VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

AKTIVNÍ IP GEOLOKACE PRO VERIFIKACI POZIC STANIC V INTERNETU

POJEDNÁNÍ O DIZERTAČNÍ PRÁCI
TREATISE ON DOCTORAL THESIS

DOKTORAND Ing. JIŘÍ BALEJ
PH.D. CANDIDATE

Brno 2017



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ
DEPARTMENT OF TELECOMMUNICATIONS

AKTIVNÍ IP GEOLOKACE PRO VERIFIKACI POZIC STANIC V INTERNETU

ACTIVE IP GEOLOCATION FOR VERIFICATION HOST POSITION IN INTERNET

POJEDNÁNÍ O DIZERTAČNÍ PRÁCI
TREATISE ON DOCTORAL THESIS

AUTOR PRÁCE
AUTHOR
Ing. JIŘÍ BALEJ

ŠKOLITEL
SUPERVISOR
doc. Ing. DAN KOMOSNÝ, Ph.D.

BRNO 2017

ABSTRAKT

Dizertační práce se zabývá způsoby nalezení geografické polohy zařízení v síti Internet při znalosti IP adresy. Tento proces se nazývá IP geolokace a je v současnosti řešen pomocí geolokačních databází nebo za využití výsledků měření síťových parametrů k cílové IP adrese. Nevýhodou dnešních geolokačních databází je, že některé poskytované polohy nejsou správné a mohou vykazovat velkou odchylku od správné polohy. Cílem této práce je vyvinout metodu, která by na základě měření dokázala ověřit správnost pozice z geolokační databáze. Z tohoto důvodu je v práci podrobně rozebrán vliv parciálních částí zpoždění, které ovlivňují výpočet maximální vzdálenosti na základě změřeného zpoždění mezi referenční stanicí a cílovou IP adresou. Ze stejného důvodu je v práci popsáno dlouhodobé měření zpoždění, kde je řešena přesnost IP geolokace za použití kalibračních dat z dřívějších měření. Navržená metoda Cable Length Based Geolocalisation (CLBG) je postavena na vlastnostech dílčích složek zpoždění, které jsou závislé na délce přenosových médií. Metoda ze změřeného obousměrného zpoždění vyloučí vliv zpoždění generovaného mezilehlými prvky a koncovými stanicemi a za použití rychlosti šíření signálu přenosovým médiem určí geografickou vzdálenost. Dále byl experimentálně zjištěn parametr nepřímého vedení kabelů, jež je použit pro určení mezních hranic. Průnik mezních hranic jednotlivých referenčních bodů je následně použit ke stanovení regionu, kde se IP adresa nachází. Výsledky této metody při geolokaci jsou lepší než jednoduché metody (ShortestPing, GeoPing a SOI) a srovnatelné s metodami pokročilejšími (CBG a Octant). Nevýhodou vytvořené metody je velikost regionu, kde se stanice nachází, což je ale dáno jejím účelem. Pro zjištění správnosti informace z geolokační databáze slouží ověření, zda její pozice leží ve zmíněném regionu.

KLÍČOVÁ SLOVA

IP geolokace, lokalizace, geolokační databáze, zpoždění, multilaterace, RTT, CBG, Octant, CLBG.

ABSTRACT

Dissertation thesis deals with methods for finding the location of the device in the Internet, based on knowledge of the IP address. The process is called IP geolocation and is currently solved by geolocation databases or by measurement of network properties to the IP address. The disadvantage of nowadays geolocation databases is an incorrect information about some locations, because they can be in large distance from correct position. The aim of the thesis is to develop a method for verification of a position from geolocation database using delay measurement. Because of it, there is a detail analysis of influence of partial delays on the distance estimation accuracy, calculated using measured delay between the landmark and the target IP address. For the same reason, long-term delay measurement was performed, where the IP geolocation accuracy was compared using calibration data from previous measurements. On this background, Cable Length Based Geolocalisation (CLBG) method is proposed. Principle of this method is built on the properties of partial delays, which depend on the length of transport media. Firstly, the method measures round trip time (rtt), which is subsequently lowered by intermediate devices and end stations delay. The geographical distance is estimated using signal speed in the transport media. Further, the winding media parameter is established, which is used to determine a constraint around the landmark. The intersection of all constraints defines the area, where the target IP is. The IP geolocation using CLBG gives better results than simpler methods (ShortestPing, GeoPing and SOI), in comparison with more advanced methods (CBG and Octant) the accuracy is similar. The disadvantage of the CLBG method is the size of region, where the target lies, but this is due to its purpose. The position found in geolocation database can be checked by evaluation if it lies in the region.

KEYWORDS

IP geolocation, localization, geolocation databases, delay, multilateration, rtt, CBG, Octant, CLBG.

BALEJ, Jiří. *Aktivní IP geolokace pro verifikaci pozic stanic v Internetu*. Brno, 2017, 38 s. Pojednání o dizertační práci. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Školitel(ka): doc. Ing. Dan Komosný, Ph.D.

Vysázeno pomocí balíčku `thesis` verze 2.63; `http://latex.feec.vutbr.cz`

OBSAH

1 Úvod **6**

2 Přehled současného stavu problematiky **7**
2.1 Pasivní IP geolokační služby 7
2.1.1 IP adresy a způsoby jejich přidělování 7
2.1.2 Analýza DNS záznamů 8
2.1.3 Databáze pro IP geolokaci 8
2.2 Aktivní IP geolokační metody 9
2.2.1 Metoda GeoPing 10
2.2.2 Metoda ShortestPing 10
2.2.3 Metoda Constraint Based Geolocation 10
2.2.4 Metoda Speed of Internet 11
2.2.5 Metoda Octant 12
2.2.6 Metoda Spotter 13
2.2.7 Srovnání aktivních geolokačních metod 13
2.3 Parametry komunikačního řetězce ovlivňující IP geolokaci 13
2.3.1 Zpoždění vznikající v koncových zařízeních 14
2.3.2 Zpoždění vznikající na přenosových linkách 15
2.3.3 Zpoždění vznikající v mezilehlých zařízeních 15
2.3.4 Zpoždění celého komunikačního řetězce 16

3 Cíle dizertační práce **17**

4 Analýza parametrů ovlivňujících odhad vzdálenosti **18**
4.1 Měření v rozlehlé síti se známou topologií 18
4.1.1 Zpoždění na jedno mezilehlé zařízení 18
4.1.2 Srovnání délky kabelu s přímou vzdáleností 19

5 Variabilita zpoždění v průběhu času **20**
5.1 Sledování změny zpoždění v Internetu v průběhu času 20
5.1.1 Změna zpoždění v průběhu času 20
5.1.2 Vliv změn obousměrného zpoždění na kalibrační funkci 21
5.2 Přesnost IP geolokace při použití kalibračních dat z předchozích měření 21
5.2.1 Přesnost IP geolokace při použití den starých kalibračních dat 21
5.2.2 Přesnost IP geolokace při použití týden starých kalibračních dat 22
5.2.3 Přesnost IP geolokace při použití čtvrt roku starých kalibračních dat 23

5.3 Poznatky užitečné pro návrh nové geolokační metody 24

6 Geolokace založená na analýze přenosového zpoždění 25
6.1 Předpoklady pro návrh geolokační techniky 25
6.2 Parciální složky zpoždění pro výpočet geografické vzdálenosti 25
6.3 Vliv nepřímého vedení kabelů na výpočet vzdálenosti 26
6.4 Geodetický aparát pro výpočet geografické polohy 26
6.4.1 Hranice okolo referenčního bodu 26
6.4.2 Region průniku všech mezních vzdáleností 27
6.4.3 Těžiště a obsah regionu průniku mezních vzdáleností 27
6.5 Kalibrační proces u metody CLBG 27
6.6 Srovnání výsledků nové metody 28
6.6.1 Porovnání výsledků CLBG s metodami GeoPing, Shortest-Ping a SOI 28
6.6.2 Porovnání výsledků CLBG s metodami Octant, CBG a SOI . 28

7 Ověření důvěryhodnosti záznamů geolokačních databází 30
7.1 Ověření důvěryhodnosti údajů z geolokačních databází 30
7.2 Srovnání přesnosti geolokačních databází 31

8 Závěr 33

Literatura 34

Publikace autora 36

1 ÚVOD

Geografická lokalizace je způsob nalezení skutečné pozice zařízení a tím pádem uživatele, který toto zařízení využívá. Poloha je určená zpravidla pomocí názvu místa (stát, region, město) nebo geografických (zeměpisných) souřadnic. Tato práce se zabývá geografickou lokalizací zařízení s IP adresou – tzv. IP geolokací, v praxi se jedná o počítače, servery, mobilní zařízení atd. Tato zařízení nejsou vždy vybavena přijímačem GPS a je tak obtížné určit přesně jejich polohu. Při použití geolokačních databází je možné zařízení zařadit do příslušné země v lepším případě určit město, ve kterém se nachází [11]. Při použití aktivní IP geolokace (založené na ICMP odpovědi IP adresy na dotaz) jsou dnes dosahované přesnosti v řádech desítek kilometrů [6].

Použitím IP geolokace můžeme zajistit například: zvýšení pohodlí uživatele (zobrazení lokálního jazyka a měny, místních zpráv, ...), zabezpečení důvěrné stránky (přihlašování k elektronickému bankovnictví a soukromým účtům), omezení přístupů k obsahu (např. z důvodu vysílacích práv), odhalení pachatelů internetové kriminality, předcházení zneužití platebních údajů a cílení relkamy.

Mimo výhod plynoucích z geolokace můžeme v souvislosti se znalostí pozice uživatele nalézt i problémy. Obecně vzato, informaci o poloze určité osoby lze považovat za důvěrnou, protože může posloužit k nekalým účelům – sledování osoby, odhalení doby, kdy se daná osoba nevyskytuje doma atd. Z tohoto důvodu je poskytnutí přesných informací, například prostřednictvím WiFi signálu, zpravidla nutné potvrdit uživatelem. Oproti tomu lokalizace pouze na základě IP adresy zařízení je dostupná vždy, neboť tyto informace (v hlavičce paketu) dojdou až příjemci zprávy. Pokud uživatel využívá NAT (Network Adresss Translation) nebo Proxy serveru, není dostupná přímo IP adresa zařízení, ale jiná adresa, často však ze zařízení v blízkosti skutečného uživatele. Možností, jak skrýt svoji identitu a tím ztížit lokalizaci, je použít anonymizační sítě – např. Tor (The Onion Router) nebo využít služeb VPN (Virtual Private Network).

Tato práce se věnuje problému ověření geografické polohy zjištěné z databáze pomocí nově navržené aktivní geolokační metody. Po úvodu do problematiky v kapitole 1, jsou v kapitole 2 podrobně rozebrány veškeré současné přístupy k nalezení polohy stanice při znalosti IP adresy – tzv. IP geolokace. Následující kapitola 3 uvádí cíl práce. Kapitola 4 pak obsahuje analýzu a měření vlastností zpoždění komunikačního řetězce, které přímo ovlivňují výpočet vzdálenosti. Následně je v kapitole 5 uveden popis výsledků dlouhodobého měření zpoždění. Kapitola 6 se pak podrobně věnuje návrhu nové metody i popisu geodetických výpočtů pro určení cílových souřadnic na povrchu Země. Použití této navržené metody v praxi je uvedeno v kapitole 7. Závěru a diskuzi výsledků je věnována kapitola 8.

2 PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY

V současnosti je pro geolokaci elektronického zařízení používáno několik přístupů. Způsoby geolokace můžeme rozdělit do tří základních skupin, dle jejich principu. První skupina zahrnuje zařízení obsahující rádiový přijímač, který je buď primárně určen ke geolokaci (GPS) nebo je geolokace jeho vedlejší funkcí (GSM, WiFi). Pro další dvě skupiny se obvykle užívá pojem IP geolokace – nalezení geografické pozice stanice za použití IP adresy. Tyto metody se dají rozdělit na pasivní a aktivní. Pasivní metody využívají statických záznamů v databázích, kdežto aktivní metody provádějí měření a na jeho základě je rozhodují o poloze. V praxi se můžeme setkat i s geolokací za využití kombinace těchto dvou přístupů.

Výsledkem geolokace je poloha, kterou je možné určit v několika úrovních – kontinent, stát, region, město, PSČ, přesná adresa a geografické souřadnice (např. ve formátu WGS84). Pasivní IP geolokační služby fungují často na komerční bázi, a tak úroveň přesnosti pozice stanice je obvykle úměrná ceně. Aktivní IP geolokace a geolokace s rádiovým přijímačem většinou vrací polohu ve formátu zeměpisných souřadnic s určitou tolerancí.

