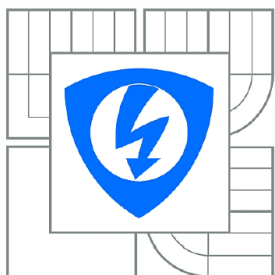


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ  
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ  
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

## SOUČASNÉ MOŽNOSTI NALEZENÍ FYZICKÉ POZICE STANICE V INTERNETU

CURRENT METHODS FOR GEOLOCATION OF INTERNET HOSTS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

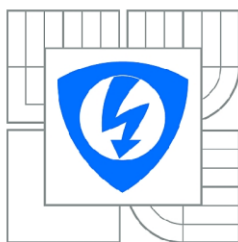
PAVEL DOLEŽEL

VEDOUcí PRÁCE

SUPERVISOR

Ing. JIŘÍ BALEJ

BRNO 2011



**VYSOKÉ UČENÍ  
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky  
a komunikačních technologií**

**Ústav telekomunikací**

# Bakalářská práce

bakalářský studijní obor  
**Teleinformatika**

**Student:** Pavel Doležel  
**Ročník:** 3

**ID:** 106409  
**Akademický rok:** 2010/2011

## NÁZEV TÉMATU:

**Současné možnosti nalezení fyzické pozice stanice v Internetu**

## POKYNY PRO VYPRACOVÁNÍ:

Prostudujte současné způsoby hledání fyzické pozice stanice v síti Internet. Zaměřte se na možnosti zjišťování fyzické pozice stanice na základě měření zpoždění. Dále nastudujte dílčí faktory způsobující zpoždění při přenosu paketů mezi stanicemi v IP sítích. V experimentální síti PlanetLab proveďte geolokaci stanic pomocí metody GeoPing. Vytvořte skript, který bude provádět automatizované měření touto metodou. Dosažené výsledky srovnajte s běžně dosahovanými výsledky metody GeoPing.

## DOPORUČENÁ LITERATURA:

- [1] SARIKAYA, B.: Geographic location in the Internet, Springer, 2002. 214s. ISBN 1402070977
- [2] PADMANABHAN, V. N.: An Investigation of Geographic Mapping Techniques for Internet Hosts, SIGCOMM, 2001. 13s. ISBN 1-58113-411-8
- [3] BASSETT, E. K.: Towards IP Geolocation Using Delay and Topology Measurements. IMC, 2006. 13 s. ISBN 1-59593-561-4

**Termín zadání:** 7.2.2011

**Termín odevzdání:** 2.6.2011

**Vedoucí práce:** Ing. Jiří Balej

**prof. Ing. Kamil Vrba, CSc.**  
*Předseda oborové rady*

## UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

# ANOTACE

Tato bakalářská práce se zabývá metodami IP geolokace, tedy určením fyzické pozice stanice v síti Internet na základě znalosti IP adresy této stanice. Mým cílem bylo provést měření zpoždění v experimentální síti PlanetLab a následně pomocí geolokační techniky GeoPing odvodit neznámou geografickou polohu daného uzlu.

Část práce je věnována popisu několika geolokačních metod založených na měření zpoždění, zejména metody GeoPing, využívající vztahu mezi zpožděním a geografickou polohou na principu „nejbližšího souseda v prostoru zpoždění“. Důležitou částí, na které je metoda GeoPing založena je soubor stanic, tzv. dataset – ten je tvořen 9 stanicemi fungujícími jako sondy a 260 stanicemi v roli referenčních bodů a cílů. Určení neznámé polohy bylo provedeno u 60 stanic.

**Klíčová slova:** IP geolokace, PlanetLab, GeoPing, měření RTT, sonda, referenční bod

# ABSTRACT

This bachelor work deals with IP geolocation, which means the estimation of geographical location of unknown host in the Internet. The aim of my thesis was to make a delay measurement in PlanetLab Network and then with the use of GeoPing technique estimate the real location of the target.

The main part of this work is focused on several different delay-based methods, especially GeoPing which employs „Nearest Neighbor in Delay Space“ algorithm in order to exploit the relationship between network delay and geographic location. Base ground for GeoPing is a database of the machines at known location. This database is composed of 9 probe machines, 260 hosts is used like a landmarks and targets. Estimation of unknown geographical location was done on 60 hosts.

**Key words:** IP geolocation, PlanetLab, GeoPing, RTT measurement, probe, landmark

DOLEŽEL, P. *Současné možnosti nalezení fyzické pozice stanice v Internetu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2011. 54 s. Vedoucí bakalářské práce Ing. Jiří Balej.

## PROHLÁŠENÍ

Prohlašuji, že svoji bakalářskou práci na téma „SOUČASNÉ MOŽNOSTI NALEZENÍ FYZICKÉ POZICE STANICE V INTERNETU“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené práce dále prohlašuji, že v souvislosti s vytvořením tohoto projektu jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 2.6.2011

.....  
podpis autora

## PODĚKOVÁNÍ

Děkuji vedoucímu bakalářské práce Ing. Jiřímu Balejovi, za ochotu, pomoc, přátelský přístup a cenné rady při zpracování této bakalářské práce.

