIDI Telecom |
| France | Tradingcom Europe |
| France | Wengo |
| France | 4 Place Felix Eboué |
| France | Telehouse 2 |
| France | Plug and Tel |
| France | WideVOIP |
| France | Kast Telecom |
| France | Callnetworks |
| France | Flexfone France |
| France | I.C.S. |
| France | ME&NA Consulting |
| France | Newtechtelecom |
| France | Top Satellite |
| France | Maghreb telecom |
| France | GlobalTransit |
| Greece | Modulus SA |
| Greece | Northwest Communicat |
| Greece | Northwest Communicat |
| Greece | Altec Telecoms |
| Greece | Neuron S.A |
| Greece | Omnivoice |
| Greece | EasyCom Telecommunic |
| Greece | Webacall |
| Greece | T Point Systems |
| Greece | NET ONE A.E. GREECE |
| Greece | Net One A.E. Greece |
| Greece | Kinetix Tele.com Hel |
| Greece | Inter Telecom |
| Greece | Interconnect |

Riferimenti bibliografici

- [1] Over-The-Top Bypass: Study of a Recent Telephony Fraud http://s3.eurecom.fr/docs/ccs16_sahin.pdf
- [2] OECD, INTERNATIONAL TRAFFIC TERMINATION [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/CISP\(2013\)9/FINAL&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/CISP(2013)9/FINAL&docLanguage=En)
- [3] FCC, Trends in Telephone Service, Table 15.3 (pag. 116) https://transition.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/IAD/trend801.pdf
- [4] CFCA, Global Fraud Loss Survey 2017 <https://www.scribd.com/document/368471387/2017-Global-Fraud-Loss-Survey-CFCA-pdf>
- [5] Simbox fraud, Akhil Singh Rawat 2016 https://www.slideshare.net/AkhilRawat/sim-box?qid=3b38b99e-3041-47d3-9501-c0d0b850b0a1&v=&b=&from_search=3
- [6] R. S. A. H. Elmi, S. Ibrahim, "Detecting SIM Box Fraud Using Neural Network," IT Converg. Secur. 2012, vol. 215, pp. 575 582, 2013
- [7] Murynets, M. Zabarankin, R. P. Jover, and A. Panagia, "Analysis and detection of SIMbox fraud in mobility networks,"
- [8] B. Reaves, E. Shernan, A. Bates, H. Carter, and P. Traynor, "Boxed Out : Blocking Cellular Interconnect Bypass Fraud at the Network Edge"
- [9] https://www.researchgate.net/publication/254047289_Scam_and_fraud_detection_in_VoIP_Networks_Analysis_and_countermeasures_using_user_profiling
- [10] Josang, Audun & Hayward, Ross & Pope, Simon. (2006). Trust Network Analysis with Subjective Logic. Proc of the 29th Australasian Computer Science Conference, CRPIT Volume 48, Hobart, Australia
- [11] D. Kamvar, Sepandar & Schlosser, Mario & Garcia-molina, Hector. (2003). The EigenTrust Algorithm for Reputation Management in P2P Networks. The EigenTrust Algorithm for Reputation Management in P2P Networks.
- [12] Telecommunications Infrastructure in Manhattan <https://cromwell-intl.com/travel/usa/new-york-internet/>
- [13] List of Major IXPs https://en.wikipedia.org/wiki/List_of_Internet_exchange_points_by_size
- [14] Mix, Milan Internet Exchange Point <https://www.mix-it.net/servizi-di-peering/#content-2>
- [15] Teclos T1 Lists https://en.wikipedia.org/wiki/Tier_1_network
- [16] Peering and transit <https://arstechnica.com/features/2008/09/peering-and-transit/4/>
- [17] Tesspay Wholesale Provider <https://www.tesspay.io/>