2.1 Pasivní IP geolokační služby

Pasivní IP geolokace je založena na vyhledávání záznamu o pozici bez použití aktivního měření. Velmi často jedná o databázi IP adres a příslušných geografických údajů, další možností je použít informací ze systému DNS nebo z databáze registrátorů IP adres. Protože jsou všechny tyto systémy založeny na statických záznamech, jsou velice náročné na správu a dochází u nich k systémovým chybám. Často se v databázi nachází chybný záznam – například stejný pro skupinu IP adres, které patří zařízením nacházejícím se na různých místech. Tyto metody mají také problém s konvergencí při přesunu stanice do nové lokace. Oproti těmto nevýhodám je hlavní výhodou rychlost a v případě správného záznamu také přesnost.

2.1.1 IP adresy a způsoby jejich přidělování

Většina geolokačních databází vychází z údajů poskytovaných registrátory IP adres (RIR). Záznam v RIR databázi obsahuje rozsah přidělených IP adres a název organizace, jež má tyto IP adresy registrovány. Velmi často se však sídlo registrující organizace neshoduje se skutečnou polohou všech stanic z přiděleného rozsahu IP adres, především z důvodu přesunů IPv4 adres mezi různými lokalitami. Naproti

tomu IPv6 má obrovský adresní prostor, kde není nutné přesouvat adresní rozsahy, na druhou stranu však narůstají nároky na velikost databází těchto adres.

2.1.2 Analýza DNS záznamů

Mezi pasivní metody pro zjištění polohy IP adresy je možné zařadit i analýzu DNS (Domain Name System) záznamů. Použitím zpětného (reverzního) překladu DNS lze pro IP adresu zjistit doménové jméno a z něj je možné vyčíst indicie o poloze. Některé koncové stanice a směrovače mají ve svém doménovém jménu uvedenu zkratku označující město, ve kterém se nachází [2]. Tyto záznamy však nejsou nijak zaručeny, aby vznikl chybný záznam, stačí přestěhovat směrovač do nové lokality a nezměnit jeho doménové jméno.

Druhou možností, jak využít DNS, je rozšíření definované v RFC 1876, které dovoluje přidat záznam o poloze DNS serveru. Bohužel většina DNS serverů nemá tento parametr vyplněn a pokud je vyplněn, tak není nijak ověřený, takže ke geolokaci jej lze použít pouze jako pomocný údaj.

2.1.3 Databáze pro IP geolokaci

Vzhledem k různému využití geolokačních databází tyto poskytují různé úrovně informací o IP adrese. V případě, že je nutné například omezit přístup z jiných zemí, postačuje znalost státu, což bývá nejjednodušší verze výstupu. Další verze databází poskytují informace o regionu, městu, PSČ, zeměpisných souřadnicích, typu a rychlosti připojení, poskytovateli připojení (ISP) nebo i demografické údaje jako velikost sídla, hustota zalidnění, bohatství, kriminalita,... Pro potřeby této práce budou využity především zeměpisné souřadnice uváděné u záznamů, přestože se tyto často vztahují k centru města, případně regionu ve kterém se tato IP nachází. V následujících kapitolách je uveden popis geolokačních databází, které jsou v kapitole 7 otestovány.

GeoIP2

Společnost MaxMind disponuje databází GeoIP2, jejíž obsah je vytvářen analýzou informací o poloze, jež o sobě uživatelé sami vyplnili na různých stránkách a dle [12] obsahuje 99.9999 % veškerých používaných IPv4 a IPv6 adres. GeoIP2 poskytuje tři placené úrovně informací – nejlevnější (stojí \$0,000 1 za záznam) poskytuje pouze kontinent a zemi ve které se IP adresa nachází. Vyšší úroveň (\$0,000 4 za záznam) poskytuje navíc město, PSČ, zeměpisné souřadnice, přesnost a další obdobné informace. Nejdražší verze (\$0,002 za záznam) obsahuje navíc informace o typu uživatele,

jeho příjmu a další demografické informace. Kromě toho MaxMind poskytuje geolokaci i bezplatně, a to prostřednictvím databáze GeoLite2, která obsahuje informace o zemi a městě, ve kterém se IP adresa nachází. Bezplatná verze je však méně přesná a méně často aktualizovaná.

IP2Location

IP2Location je geolokační databáze pro zjištění různých dat o poloze IPv4 i IPv6 adresy. Celkem je k dispozici 24 různých variant poskytovaných informací počínaje informací o zemi za $49 za rok [7]. Oblíbená verze stojí desetkrát tolik a zahrnuje zemi, region, město, PSČ a zeměpisné souřadnice. Nejvyšší verze stojí $1849 za rok, a kromě již zmíněného poskytuje například informace o časovém pásmu, ISP, rychlosti připojení a kódu nejbližší meteorologické stanice. Databázi je možné si vyzkoušet, zdarma je poskytováno 50 záznamů denně (pro registrované uživatele 200 záznamů denně). Kromě komerční verze a demo přístupů je možné využít i Open source verzi databáze s názvem IP2Location Lite, která má však deklarovanou nižší přesnost a stejnou polohu uvádí vždy pro blok 256 IP adres (IPv4 prefix /24).

Komerční verze databáze dle [7] určí správně zemi ve $>99{,}5$ % případů (město ve >80 %, oproti tomu Lite verze má přesnost státu >98 % a města >60%. Dle webových stránek [7] je každý měsíc aktualizováno 15 % záznamů, což znamená průběžné zpřesňování databáze, případně aktualizace některých záznamů z důvodu přesunu zařízení nebo změny jeho IP adresy.

Geobytes

Geobytes je jeden z nejstarších (od roku 1999) poskytovatelů geolokačních informací na Internetu a obsahuje informace o poloze všech rozsahů IPv4 adres, které se objevují v BGP tabulkách směrovačů v Internetu [4]. Kromě informací o poloze (stát, město, zeměpisné souřadnice) obsahuje Geobytes databáze i demografické informace jako (národnost, měnu, populaci a další). Informace z databáze jsou poskytovány zdarma až do 16 384 dotazů za hodinu, více je však možné si zaplatit pomocí VIP přístupu ($9.99 za 100 000 dotazů). Dle [4] má databáze přesnost 97 % ve správném určení země a 75 % v určení města (s tolerancí 50 km).

2.2 Aktivní IP geolokační metody

Aktivní IP geolokace je založena na měření zpoždění a případně dalších síťových parametrů mezi stanicí se známou polohou (referenčním bodem – landmarkem) a lokalizovanou stanicí. Většinou se k lokalizaci jedné stanice používá větší množství referenčních bodů (řádově desítky). Jelikož probíhá několik měření mezi referenčními

body a lokalizovanou stanicí, dochází k navýšení síťového provozu úměrně počtu měření a počtu referenčních bodů.

Princip většiny aktivních metod je založen na korelaci mezi zpožděním a geografickou vzdáleností. To je dáno vlivem zpoždění způsobeného rychlostí šíření signálu v médiu, které je hlavní složkou zpoždění na dlouhých vzdálenostech. Podrobněji se o zdrojích a obvyklé velikosti parciálních zpoždění se píše v kapitole 2.3.

2.2.1 Metoda GeoPing

Geoping [14] je nejstarší geolokační metoda založená na měření zpoždění. Ke své činnost potřebuje velké množství pasivních referenčních bodů (uzlů se známou polohou) a několik aktivních sond (uzlů provádějících měření). Nevýhodou této metody je určení výsledné polohy jako místa, kde leží jeden z referenčních bodů, čímž je omezena přesnost metody.

Princip metody je v porovnání vektorů zpoždění příslušejících referenčním bodům a lokalizované stanici. Vektor zpoždění obsahuje změřenou dobu přenosu informace mezi referenčním bodem a všemi sondami. Stejný vektor je změřen pro lokalizovanou stanici a následně je srovnán s vektory referenčních bodů k nalezení nejvíce podobného vektoru. Pro určení nejpodobnějšího vektoru je vytvořen M rozměrný prostor (rovný počtu sond), v němž je nalezen vektor s nejmenší eukleidovskou vzdáleností k hledanému vektoru [14].

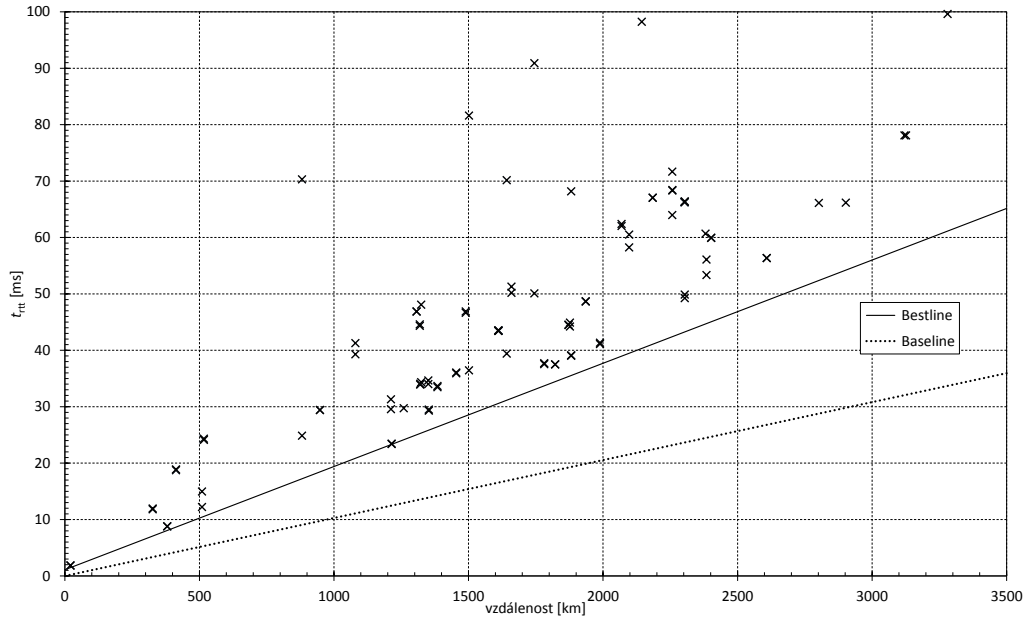
2.2.2 Metoda ShortestPing

Principiálně nejjednodušší IP geolokační metodou založenou na měření zpoždění je ShortestPing [8]. Tato metoda vyžaduje velké množství rovnoměrně rozmístěných referenčních bodů se známou polohou. Metoda zjišťuje zpoždění mezi lokalizovanou stanicí a všemi referenčními body, výsledná pozice je přisouzena poloze referenčního bodu s nejmenší hodnotou zpoždění. Přestože je tato metoda jednoduchá a výslednou pozici přisuzuje jednomu z referenčních bodů, v některých případech dosahuje tato metoda lepších výsledků než některé složitější metody (např. GeoPing).

2.2.3 Metoda Constraint Based Geolocation

Constraint Based Geolocation (CBG) [5] ke své činnosti využívá multilaterace známé z rádiového určování polohy. Princip metody tkví ve využití vztahu mezi geografickou vzdáleností a zpožděním k vytvoření tzv. hranice nejvzdálenějšího možného umístění stanice. Tato hranice je určena přepočtem zpoždění na základě tzv. Bestline, což je přímka vytvořená při kalibraci a udává vztah mezi zpožděním a vzdáleností

pro příslušný referenční bod. CBG tedy pro činnost potřebuje množinu aktivních referenčních bodů se známou polohou (landmarků).



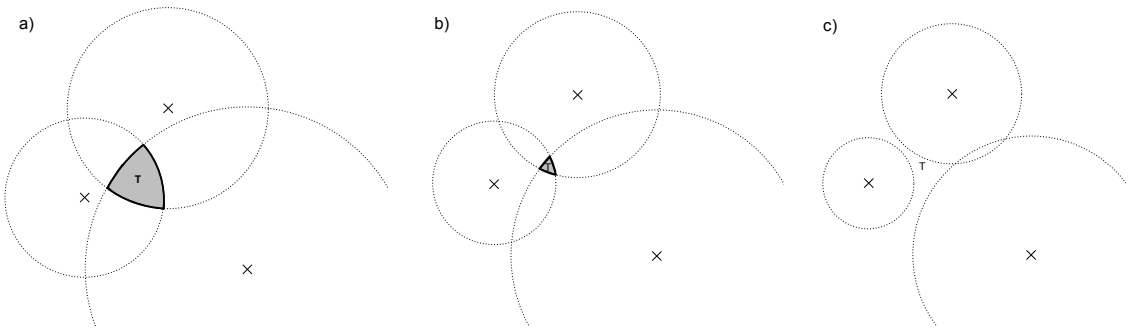
Obr. 2.1: Graf zpoždění v závislosti na geografické vzdálenosti zjištěný při kalibraci metody CBG.

Před měřením je provedena kalibrace – každý landmark změří zpoždění k ostatním referenčním bodů a k naměřené hodnotě zpoždění přiřadí geografickou vzdálenost (viz obrázek 2.1). Pro každý landmark je pak je nalezena přímka (tzv. Bestline), která leží pod všemi body grafu a zároveň k nim má nejblíže – tím reprezentuje největší poměr zpoždění a vzdálenosti zjištěné kalibrací [5].

Samotná lokalizace pak probíhá tak, že každý referenční bod změří zpoždění k cílové stanici. Toto zpoždění následně referenční bod přepočítá pomocí vlastní Bestline přímky na vzdálenost, která se rovná poloměru kruhu, ve kterém se cílová stanice nachází. Cílová pozice stanice je pak určena průnikem kruhů jednotlivých referenčních bodů a nalezením těžiště této oblasti průniku. Velikost průniku určuje také chybovou oblast, ve které se cílová stanice může nacházet. Na obrázku 2.2 b) je průnik kruhů – oblast, kde se nachází cílová stanice.

2.2.4 Metoda Speed of Internet

Metoda Speed of Internet (SOI) [8] je založena na podobném principu jako CBG – vytvoření hranice nejzazší vzdálenosti, kde se cíl může nacházet. Ke své činnosti tedy také potřebuje množinu aktivních referenčních bodů se známou polohou. Rozdíl oproti CBG je v přepočítání zpoždění na vzdálenost, kdy je použita konstanta $\frac{4}{9}$c



Obr. 2.2: Zobrazení principu geolokačních metod založených na vytváření hranic (kruhů) okolo referenčních bodů (křížky). Průnik kruhů definuje oblast, ve které se cíl nachází. Na obrázku a) je nadhodnocení velikosti kruhů, zde je cíl bezpečně uvnitř průniku, b) zobrazuje minimalizování velikostí kruhů, stále však bezpečné pro lokalizaci a na c) je chyba při lokalizaci, kdy některé hranice (kruhy) byly podhodnoceny a průnik všech kruhů není možný stejně jako lokalizace cíle (T).

(přímka Baseline na obrázku 2.1) namísto přímky vypočítané z kalibračních dat [8]. SOI tedy nepotřebuje kalibrační měření, čímž je zmenšena zátěž sítě. Nevýhodou je poté menší přesnost, větší oblast průniku – obr. 2.2 a) a také možnost, že se kruhy neprotnou. To může nastat při podhodnocení vzdáleností, podobně jako na obrázku 2.2 c), kde neexistuje průnik oblastí a není tedy možné určit pozici cíle [5].

2.2.5 Metoda Octant

Metoda Octant [19] principiálně vychází z metody CBG, oproti ní však může referenční bod zjistit nejen oblast, kde se cílová stanice nachází, ale navíc i oblast, kde se cílová stanice nemůže nacházet. Tyto oblasti se označují jako pozitivní a negativní vzdálenosti. Negativní vzdálenost udává hranice, za kterými se stanice nemůže nacházet – jedná se o kruhovou oblast blízko referenčnímu bodu. Spojením s pozitivní informací (známou z CBG) vznikne mezikruží, ve kterém se cílová stanice může nacházet. Cílová poloha je pak určena jako průnik těchto mezikruží, čímž může vzniknout i nekonvexní oblast, která je pak popsána Beziérovými křivkami.

Pro přepočet zpoždění na vzdálenost se obdobně jako u CBG používá kalibračních dat mezi referenčními body. K přepočtu je však využita konvexní obálka všech změřených dat, kde horní část konvexní obálky popisuje přepočet obousměrného zpoždění na pozitivní mezní hranice (obdobně jako u metody CBG). Dolní část konvexní obálky definuje negativní mezní hranice, kde se cílová stanice nenachází [19].

Octant umožňuje také jako negativní informaci použít obydlenost území a vyřadit

tak z výsledku moře a další nepravděpodobné oblasti. Dále Octant zjištěnou polohu zpřesňuje použitím zpětného převodu DNS a hledáním polohy mezilehlých prvků (směrovačů).

2.2.6 Metoda Spotter

Metoda Spotter [10] také vychází z metody CBG a využívá kalibrace mezi jednotlivými referenčními body. Kalibrační data jsou poté podrobena statistické analýze a na tomto základě je vytvořeno normální (Gaussovo) rozdělení pravděpodobnosti vzdáleností. Následně je změřeno zpoždění od referenčního bodu k lokalizované stanici a dle toho je vytvořena kružnice (se středem v referenčním bodě). Na jejím blízkém okolí je definována hustota pravděpodobnosti dle zkalibrovaného Gaussova rozdělení. Průnik hustot pravděpodobnosti všech referenčních bodů vytvoří v místě průniku region s vysokou pravděpodobností (součet všech pravděpodobností) výskytu cílové stanice.

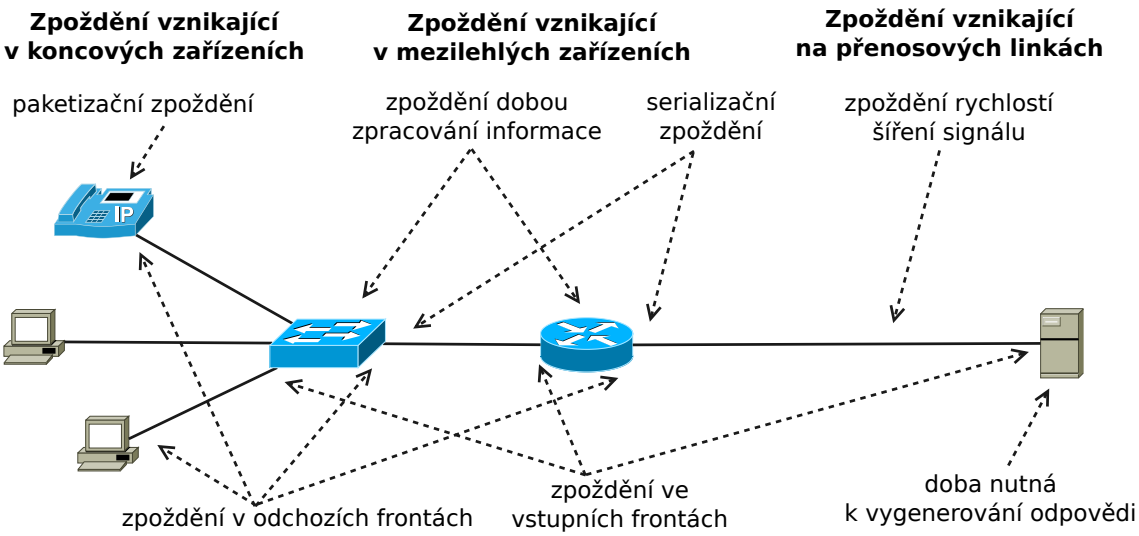
2.2.7 Srovnání aktivních geolokačních metod

Aktivní geolokační metody uvedené výše jsou referovány dále, ale tento výčet není kompletní. Srovnání jednotlivých metod je možné najít u autorů metody novější, jejíž výsledky jsou porovnávány s metodami staršími – např. [19, 3].

2.3 Parametry komunikačního řetězce ovlivňující IP geolokaci

Zpoždění vznikající na jednotlivých částech komunikačního řetězce (obrázek 2.3) má různý charakter, velikost i vliv na celkové zpoždění [23].

V publikaci [2] autoři rozdělují celkové zpoždění na jeho *deterministickou* část a *stochastickou* část. Deterministické zpoždění má konstantní velikost, kterou lze vypočítat (jedná se o minimální čas potřebný pro přenos zprávy). Hodnota celkového zpoždění nemůže být nikdy menší než velikost deterministické části zpoždění. Oproti tomu stochastické zpoždění má náhodný charakter a je ovlivněné aktuálním stavem sítě (dobou zpracování v odchozích frontách, zatížením síťových prvků, rušením na přenosovém médiu – nutnost retransmise). Toto zpoždění může mít i nulovou hodnotu a nemusí celkové zpoždění ovlivňovat, naopak v případě velkého zatížení sítě může velikost tohoto zpoždění převyšovat deterministickou část [2].



Obr. 2.3: Zdroje zpoždění a místo jejich vzniku.

2.3.1 Zpoždění vznikající v koncových zařízeních

Koncovými zařízeními jsou zdroj a cíl přenášené informace, které předpřipraví zprávu pro přenos, označí ji adresou svojí i příjemce a zapouzdří ji do paketů, následně do rámců nebo buňek a poté jej vyšle bit po bitu na přenosovou linku. Přijímací stanice k získání přenášené informace musí provést opačný postup. Na celkové zpoždění koncového zařízení t_{KZ} má vliv několik dílčích zpoždění, která jsou rozebrána v následujících podkapitolách.

Paketizační zpoždění

Paketizační zpoždění (t_{pak}) je doba, potřebná pro vytvoření paketu – přenosové jednotky v IP sítích. Závisí na rychlosti zpracování informace počítačem. Pokud se však jedná o interaktivní komunikaci, závisí toto zpoždění též na velikosti paketu, objemu informace a vzorkovací frekvenci. Při depaketizaci je toto zpoždění vyrovnáno pozdržením ostatních vzorků paketu tak, aby byly stejně zpožděny jako první vzorek.

Doba nutná k vygenerování odpovědi

Při analýze obousměrného zpoždění t_{rtt} (RTT – Round Trip Time) patří mezi složky zpoždění i doba potřebná k vygenerování odpovědi cílové stanice (t_{odp}). Velikost tohoto zpoždění je závislá na koncové stanici, operačním systému a jejím vytížení. Tím pádem ji není možné snadno definovat a k jejímu zjištění je potřebné provést měření na skutečných zařízeních.

Laki et al. [9] provedl měření doby nutné k vygenerování odpovědi stanic v síti GEANT2 pomocí nástrojů `ping` a `traceroute`. Doba nutná k vygenerování odpovědi byla ve většině případů mezi 300 μs a 1000 μs, přičemž nejvíce zastoupena byla hodnota okolo 500 μs. V této hodnotě jsou zahrnuty i paketizační, serializační a deserializační zpoždění obou koncových zařízení t_{KZ2}.

2.3.2 Zpoždění vznikající na přenosových linkách

Přenosové linky ke stanicím nejsou vedeny nejkratší cestou, ale kabely jsou pokládány na vhodných místech (například podél silnic, železnic a spolu s dálkovým vedením elektrické energie) [2]. Další prodloužení cesty paketu může způsobit směrování. Tímto narůstá délka trasy a roste i zpoždění způsobené dobou šíření signálu. Kromě fyzické vzdálenosti stanic je důležitá také technologie přenosu signálu, která ovlivňuje především přenosovou rychlost, a tedy i serializační zpoždění. Doba zpoždění přenosově linky (t_{PL}), zahrnuje zpoždění způsobené dobou šíření signálu. Serializační zpoždění ovlivněné rychlostí linky je uvedeno v rámci zpoždění mezilehlých zařízení.

Zpoždění vzniklé rychlostí šíření signálu

Zpoždění vzniklé rychlostí šíření signálu v přenosovém médiu t_{rs} (v jiné literatuře nazývané také propagační zpoždění) je přímo závislé na délce přenosové linky l a rychlosti šíření signálu v_{rs} v ní, dle rovnice

$$t_{rs} = \frac{l}{v_{\mathrm{rs}}} \text{ . [9]} \tag{2.1}$$

Elektromagnetický signál se šíří médiem rychlostí, která je odvozená od rychlosti světla ve vakuu, respektive c je maximální možná rychlost šíření. Protože je nejrozšířenějším médiem pro transportní sítě optické vlákno, je vhodné na dlouhých trasách počítat s rychlostí šíření signálu přibližně 0,65c [16]. Pro dlouhé vzdálenosti (nad 1000 km) má toto zpoždění hlavní podíl na celkovém zpoždění [24].

2.3.3 Zpoždění vznikající v mezilehlých zařízeních

Mezilehlé zařízení je jakýkoliv aktivní prvek na cestě mezi koncovými zařízeními. Zpoždění v mezilehlých zařízení t_{MZ} se skládá ze serializačního a deserializačního zpoždění, doby potřebné pro přesunutí datové jednotky ze vstupu na výstup, a doby čekání ve výstupních frontách.

Vzhledem k složitosti všech složek zpoždění vznikajících v mezilehlých zařízeních (více viz následující podkapitoly) autoři publikace [9] využívají, jako aproximaci zpoždění jednoho L3 prvku, vlastním měřením zjištěnou hodnotu $t_{\mathrm{MZ}} = 100$ μs.

Serializační a deserializační zpoždění

Serializační zpoždění (t_{ser}) je doba od vyslání prvního bitu do odeslání posledního bitu rámce na linku. Hodnotu tohoto zpoždění ovlivňuje velikost rámce a přenosová rychlost linky. Serializační zpoždění vzniká za každým aktivním prvkem, ve kterém dojde k uložení rámce do paměti.

Deserializační zpoždění (zpoždění ve vstupních frontách) t_{des} je doba, kterou rámec stráví ve vstupní vyrovnávací paměti (buffer), než jsou načteny všechny potřebné bity. Velikost deserializačního zpoždění rovna serializačnímu zpoždění t_{ser} příchozí linky vztažené na počet bitů ukládaných do vyrovnávací paměti.

Doba zpracovávání informace

Je to čas t_{z}, který zařízení potřebuje pro přesunutí paketu ze vstupu do výstupní fronty. Minimální hodnota závisí na výkonu zařízení a prováděných operacích. Celková doba zpracování se zvyšuje se zvyšujícím se zatížením prvku. Moderní síťové prvky mají většinu funkcionalit podporovaných hardwarovým čipem, a tak při standardním provozu bude tato doba minimální. Každopádně je nemožné stanovit deterministicky tuto dobu pro různé typy zařízení a jejich funkcionality.

Zpoždění v odchozích frontách

Hodnotu zpoždění v odchozích frontách (t_{f}) není možné deterministicky určit, protože je závislá na aktuálním zatížení prvku. Zpoždění v odchozích frontách se uplatňuje nejen směrovače a přepínače, ale i pro koncové stanice. Velikost tohoto zpoždění je proměnná a dá se aktuálně zjistit podle celkové velikosti paketů ve frontě a rychlosti odchozí linky.

2.3.4 Zpoždění celého komunikačního řetězce

Jak je zobrazeno na obrázku 2.3, zdrojem zpoždění v síti je každá část sítě. Celkové zpoždění je dáno součtem zpoždění za jednotlivé části sítě. V závislosti na počtu N mezilehlých zařízeních (počet linek $N + 1$) je možné vyjádřit obousměrné zpoždění t_{rtt} (RTT) jako

$$t_{\text{rtt}} = t_{\text{pak,zdroj}} + t_{\text{f,zdroj}} + t_{\text{ser,zdroj}} + t_{\text{des,zdroj}} + 2 \cdot \sum_{i=0}^{N} t_{\text{rs},i} + \\ + 2 \cdot \sum_{i=0}^{N-1} (t_{\text{des},i} + t_{\text{z},i} + t_{\text{f},i} + t_{\text{ser},i}) + t_{\text{des,cil}} + t_{\text{odp,cil}} + t_{\text{pak,cil}} + t_{\text{ser,cil}} + t_{\text{f,cil}} \,. \tag{2.2}$$

3 CÍLE DIZERTAČNÍ PRÁCE

V dnešní době je využití IP geolokace na denním pořádku, vzhledem k velkému počtu webových stránek, které ji používají k přizpůsobení stránky návštěvníkovi nebo naopak k omezení služeb z některých zemí. Pro správce sítě a experty z oblasti kyberkriminality je důležité také zjistit, odkud síťový provoz pochází (na základě zdrojové IP adresy). K tomu všemu jsou v převážné míře využívány IP geolokační databáze. Jejich přesnost je velmi dobrá – s jistotou určí zemi a při určení místa (města, popř. regionu) jsou přesné v cca 75 % případů [11]. Problematické však jsou v těchto případech záznamy o IP adresách, které nejsou pravdivé. Tyto vznikají především z důvodu změny IP adresy zařízení, při přesunu zařízení či při vložení chybného záznamu do databáze [15]. Z tohoto ohledu je důležité najít způsob, jak tento chybný záznam s jistotou rozpoznat a případně předat informace, které pomůžou vlastníkům databází záznam opravit.

Dizertační práce se zabývá aktivními IP geolokačními metodami a jejich využitím pro ověření záznamů z IP geolokačních databází. Aktivní metody jsou založeny na měření síťových parametrů z referenčních bodů zvaných landmarky. Při využití fyzikálních zákonů (např. rychlost šíření signálu v médiu) a statistických údajů komunikace v Internetu je možné stanovit vzdálenost od referenčního uzlu (landmarku), za kterou již lokalizovaná stanice nemůže ležet. Pokud vytvoříme průnik takto vzniklých oblastí okolo jednotlivých landmarků získáme oblast, ve které se s jistotou cílová stanice nachází. Když pro tuto stanici vezmeme zeměpisné souřadnice z geolokační databáze, můžeme snadno ověřit, zda tyto souřadnice leží uvnitř vymezené oblasti a rozhodnout, jestli je informace o poloze z geolokační databáze důvěryhodná.

Hlavním cílem dizertační práce je navrhnout geolokační metodu, která bude s jistotou definovat oblast, ve které se hledaná stanice nachází a následně tuto metodu aplikovat na záznamy z vybraných geolokačních databází a nalézt tak záznamy, které nejsou důvěryhodné. Tato metoda bude založena na fyzických vlastnostech síťových prvků – jak aktivních (router, switch, ...), tak pasivních (přenosová média). Funkčnost metody bude ověřena na vybraném vzorku testovacích adres a její výsledky budou porovnány s obdobnými metodami. Následně bude vytvořen dataset IP adres, pro které budou získány záznamy z geolokačních databází k ověření vytvořenou metodou. Výstupem bude především statistika počtu adres mimo oblast zaručené polohy, druhotně pak přesnost záznamů vzhledem k výsledkům aktivního měření.

4 ANALÝZA PARAMETRŮ OVLIVŇUJÍCÍCH ODHAD VZDÁLENOSTI

Tato kapitola se věnuje analýze síťových parametrů mezi stanicemi se známou polohou za účelem zjištění typických hodnot parametrů, které ovlivňují přepočet zpoždění na vzdálenost pro účely IP geolokace. Většina aktivních geolokačních metod uvedených v kapitole 2.2 potřebuje pro svoji činnost odvodit vzdálenost mezi landmarkem (stanicí se známou polohou) a lokalizovanou stanicí. K tomu je využito měření zpoždění při přenosu informace. K odvození velikosti jednosměrného zpoždění je nutné od obousměrného zpoždění odpočítat čas nutný k vygenerování odpovědi cílovou stanicí a také čas zpracování zdrojovou stanicí (odeslání a příjem), které je podrobněji popsané v kapitole 2.3.1. Pro výpočty v této kapitole vyjdeme ze závěrů autorů publikace [9], kteří využívají nejnižší jimi zjištěnou dobu nutnou k vygenerování odpovědi ($t_{\text{odp}} = 300\ \mu\text{s}$). Díky tomu je možné naměřené hodnoty obousměrného zpoždění dále analyzovat jako jednotlivé složky jednosměrného zpoždění tak, jak je uvedeno v kapitole 2.3.

4.1 Měření v rozlehlé síti se známou topologií

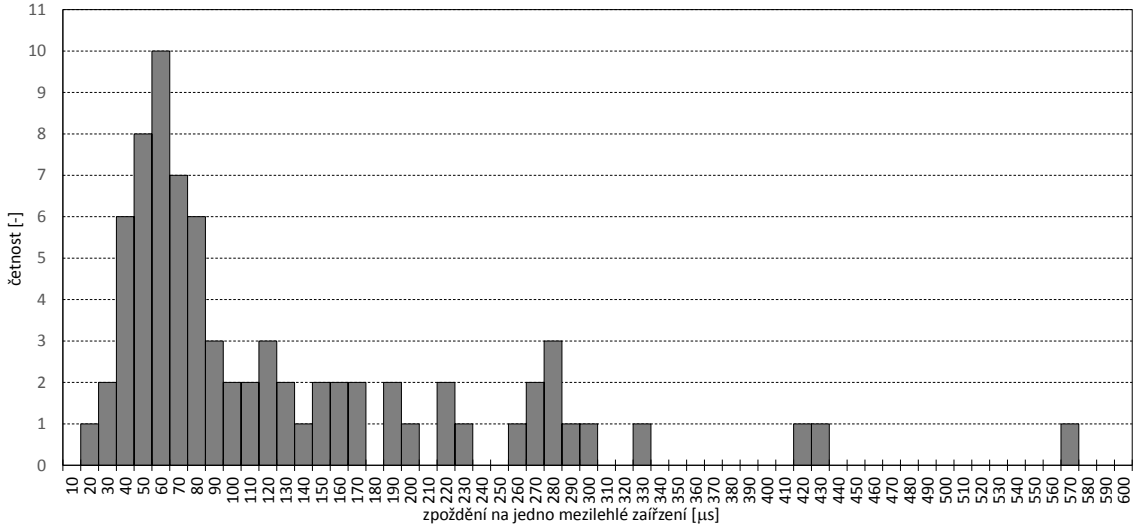
Cílem měření v síti CESNET2 bylo srovnat naměřené zpoždění s předpokládanou hodnotou přenosového zpoždění a odvodit minimální, průměrnou a střední hodnotu zpoždění pro jedno zařízení pracující na síťové vrstvě (L3) referenčního modelu ISO/OSI. Celá síť CESNET2 je bohatě dokumentovaná a je možné zjistit mnoho informací o topologii sítě i parametrech jednotlivých linek. Pro další použití byla v dokumentaci nalezena skutečná délka optických kabelů.

4.1.1 Zpoždění na jedno mezilehlé zařízení

Pro výpočet zpoždění je použito známé délky optického kabelu (přenosové cesty), počtu mezilehlých zařízení a změřeného obousměrného zpoždění. Počet zařízení pracujících na třetí vrstvě byl zjištěn z velikosti TTL příchozího paketu. Vzhledem k tomu, že obvyklé hodnoty TTL (255, 128, 64 a 32) používané operačními systémy jsou známé [13], dopočet není složitý. Zjištěný počet mezilehlých zařízení byl také ověřen programem `traceroute`.

Histogram na obrázku 4.1 zobrazuje statistické rozložení spočítaných hodnot zpoždění na jeden L3 prvek (zařízení pracující na třetí vrstvě referenčního modelu ISO/OSI). Dle histogramu můžeme prohlásit, že pro 75 % hodnot je zpoždění nižší než $160\,\mu$s, což se rovná přenosovému zpoždění na vzdálenost 31 km. Minimální vypočítané zpoždění bylo $15\,\mu$s a hodnota dolního kvartilu byla $53\,\mu$s. Průměrná

hodnota vypočítaných zpoždění na jeden prvek je 124 μs, medián je roven 77 μs [22]. Tyto hodnoty jsou srovnatelné s výsledky z článku [9], kde autoři uvádějí průměrnou hodnotu zpoždění směrovače 101 μs.



Obr. 4.1: Histogram zpoždění pro jedno mezilehlé zařízení zjištěné měřením a výpočtem, každý sloupec je široký 0,01 ms [22].

4.1.2 Srovnání délky kabelu s přímou vzdáleností

Aktivní geolokační techniky určí polohu cílové stanice jako průnik oblastí zjištěných jednotlivými referenčními body. Poloměr oblasti je zjištěn přepočtem zpoždění na geografickou vzdálenost, která je ovlivňována nejen přenosovým zpožděním, ale i skutečnou délkou, která se může od přímé vzdálenosti výrazně lišit. Proto byl vyjádřen poměr skutečné délky kabelů ku přímé vzdálenosti. Pokud výsledky vyjádříme statisticky, je poměr v průměru 1,99, medián 1,92, minimum 1,38 a maximum 4,64. Srovnatelné výsledky uvádí Laki et al. [9], kde je uveden průměrný poměr 2,17 mezi přímou vzdáleností a skutečnou délkou kabelů.

5 VARIABILITA ZPOŽDĚNÍ V PRŮBĚHU ČASU

Tato kapitola se zabývá možností využití kalibračních dat z předešlé kalibrace za předpokladu zachování stejné přesnosti geolokace a také tím, nakolik neprovedení kalibrace degraduje výsledky. Měření v této kapitole probíhalo v experimentální síti PlanetLab, konkrétně na serverech PlanetLabu umístěných v Evropě. Měření bylo opakovaně spouštěno po dobu necelých 4 měsíců (15 týdnů) tak, aby byla posbírána veškerá data ke každé celé hodině. Při měření mezi každou dvojicí stanic bylo odesláno 30 dotazů v 0,5 sekundových intervalech, z nichž byla vybrána minimální hodnota. Zároveň dle velikosti TTL u odpovědi byl určen počet mezilehlých zařízení.

5.1 Sledování změny zpoždění v Internetu v průběhu času

Při analýze sesbíraných dat bylo v prvé řadě přistoupeno ke sledování změny zpoždění mezi stejnými uzly v Internetu v průběhu časového úseku. Převážná většina aktivních geolokačních metod provádí přepočet RTT (obousměrného zpoždění) na vzdálenost za pomoci kalibrační funkce sestavené z párů hodnot RTT a geografické vzdálenosti. Vzhledem k tomu, že geografická vzdálenost mezi místy je konstantní, je potřebné zjistit, nakolik změna zpoždění vyvolá změnu kalibrační funkce.

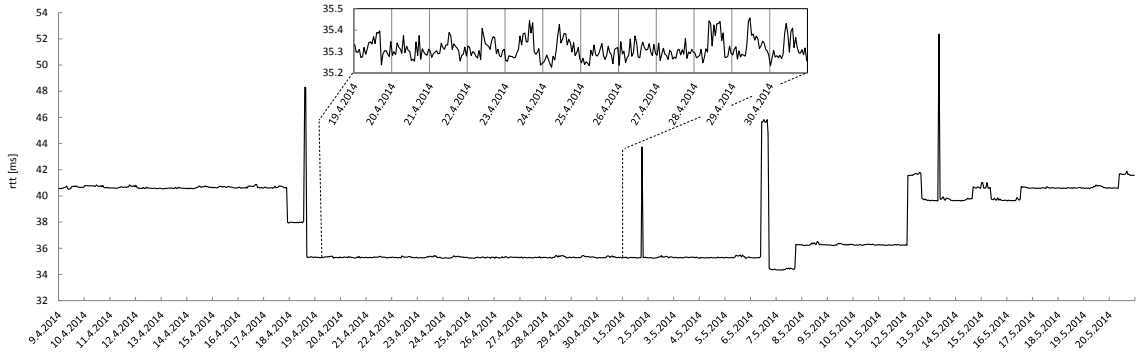
5.1.1 Změna zpoždění v průběhu času

Vzhledem k přesnosti ±0,1 ms nástroje `ping` zmiňované v [1], byla považována změna velikosti zpoždění o 0,2 ms za nepřesnost měření a v dalších úvahách je jako změna zpoždění považovaná změna větší než 0,2 ms. Takovéto změny byly identifikovány u 195 tisíc po sobě jdoucích vzorků. Celkem 125 tisíc měření (2 % ze všech) vykazovalo změnu zpoždění, která trvala alespoň dvě hodiny. U zbytku změn (1 % ze všech měřených případů) se jednalo pouze o krátkodobé změny vyskytující se právě v jednom vzorku.

Po dobu šesti týdnů došlo k trvalejší změně (stejná hodnota po alespoň 2 hodiny) u každého páru v průměru 15krát (medián 9). Dočasné změny, kdy hodnota RTT měla při dalším měření hodnotu jinou, proběhly v průměru 7,5krát u každého páru (medián byl 2).

K ilustraci typického průběhu zpoždění mezi místy bylo vybráno měření mezi městy Modena v Itálii (`planetlab-2.ing.unimo.it`) a Madrid ve Španělsku

(`utet.ii.uam.es`). Průběh je zobrazen na obrázku 5.1, kde je zobrazen vývoj velikosti obousměrného zpoždění po dobu šesti týdnů (měřeno každou hodinu), přičemž graf začíná v 0:00 středoevropského času. V části grafu přibližující téměř konstantní průběh lze pozorovat výchylku pohybující se maximálně ±0,1 ms a tato může být způsobena chybou nástroje, popř. zvýšeným zatížením v průběhu pracovního dne.



Obr. 5.1: Minimální změřené RTT mezi uzly `planetlab-2.ing.unimo.it` (Modena, Itálie) a `utet.ii.uam.es` (Madrid, Španělsko) v průběhu času.

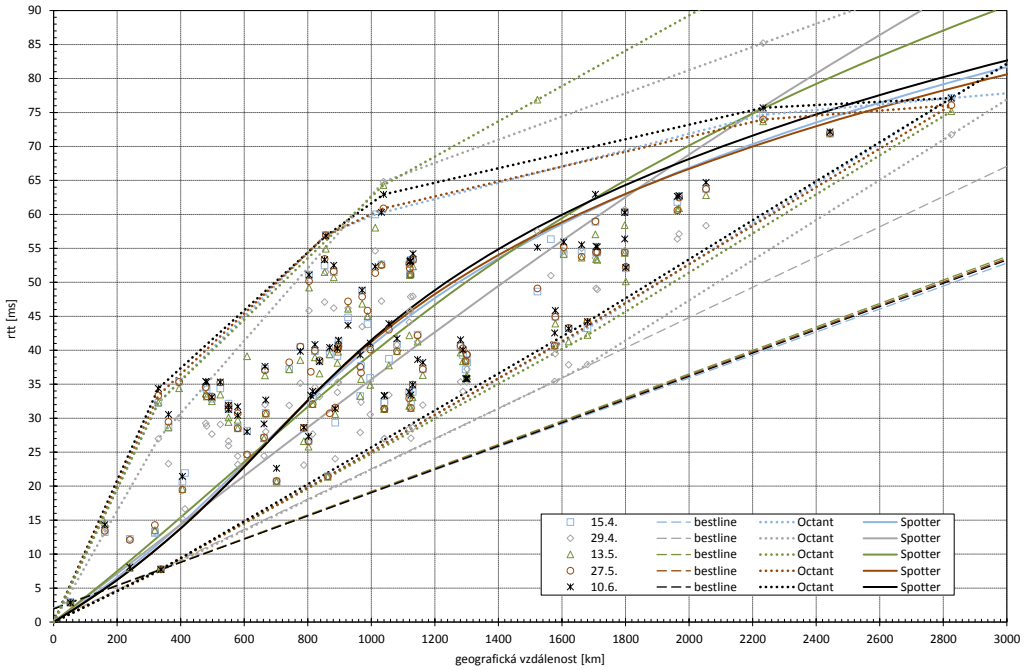
5.1.2 Vliv změn obousměrného zpoždění na kalibrační funkci

Kalibrační funkce IP geolokačních metod založených na hranicích (např. CBG a Octant) berou v úvahu především maximální, případně minimální poměr mezi obousměrným zpožděním a vzdáleností. Na obrázku 5.2 je zobrazený kalibrační graf pro referenční uzel `planetlab2.ing.unimo.it` umístěný v Modeně (Itálie). Většina hodnot RTT kolísá okolo stejné úrovně a tím pádem vliv na kalibrační funkce IP geolokačních metod je malý, což dokládá i změna přímky Bestline popř. konvexní obálky u metody Octant. Výjimkou jsou data z 29. 4., kdy došlo k větší změně RTT (pro mnoho párů je nižší) a tím i převodní funkce vypočítané z tohoto měření jsou více odlišné. Vliv těchto odlišností na přesnost geolokace je podrobně ukázán v dalších kapitolách.

5.2 Přesnost IP geolokace při použití kalibračních dat z předchozích měření

5.2.1 Přesnost IP geolokace při použití den starých kalibračních dat

Tato část se zaměřuje na vliv stáří kalibračních dat v řádech hodin na výslednou polohu určenou IP geolokací. V grafu na obrázku 5.3. jsou zobrazeny kvartily rozložení

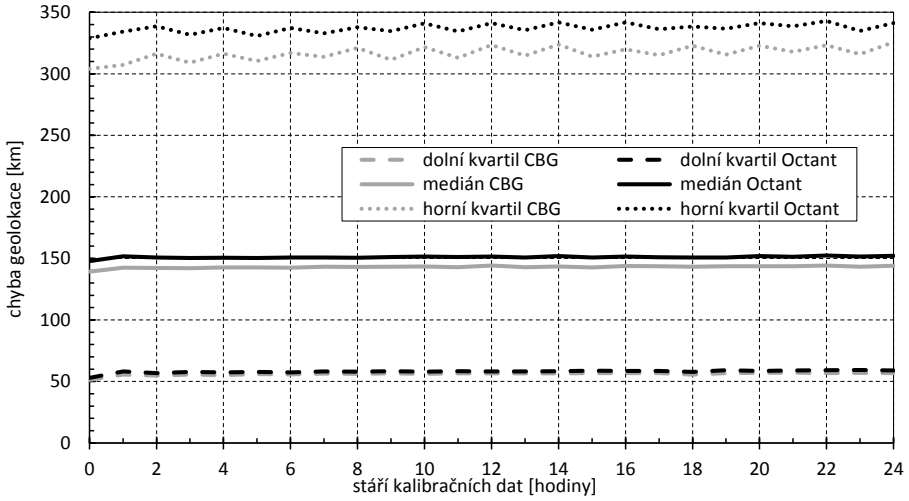


Obr. 5.2: Graf kalibrační funkce pro referenční bod v Modeně, Itálie (`planetlab-2.ing.unimo.it`). Zobrazuje vývoj kalibračních dat v průběhu 6 týdnů – zobrazena jsou data sesbíraná v úterý ve 21:00 každých 14 dní.

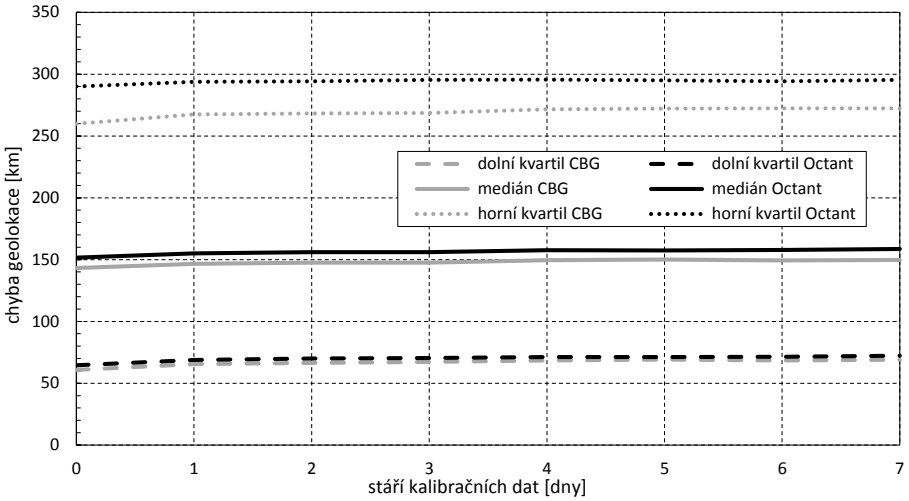
chyby pro metody CBG a Octant. Nejnižší chyba je při použití aktuálních kalibračních dat (hodnota 0 na ose x) a následně mírně narůstá (obdobně jako chyba výpočtu vzdálenosti). Pro dolní kvartil je tento nárůst o 11,4 % pro CBG a o 12,1 % pro Octant, medián narostl o 3,6 % respektive o 3,1 % a horní kvartil o 7,1 % respektive o 4,3 %.

5.2.2 Přesnost IP geolokace při použití týden starých kalibračních dat

Obrázek 5.4 ukazuje obdobný graf jako v předchozí kapitole – průběh kvartilů chyby geolokace pomocí metod CBG a Octant. Obdobně jako u chyby geolokace při použití den starých dat (předchozí kapitola) je i zde znatelný mírný nárůst chyby IP geolokace. Konkrétně relativní nárůst chyby byl pro dolní kvartil o 13,7 % (CBG) a o 12 % (Octant), pro medián o 4,8 % respektive o 4,6 % a pro horní kvartil o 4,8 % a o 1,9 %.



Obr. 5.3: Graf chyby geolokace za použití kalibračních dat získaných v časovém rozmezí hodina až 24 hodin. Zobrazení vývoje kvartilů chyby pro metody CBG a Octant.

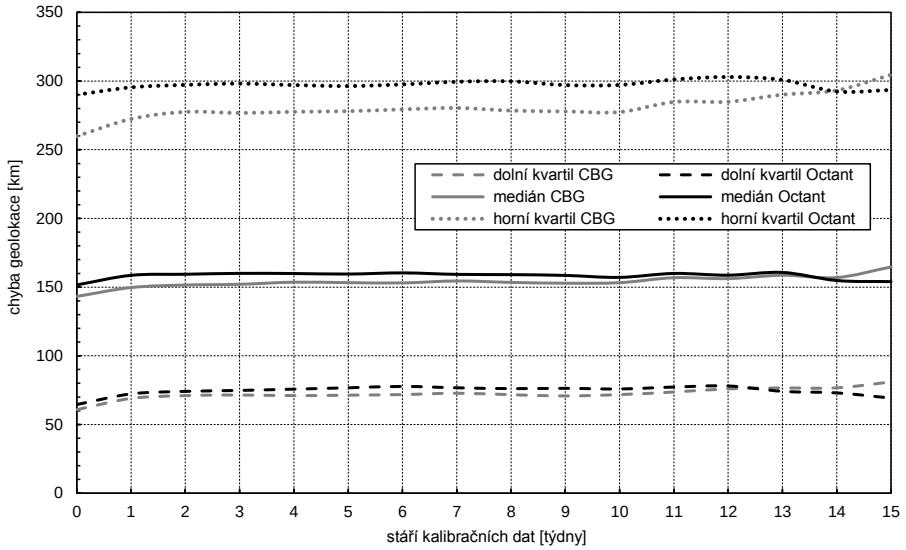


Obr. 5.4: Graf chyby geolokace za použití kalibračních dat den až týden starých. Zobrazení vývoje kvartilů chyby pro metody CBG a Octant.

5.2.3 Přesnost IP geolokace při použití čtvrt roku starých kalibračních dat

Vzhledem k tomu nakolik kolísá přesnost výpočtu vzdálenosti, je důležité ukázat, jaký vliv to má na přesnost samotné geolokace. Vzdálenost zjištěná při použití starých kalibračních dat kolísá obdobným způsobem chyby IP geolokace (viz obrázek 5.5). Dolní kvartil lokalizovaných zařízení má chybu vyšší o 33,2 % (u CBG) a o 20,9 % (pro Octant), u mediánu vzrostla chyba o 15 % respektive o 6 % a horní

kvartil má chybu vyšší o 17,2 % u metody CBG a o 4,5 % u metody Octant.



Obr. 5.5: Graf chyby geolokace za použití čtvrt roku starých kalibračních dat. Zobrazení vývoje kvartilů chyby pro metody CBG a Octant.

5.3 Poznatky užitečné pro návrh nové geolokační metody

Z výsledků prezentovaných v této kapitole je zřejmé, že stáří kalibračních dat má vliv na výpočet vzdálenosti i na přesnost geolokace samotné. Pro návrh nové geolokační metody je důležité také uvést, že charakter zpoždění je po většinu času trvalý. Ke změnám došlo pouze ve 3 % po sobě jdoucích vzorcích změřených dat. U těchto změn byla zaznamenána i změna počtu mezilehlých zařízení a pravděpodobně došlo ke změně směrovací cesty. Z tohoto můžeme potvrdit předpoklad, že změna změřeného zpoždění, je ovlivněna počtem mezilehlých zařízení a délkou přenosových médií. Těchto vlastností bude využito při návrhu metody, která se bude opírat o rychlost šíření signálu přenosovým médiem a dle analýzy zpoždění odpočítávat dobu strávenou v mezilehlých zařízeních. Z těchto důvodů není předpoklad zásadního vlivu kalibrace na novou metodu jako je tomu u metod CBG, Octant a Spotter, u kterých je přepočet zpoždění na vzdálenost přímo ovlivněn zjištěnými daty z kalibrace.

6 GEOLOKACE ZALOŽENÁ NA ANALÝZE PŘENOSOVÉHO ZPOŽDĚNÍ

6.1 Předpoklady pro návrh geolokační techniky

Při návrhu metody budeme na jedné straně vycházet z komplexního popisu zpoždění, který je podrobně uveden v kapitole 2.3. Základním předpokladem pro návrh metody je nemožnost využít měřenou stanici (cíl) k provedení aktivního měření. Z tohoto důvodu jsme při návrhu metody omezeni na obecně podporované mechanismy jako je překlad IP adresy na doménové jméno a odpověď stanice na ICMP dotazy. Dále je možné zjistit počet mezilehlých L3 zařízení pomocí nástroje `traceroute` nebo výpočtem z hodnoty TTL uvedené ve zprávě odpovídajícího cíle.

6.2 Parciální složky zpoždění pro výpočet geografické vzdálenosti

Dle analýzy složek zpoždění provedené v kapitole 2.3 můžeme obousměrné zpoždění t_{rtt} definovat rovnicí 2.2. Zpoždění způsobené rychlostí šíření signálu závisí pouze rychlosti šíření signálu médiem v_{rs} a délce přenosového média l. Pokud budeme na dálkových trasách předpokládat majoritní zastoupení optických vláken, můžeme pro rychlost šíření v médiu použít hodnotu $v_{\mathrm{rs}} = 0,65\mathrm{c}$. Výpočet maximální vzdálenosti l_{celk}, skládající se ze součtu všech délek přenosových médií, je poté možné zapsat po úpravě předcházející rovnice takto

$$l_{\mathrm{celk}} = 2(N+1)v_{\mathrm{rs}}\left(t_{\mathrm{rtt}} - t_{\mathrm{KZ,zdroj}} - 2\sum_{i=0}^{N-1}\left(t_{\mathrm{des},i} + t_{\mathrm{z},i} + t_{\mathrm{f},i} + t_{\mathrm{ser},i}\right) - t_{\mathrm{KZ,cil}}\right). \quad (6.1)$$

Dalším zpožděním, které ovlivňuje výpočet vzdálenosti je doba nutná k předání zprávy ze vstupu zařízení jeho na výstup t_{z}. Tuto dobu lze u některý produktů najít v katalogových listech, ale vzhledem nemožnosti rozpoznat výrobce mezilehlých uzlů, vyjdeme v dalších úvahách z výsledků uvedených v kapitole 4.1.1. Konkrétně pro prvotní návrh metody budeme pracovat se změřeným prvním kvartilem, tj. $t_{\mathrm{MZ}} = 53\ \mu\mathrm{s}$. Tato hodnota vzhledem ke způsobu zjištění zahrnuje nejen dobu nutnou k předání zprávy mezi vstupem a výstupem, ale serializační zpoždění t_{z} i zpoždění v odchozích frontách t_{f}.

Koncová zařízení, kromě parametrů stejných jako pro mezilehlá zařízení ovlivňují výpočet také dobou nutnou k vygenerování zprávy. V dalších úvahách budeme předpokládat všechny tyto hodnoty jako jedno zpoždění vzniklé v koncových zařízeních $t_{\mathrm{KZ,zdroj}} + t_{\mathrm{KZ,cil}} = t_{\mathrm{KZ2}}$, která dle [9] má minimální hodnotu $300\,\mu\mathrm{s}$.

6.3 Vliv nepřímého vedení kabelů na výpočet vzdálenosti

V kapitole 4.1.2 je v síti se známou topologií linek ukázán vliv jejich nepřímého vedení. Konkrétně využijeme minimální zjištěnou hodnotu $\frac{l_{\text{skut}}}{l_{\text{prima}}} = 1,38$, která pro vypočtené hodnoty představuje minimální násobek prodloužení trasy. Tuto hodnotu dále použijeme jako koeficient k_{nv} pro určení přímé vzdálenosti l_{prima} z vypočítané celkové vzdálenosti l_{celk}.

Výpočet přímé geografické vzdálenosti l_{prima}, při znalosti obousměrného zpoždění t_{rtt} a počtu mezilehlých L3 zařízení N, poté můžeme provést pomocí této rovnice

$$l_{\text{prima}} = \frac{2(N+1)v_{\text{rs}}\left(t_{\text{rtt}} - 2Nt_{\text{MZ}} - t_{\text{KZ2}}\right)}{k_{\text{nv}}}\,. \tag{6.2}$$

6.4 Geodetický aparát pro výpočet geografické polohy

Vypočítaná vzdálenost se promítá na povrch Země, jejíž tvar lze v globále aproximovat modely, kterými můžou být koule, elipsoid nebo geoid. S rozvojem GPS se rozšířil popis polohy pomocí zeměpisných souřadnic na referenčním elipsoidu WGS84. K výpočtu vzdáleností na povrchu tohoto rotačního elipsoidu je možné využít tzv. Vincentyho rovnic [18].

Pro účely IP geolokace je nezbytné nalézt hranici mezní vzdálenosti (constraint) od referenčního (měřícího) bodu. Pomocí průniku takto nalezených hranic se vytvoří region, ve kterém se cílová stanice nachází. Zeměpisné souřadnice polohy cíle jsou následně určeny jako těžiště nalezeného regionu.

Souřadnice bodu v definované vzdálenosti od referenčního bodu

Prvním krokem bude nalezení hranice mezní vzdálenosti, kterou pro další použití popíšeme jako polygon spojující body na této hranici. Pro nalezení bodu ve vzdálenosti l po zemském povrchu pod úhlem α_1 od počátečního bodu se souřadnicemi ϕ_1 (zeměpisná šířka) a L_1 (zeměpisná délka) je možné použít přímé Vincentovy formule (Direct Vincenty Formula) [18].

6.4.1 Hranice okolo referenčního bodu

Jestliže aplikujeme rovnice z předchozí kapitoly, za použití různých hodnot azimutů z výchozího bodu, získáme souřadnice bodů na mezní hranici od referenčního bodu. Hustotu těchto bodů je možné volně nastavit pomocí k_α, kterým definujeme počet

bodů na hranici. Získané zeměpisné souřadnice jednotlivých bodů (ϕ_i, L_i) pak tvoří polygon popisující mezní vzdálenost okolo referenčního bodu.

6.4.2 Region průniku všech mezních vzdáleností

Vzhledem k dostupnému výpočetnímu výkonu dnešní doby, bylo přistoupeno k nalezení regionu průniku všech hranic referenčních bodů M pomocí ověření, zda jednotlivé body z polygonů popisujících hranice mezní vzdálenosti leží ve vypočítané vzdálenosti od jednotlivých referenčních uzlů. Pro každou dvojici tvořenou referenčním bodem a hraničním bodem jiného referenčního bodu byla pomocí inverzní Vincentyho formule [18] zjištěna vzdálenost. Pokud tato vzdálenost byla menší nebo rovna než vypočtená vzdálenost pro všechny referenční body (LM), došlo k zahrnutí tohoto bodu do množiny $\boldsymbol{R}$, která tvoří region průniku.

6.4.3 Těžiště a obsah regionu průniku mezních vzdáleností

Výstupem většiny aktivních IP geolokačních metod je vypočtená poloha měřené stanice a pro definici přesnosti plošná výměra vypočteného regionu. Zjištění souřadnic měřeného zařízení probíhá pomocí nalezení těžiště (centroidu).

Pro výpočet centroidu a obsahu polygonu nad geoprostorovými daty je možné využít geodetických knihoven pro různé programovací jazyky. Tyto funkce jsou připravené například v knihovně `org.geotools` pro programovací jazyk *Java*. Tato knihovna obsahuje též funkci, která ověří, zda pozice leží uvnitř vypočítaného regionu. Toho je možné využít při ověřování platnosti geografických souřadnic, které poskytují IP geolokační databáze.

6.5 Kalibrační proces u metody CLBG

Jak bylo ukázáno v kapitole 5 pro zajištění přesnosti geolokačních metod CBG a Octant je nutná kalibrace. Tyto metody jsou však založeny na převodu zpoždění na vzdálenost pomocí aktuálně změřených dat. Metoda CLBG je oproti tomu postavená na fyzických charakteristikách zpoždění (převážně jeho deterministické části) a vliv na výsledný výpočet mají pouze parametry velikosti zpoždění na jedno mezilehlé zařízení t_{MZ}, zpoždění koncových zařízení t_{KZ2} a koeficient nepřímého vedení kabelů k_{nv}. Protože pro zjištění těchto hodnot bylo použito minimálních zjištěných hodnot, není nutné provádět kalibraci těchto hodnot před každým měřením. Zároveň předpokládáme, že vlastnosti, na kterých jsou závislé (vnitřní architektura síťových zařízení, poloha dálkových vedení), se nemění dramaticky v čase.

Kalibraci velikosti zpoždění na jedno mezilehlé zařízení t_{MZ} je možné provést podle postupu uvedeného v kapitole 4.1.1 nebo podle experimentálního měření provedeného v [9]. Pro zjištění hodnoty zpoždění způsobené koncovými zařízeními t_{KZ2} je potřeba postupovat podle [9]. Ke zjištění koeficientu nepřímého vedení kabelů k_{nv} je možné využít postupu uvedeného v kapitole 4.1.2, nebo vyjít z měření uvedeného v [9] a výsledný parametr r_{nv} následně přepočítat. U všech hodnot je důležité použít minimální zjištěné hodnoty, případně nízkého percentilu (např. 1 %) zjištěných hodnot.

6.6 Srovnání výsledků nové metody

V kapitolách 6.2 a 6.3 je popsán postup přepočtu zpoždění na geografickou vzdálenost a je k tomu využito rovnice 6.2. Tato metoda je založena přepočtu zpoždění na vzdálenost dle velikosti propagačního zpoždění [26]. Protože její výsledky jsou závislé na délce přenosových médií (kabelů), byla nazvána Cable Length Based Geolocalization (CLBG) [27]. Následující podkapitoly poskytují srovnání této metody s jinými aktivními geolokačními metodami.

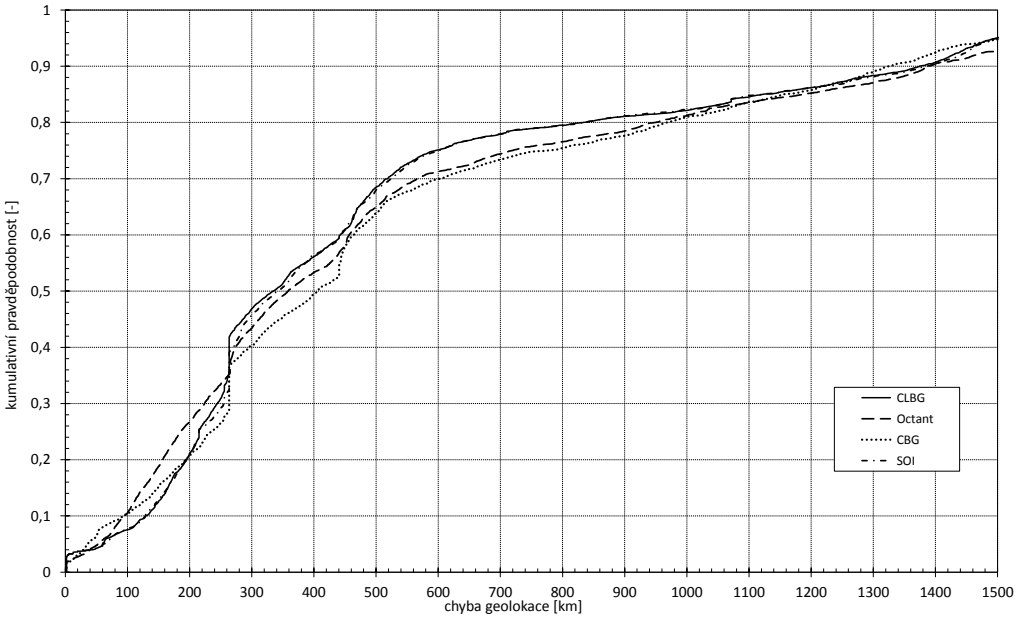
6.6.1 Porovnání výsledků CLBG s metodami GeoPing, ShortestPing a SOI

Výsledky metody CLBG (navržené v rámci doktorského studia) jsou srovnávány s geolokačními metodami GeoPing, ShortestPing a SOI. Nejlépe při srovnání přesnosti vychází metoda CLBG, která má chybu pro dolní kvartil 102 km, medián chyby je 213 km, horní kvartil má chybu 290 km a průměrná chyba má hodnotu 208 km. Při porovnání křivek v grafu můžeme určit pořadí dalších metod SOI, ShortestPing a GeoPing (ve stejném pořadí medián chyby – 219 km, 248 km a 285 km).

6.6.2 Porovnání výsledků CLBG s metodami Octant, CBG a SOI

Graf kumulativní pravděpodobnosti na obrázku 6.1 ukazuje dosahované přesnosti jednotlivých metod. Z výsledků není možné jednoznačně určit metodu s nejnižší chybou, protože pro různé vzorky hodnot se pořadí různí. Pro dolní kvartil hodnot dosahuje nejpřesnějších výsledků metoda Octant (chyba 186 km), stejnou chybu měla metoda CLBG a SOI (215 km), nejvyšší chybu (235 km) měla CBG. Medián chyby byl nejnižší pro metodu CLBG (334 km), následována metodou SOI (343 km), Octant (358 km) a CBG (406 km). V horním kvartilu byla nejpřesnější metoda

CLBG (596 km), těsně za ní byla metoda SOI (601 km), dále Octant (719 km) a CBG (775 km).



Obr. 6.1: Kumulativní funkce pravděpodobnosti chyby geolokace metody CLBG ve srovnání s metodami Octant, CBG a SOI [25].

Přesnost metody CLBG nebyla hlavním kritériem při jejím návrhu. Především bylo důležité navrhnout metodu, která s jistotou určí region, ve kterém se stanice nachází. Zároveň je podmínkou spolehlivé funkce metody to, aby bylo možné vypočítat průnik mezních hranic – minimalizovat možnost podhodnocení hranic. Metoda CLBG a také metoda SOI měly pouze 0,48 % cílových adres, které nebylo možné lokalizovat a to z důvodu dočasné nedostupnosti IP adres při měření. Oproti tomu metoda Octant nebyla schopna lokalizovat 9 % cílových adres, ve většině případů z důvodu podhodnocení hranic. Metoda CBG takovýchto adres měla dokonce 24 %.

Nevýhodou metody CLBG a potažmo i metody SOI je velikost regionu, který zahrnuje možné pozice cíle. Důvodem je, že metoda CLBG byla navržena tak, aby generovala region, ve kterém se na základě analýzy zpoždění cílová stanice vždy nachází. Ve většině případů je tento region velký a často zahrnuje území celého státu, někdy i několika.

7 OVĚŘENÍ DŮVĚRYHODNOSTI ZÁZNAMŮ GEOLOKAČNÍCH DATABÁZÍ

Cílem kapitoly je pomocí vytvořené metody CLBG provést ověření důvěryhodnosti záznamů z geolokačních databází. Jak je blíže popsáno v kapitole 2.1.3, jedná se o statické záznamy o poloze IP adresy, které v některých případech obsahují chybné záznamy s výraznou odchylkou od skutečné polohy. V této kapitole bude ověřeno pomocí metody CLBG, zda poloha IP adresy, oznamovaná různými geolokačními databázemi, je reálná. Tedy zda v oblasti, kterou ohlašuje CLBG leží dané souřadnice a pokud neleží je možné databázi navrhnout opravu záznamu, například použitím jiného zdroje informace. Pro srovnání výsledků metody CLBG s jinými, jsou v této kapitole referovány i metody CBG a Octant.

Všechny použité aktivní IP geolokační metody (CBG, Octant i CLBG) využívají měřících stanic se známou polohou (landmarků). Pro tuto část práce byla použita experimentální síť PlanetLab, která obsahuje servery rozmístěné po celém světě a umožňuje přístup k nim za účelem provádění výzkumů. Dataset cílových uzlů byl získán ze stránek projektu iPlane [17] provozovaného Univerzitou ve Washingtonu. V tomto datasetu byly ponechány pouze IP adresy umístěné na evropském kontinentě, aby se z nich následně náhodnou funkcí vybralo 5000, které budou otestovány. Pro tyto adresy byly zjištěny zeměpisné souřadnice jejich polohy pomocí tří geolokačních databází. Konkrétně se jednalo o veřejné databáze *Geobytes* a *GeoLite2* a demoverzi placené databáze *IP2Location*. Jako údaje z další „databáze“ byly použity zeměpisné souřadnice uvedené přímo v datasetu projektu iPlane – dále označované jako *poloha z datasetu*.

7.1 Ověření důvěryhodnosti údajů z geolokačních databází

Hlavním cílem této kapitoly bylo ověřit důvěryhodnost dat poskytnutých geolokačními databázemi. Na základě změřených dat (obousměrné zpoždění a počet mezilehlých uzlů) mezi referenčními uzly (landmarky) a adresami z datasetu, byla pomocí výše popsaných metod vypočítána předpokládaná poloha cíle a především region, kde se cíl nachází. Následně bylo vyhodnoceno, zda se souřadnice získané z jednotlivých geolokačních databází nacházejí ve vytyčené oblasti. Výpočet oblasti, ve které se cíl nachází, v některých případech selhal kvůli podhodnocení vzdáleností – ty neutvořily průnik. Pro metodu CLBG k tomuto došlu v 0,5 % případů (cíl neodpověděl na ICMP zprávu), u metody CBG u 24 % a u metody Octant u 9 % [25].

Tab. 7.1: Procentuální vyjádření množství IP adres ležících v regionu, vypočítaném pomocí metod CBG, Octant a CLBG [25].

	CLBG	CBG	Octant
poloha z datasetu	91.9 %	46.5 %	44.0 %
Geobytes	68.0 %	39.4 %	36.5 %
GeoLite2	83.1 %	48.7 %	45.0 %
IP2Location	91.7 %	45.2 %	45.1 %

V tabulce 7.1 jsou uvedeny procentuální počty záznamů z geolokačních databází, jejichž poloha byla uvnitř vypočítané oblasti. Při srovnání výsledků jednotlivých metod můžeme pozorovat, že metoda CLBG má nejvyšší procento záznamů z databází ve vypočítaném regionu. Toto je dáno tím, že tato metoda vytváří větší region, který však udává hranice, kde se dle fyzikálních zákonů stanice může nacházet. Metody CBG a Octant jsou svou podstatou vytvořeny pro co nejpřesnější geolokaci a tak je jejich region menší. Z tohoto důvodu je i více záznamů z datasetů, které v jejich regionech neleží. Nevyplývá z toho ale jistota, že poloha získaná z databáze je špatná. Proto v dalších srovnáních budeme upřednostňovat výsledky získané metodou CLBG.

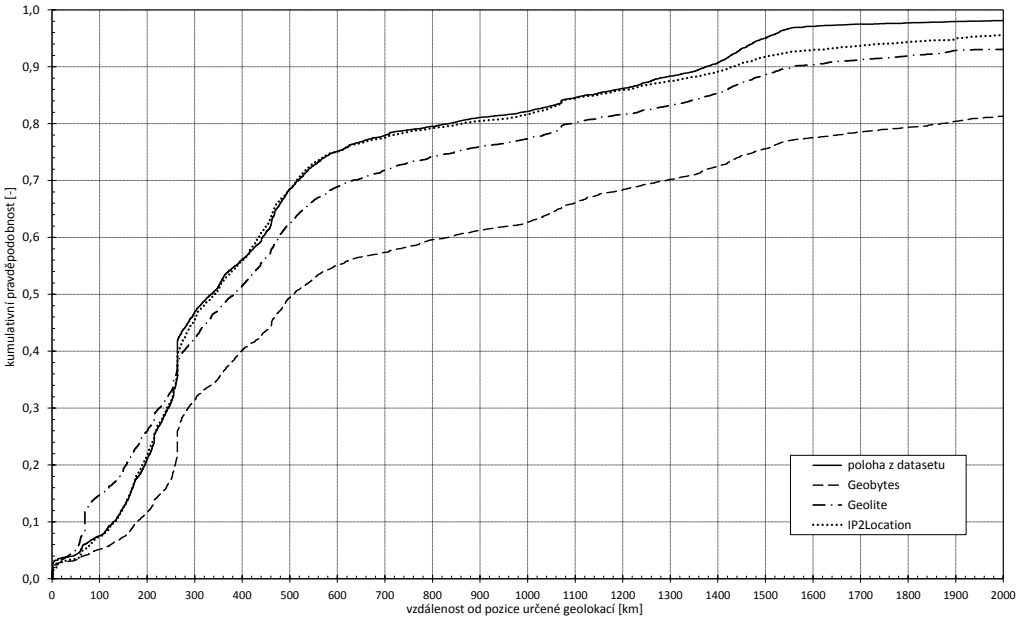
Při porovnání výsledků jednotlivých databází v tabulce 7.1 můžeme dojít k závěru, že údaje z databáze *Geobytes* jsou nejméně důvěryhodné, neboť až 32 % záznamů leží mimo vypočtenou oblast. Databáze *GeoLite2*, která je neplacenou verzí komerční databáze *GeoIP2* je důvěryhodná pro více než 83 % záznamů. A nejlépe z uvedeného porovnání vychází demo verze placené databáze *IP2Location*, poloha jejíchž uzlů je v téměř 92 % případů ve vypočteném regionu.

Srovnání databází, provedené v této kapitole, nebylo vytvořeno za účelem diskreditovat některé databáze, ale především pro ověření funkčnosti navržené metody. Celkově je však možné na základě těchto výsledků uvést, že obecně mají geolokační databáze při ohlašování přesné polohy chybovost – v našem případě minimálně 8 %. Cílem práce bylo nabídnout nástroj pro vylepšení těchto databází odstraněním chybných záznamů a jejich nahrazením informacemi z jiných zdrojů.

7.2 Srovnání přesnosti geolokačních databází

Přestože v předchozí kapitole je uvedeno, že cílem práce není hodnocení jednotlivých geolokačních databází, je možné díky změřeným datům porovnat i přesnost jednotlivých databází. Graf kumulativní pravděpodobnosti odchylky pozice vypočítané

aktivní metodou CLBG od skutečné polohy je na obrázku 7.1. Z grafu a výsledků je patrné, že pouze 3 % vypočítaných pozic metodou CLBG se shoduje s pozicí udávanou databázemi (tolerance 5 km).



Obr. 7.1: Kumulativní distribuční funkce pravděpodobnosti chyby polohy vypočítané metodou CLBG od polohy získané z geolokačních databází [25].

Obecně je možné z grafu vyvodit, že databáze *Geobytes* dosahuje nejmenší přesnosti – při porovnání polohy s aktivními metodami. U databáze GeoLite2 je možné uvést, že je pro dolní kvartil záznamů nejpřesnější, dále však její přesnost klesá. Toto může být dáno tím, že se jedná o neplacenou verzi databáze, která může obsahovat některé záznamy více přesné a jiné záměrně méně (donucení ke koupi přístupu do databáze). Velmi podobnou přesnost vykazuje demo verze databáze IP2Location a poloha uvedená v originálním datasetu.

8 ZÁVĚR

Hlavním cílem práce bylo vytvořit nástroj pro nalezení chybných záznamů v geolokačních databázích. Chybné záznamy jsou pak na základě zjištění nástroje odstraněny nebo nahrazeny informacemi z jiných zdrojů. Zmiňovaným nástrojem je nová geolokační metoda, která je schopna určit region, ve kterém se stanice musí nacházet.

Prvním předpokladem pro návrh metody byla znalost jednotlivých složek zpoždění a především měření, za účelem nalezení jejich obvyklých hodnot. Výsledkem měření v síti se známou topologií byla hodnota zpoždění na jeden mezilehlý prvek a parametr nepřímého vedení kabelů.

V dalším kroku bylo provedeno dlouhodobé sledování variace zpoždění a vliv této variace na IP geolokaci. Na základě měření v průběhu 4 měsíců bylo zjištěno, že zpoždění má po delší časový úsek (dny až týdny) konstantní úroveň. Změny se vyskytují u přibližně 3 % po sobě jdoucích hodnot, kdy 2 % znamenají skok zpoždění na jinou „stabilní“ úroveň, zbytek jsou jednorázové změny. Na samotnou IP geolokaci mělo stáří kalibračních dat největší vliv u přesných hodnot (dolní kvartil) chyba byla vyšší až o 33 %.

Hlavní pozornost v dizertační práci je věnovaná návrhu nové geolokační metody, která byla dle své závislosti na propagačním zpoždění přenosových linek nazvána Cable Length Based Geolocalisation (CLBG) [27]. Princip metody tkví ve víceparametrovém přepočtu obousměrného zpoždění na přímou geografickou vzdálenost. Přepočet bere v úvahu počet zařízení na třetí vrstvě OSI a jejich minimální zpoždění. Dále také zpoždění vzniklé v koncových zařízeních a parametr nepřímého vedení kabelů ku přímé vzdálenosti. Po výpočtu maximální vzdálenosti od měřícího bodu (ve které se cílová stanice může nacházet) je nutné zjistit region průniku všech takto získaných hranic. K tomu je využito geodetických výpočtů popsaných v kapitole 6.4. Výsledná geografická poloha cílové stanice je určena v těžišti zjištěného regionu.

Metoda CLBG byla následně porovnána s ostatními používanými metodami, přičemž jednodušší metody (ShortestPing, GeoPing a SOI) překonávala v přesnosti. Při srovnání se složitějšími metodami (CBG a Octant) vykazovala měla srovnatelnou přesnost.

Návrh metody CLBG probíhal s ohledem na to, aby byla schopna určit oblast, kde se na základě přenosových charakteristik musí zařízení nacházet. Uplatnění metoda najde v případech, kde je třeba ověřit polohu IP adresy, získané například z geolokační databáze. Toto bylo ukázáno v poslední kapitole, kde na vzorku 5000 cílových adres byla ověřena jejich poloha zjištěná ze tří geolokačních databází.

LITERATURA

[1] Arlos, P. – Fiedler, M.: Accuracy Evaluation of Ping and J-OWAMP. In *Swedish National Computer Networking Workshop*, Luleå, 2006.

[2] Bovy, C. J. – Mertodimedjo, H. T. – Hooghiemstra, G., et al.: Analysis of End-to-end Delay Measurements in Internet. In *Passive and Active Measurement Workshop-PAM*, Vol. 2002, s. 8, URL `http://www.pamconf.net/2002/Analysis_of_End_to_end_Delay_Measurements_in_Internet.pdf`

[3] Eriksson, B. – Barford, P. – Maggs, B. et al.: *Posit: An Adaptive Framework for Lightweight IP Geolocation.* Technická zpráva, BUCS-TR-2011-018, Computer Science Department, Boston University, Červenec 2011. URL `http://www.cs.bu.edu/techreports/pdf/2011-018-posit.pdf`

[4] Geobytes: *Simple Solutions.* [online]. URL `http://geobytes.com` [cit. 3. 5. 2017]

[5] Gueye, B. – Ziviani, A. – Crovella, M. et al.: Constraint-based geolocation of internet hosts. In *IEEE/ACM Trans. Netw.*, ročník 14, č. 6, Prosinec 2006: s. 1219–1232, ISSN 1063-6692. URL `http://dx.doi.org/10.1109/TNET.2006.886332`

[6] Hillmann, P. – Stiemert, L. – Dreo, G. et al.: On the Path to High Precise IP Geolocation: A Self-Optimizing Model. *International Journal of Intelligent Computing Research (IJICR)*, ročník 7, číslo 1, březen 2016, ISSN: 0976-9013.

[7] IP2Location: *IP Address to Identify Geolocation Information*, [online], c2017. URL `http://www.ip2location.com` [cit. 24. 5. 2017]

[8] Katz-Bassett, E. – John, J. P. – Krishnamurthy, A. et al.: Towards IP geolocation using delay and topology measurements. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, IMC '06, New York, NY, USA: ACM, 2006, ISBN: 1-59593-561-4, s. 71–84. URL `http://doi.acm.org/10.1145/1177080.1177090`

[9] Laki, S. – Mátray, P. – Hága, P. et al.: A model based approach for improving router geolocation. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, ročník 54, č. 9, Červen 2010: s. 1490–1501, ISSN 1389-1286.URL `http://dx.doi.org/10.1016/j.comnet.2009.12.004`

[10] Laki, S. – Mátray, P. – Hága, P. et al.: Spotter: A model based active geolocation service. In *Proceedings IEEE INFOCOM, Shanghai, 2011*, s. 3173–3181. URL http://dx.doi.org/10.1109/INFCOM.2011.5935165

[11] Li, H. – He, Y. – Xi, R. et al.: A Complete Evaluation of the Chinese IP Geolocation Databases. In *2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Červen 2015, s. 13–17, doi:10.1109/ICICTA.2015.13.

[12] MaxMind: *IP Geolocation and Online Fraud Prevention*, [online], c2017. URL https://www.maxmind.com [cit. 8. 6. 2017]

[13] Mukaddam, A. – Elhajj, I. H.: Hop count variability. In *2011 International Conference for Internet Technology and Secured Transactions*, Prosinec 2011, s. 240–244.

[14] Padmanabhan, V. N. – Subramanian, L.: An investigation of geographic mapping techniques for internet hosts. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, San Diego, California, USA, ročník 31, č. 4, Srpen 2001: s. 173–185, ISSN 0146-4833. URL http://doi.acm.org/10.1145/964723.383073

[15] Poese, I. – Uhlig, S. – Kaafar, M. A. et al.: IP geolocation databases: unreliable?. In *ACM SIGCOMM Computer Communication Review*, ročník 41, č. 2, Duben 2011: s. 53–56, ISSN 0146-4833, doi:10.1145/1971162.1971171. URL http://doi.acm.org/10.1145/1971162.1971171

[16] Pužmanová, R.: *Moderní komunikační sítě od A do Z*. 2. vydání, Brno, Computer Press, 2006, s. 432 ISBN 978-8-025-11278-6.

[17] University of Washington: *iPlane: Datasets*, University of Washington. URL http://iplane.cs.washington.edu/data/data.html [cit. 11. 8. 2013]

[18] Vincenty, T.: Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. In *Survey review*, ročník 23, č. 176, 1975: s. 88–93.

[19] Wong, B. – Stoyanov, I. – Sirer, E. G.: Octant: a comprehensive framework for the geolocalization of internet hosts. In *Proceedings of the 4th USENIX conference on Networked systems design & implementation*, NSDI'07, Berkeley, CA, USA: USENIX Association, 2007, s. 23–23. URL https://www.usenix.org/conference/nsdi-07/octant-comprehensive-framework-geolocalization-internet-hosts

PUBLIKACE AUTORA

[20] Balej, J.: Improvement of Geolocation Methods. In *Proceedings of the 17th conference Student EEICT 2011*, 2011, s. 16–20. URL http://www.feec.vutbr.cz/EEICT/2011/sbornik/03-Doktorske%20projekty/01-Elektronika%20a%20komunikace/01-xbalej02.pdf

[21] Balej, J.: Srovnaní přesnosti aktivních geolokačních technik, [online], *Access server*, 2012, ISSN 1214-9675. URL http://access.feld.cvut.cz/view.php?cisloclanku=2012070001

[22] Balej, J. – Komínek, O. – Rajnoha, M.: Geographic Distance Estimation for IP Geolocation. In *Proceedings in Electronic International, Interdisciplinary Conference EIIC 2013*, ročník 2, číslo 1, 2013, s. 584-588, ISBN 978-80-554-0762-3.

[23] Balej, J. – Komosný, D.: Zdroje zpoždění při komunikaci v Internetu. *Elektrorevue - Internetový časopis (http://www.elektrorevue.cz)*, ročník 2010/42, s. 1–7, ISSN: 1213-1539. URL http://elektrorevue.cz/cz/clanky/komunikacni-technologie/0/zdroje-zpozdeni-pri-komunikaci-v-internetu/

[24] Balej, J. – Komosný, D. – Kathiravelu, G.: Mapping round-trip time into length suitable for geolocation. In *ICT2011 proceeding*, Únor 2011, s. 63–68, ISBN: 978-80-214-4231-3.

[25] Balej, J. – Komosný, D. – Zach, P.: How can measurement of delay be helpful for geolocation databases? In *Enterprise and Competitive Environment: Conference Proceedings*, 2017, s. 24–32, ISBN 978-80-7509-499-5. URL https://ece.pefka.mendelu.cz/sites/default/files/imce/ECE2017_fin.pdf

[26] Komosný, D. – Balej, J. – Sathu, H. et al.: Impact of Intermediate Device Latency on IP Geolocalisation. In *Proceedings of the 2011 International Conference on Telecommunication Systems, Modeling and Analysis*, Praha, 2011, s. 39–48, ISBN 978-0-9820958-4-3.

[27] Komosný, D. – Balej, J. – Sathu, H. – aj.: Cable length based geolocalisation. *Przeglad Elektrotechniczny*, ročník 88, č. 7 A, Červenec 2012: s. 26–32. (IF=0,244)

Curriculum Vitæ

Jiří Balej

Osobní informace

Datum narození:	28. července, 1986
Místo narození:	Brno
Adresa:	Vlnitá 4, 627 00 Brno
E-mail:	jirka.balej@gmail.com

Vzdělání

2010–2017	Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Technická 3058/10, 616 00 Brno, obor: Teleinformatika, titul: Ph.D. (předpokládané ukončení: 2017)
2008–2010	Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Technická 3058/10, 616 00 Brno, obor: Telekomunikační a informační technika, titul: Ing.
2005–2008	Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Technická 3058/10, 616 00 Brno, obor: Teleinformatika, titul: Bc.
2001–2005	Střední průmyslová škola elektrotechnická v Brně, Kounicova 684/16, 602 00 Brno, obor: Slaboproudá elektrotechnika.

Zahraniční stáže

2017 (únor–červen)	University of Patras (Řecko), Department of Electrical & Computer Engineering.

Odborné certifikace

2016	CCNA Security – Cisco Certified Network Associate Security
2015	MTCWE – MikroTik Certified Wireless Engineer
2014	JNCIA-Junos – Juniper Network Certified Associate JunosOS
2013	MTCTCE – MikroTik Certified Traffic Control Engineer
2013	MTCNA – MikroTik Certified Network Associate
2013	CCNA R&S – Cisco Certified Network Associate Routing & Switching

Zaměstnání

2014–	*Projektant a analytik informačních systémů*, Mendelova univerzita v Brně, Ústav informačních technologií, Zemědělská 1655/1, 613 00 Brno.
2013–2014	*Technicko hospodářský pracovník*, Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, Technická 3082/12, 616 00 Brno.
2012–	*Akademický pracovník – asistent*, Mendelova univerzita v Brně, Provozně ekonomická fakulta, Ústav informatiky, Zemědělská 1655/1, 613 00 Brno.

Publikace

- Publikace v časopisech s impaktním faktorem: 1 + 1 (*odesláno k publikování*)
- Publikace v časopisech indexovaných databází Scopus: 4
- Publikace v jiném odborném periodiku: 2
- Publikace v konferenčních sbornících: 7
- Konferenční příspěvky indexované databází WoS: 2 + 1 (*odeslán k indexaci*)