V Brně dne 2.6.2011

.....  
podpis autora

# OBSAH

|                                                                                            |           |
|--------------------------------------------------------------------------------------------|-----------|
| <b>Seznam obrázků .....</b>                                                                | <b>8</b>  |
| <b>Seznam tabulek .....</b>                                                                | <b>8</b>  |
| <b>Seznam zkratk .....</b>                                                                 | <b>9</b>  |
| <b>Úvod .....</b>                                                                          | <b>10</b> |
| <b>1 Zpoždění v síti (Latence).....</b>                                                    | <b>11</b> |
| <b>1.1 Pojem zpoždění .....</b>                                                            | <b>11</b> |
| <b>1.2 Dílčí faktory způsobující zpoždění .....</b>                                        | <b>11</b> |
| 1.2.1 Zpoždění na koncových zařízeních .....                                               | 11        |
| 1.2.2 Zpoždění na přenosových linkách .....                                                | 11        |
| 1.2.3 Zpoždění na mezilehlých zařízeních .....                                             | 12        |
| <b>2 Měření zpoždění.....</b>                                                              | <b>13</b> |
| <b>2.1 Protokol ICMP (Internet Control Message Protocol) .....</b>                         | <b>13</b> |
| <b>2.2 Program Ping (Packet InterNet Groper).....</b>                                      | <b>13</b> |
| <b>2.3 RTT (Round Trip Time) .....</b>                                                     | <b>14</b> |
| <b>2.4 Jednosměrné zpoždění (one-way delay) .....</b>                                      | <b>14</b> |
| <b>3 Geolokace .....</b>                                                                   | <b>15</b> |
| <b>3.1 IP geolokace.....</b>                                                               | <b>15</b> |
| <b>3.2 Základní rozdělení metod IP geolokace.....</b>                                      | <b>16</b> |
| 3.2.1 Metody využívající měření v síti .....                                               | 16        |
| 3.2.2 Metody využívající informace o IP adrese .....                                       | 17        |
| 3.2.3 Názvosloví.....                                                                      | 17        |
| <b>4 Geolokační metody využívající měření zpoždění.....</b>                                | <b>18</b> |
| <b>4.1 Současné omezení geolokačních technik.....</b>                                      | <b>18</b> |
| <b>4.2 GeoPing.....</b>                                                                    | <b>19</b> |
| 4.2.1 NNDS – Nearest Neighbor in Delay Space („Nejbližší soused v prostoru zpoždění“)..... | 19        |
| 4.2.2 Mapa zpoždění (delay map) .....                                                      | 21        |
| 4.2.3 Hledání vektoru zpoždění co nejvíce shodného s vektorem zpoždění cíle.....           | 22        |
| 4.2.4 Teoretické předpoklady pro přesné měření .....                                       | 23        |
| 4.2.5 Poslední míle .....                                                                  | 23        |
| 4.2.6 Efektivita .....                                                                     | 23        |

|            |                                                       |           |
|------------|-------------------------------------------------------|-----------|
| <b>4.3</b> | <b>Shortest Ping .....</b>                            | <b>24</b> |
| <b>4.4</b> | <b>Simple GeoPing (SGP) .....</b>                     | <b>25</b> |
| <b>4.5</b> | <b>Constraint Based Geolocation (CBG).....</b>        | <b>26</b> |
| 4.5.1      | Hranice vzdálenosti (distance constraint).....        | 26        |
| 4.5.2      | Kalibrování .....                                     | 27        |
| 4.5.3      | Multilaterace.....                                    | 29        |
| 4.5.4      | Možné případy při vytváření hranice vzdálenosti ..... | 31        |
| 4.5.5      | Efektivita .....                                      | 32        |
| <b>4.6</b> | <b>Speed of Light (SOI).....</b>                      | <b>33</b> |
| 4.6.1      | Podhodnocení hranic .....                             | 33        |
| 4.6.2      | Efektivita .....                                      | 34        |
| <b>5</b>   | <b>Měření pomocí metody GeoPing.....</b>              | <b>35</b> |
| <b>5.1</b> | <b>Geografické umístění.....</b>                      | <b>35</b> |
| <b>5.2</b> | <b>PlanetLab .....</b>                                | <b>36</b> |
| <b>5.3</b> | <b>Vytváření mapy zpoždění .....</b>                  | <b>37</b> |
| 5.3.1      | Sondy .....                                           | 37        |
| 5.3.2      | Referenční body.....                                  | 38        |
| 5.3.3      | Zjišťování polohy stanic.....                         | 39        |
| 5.3.4      | Měření RTT .....                                      | 40        |
| 5.3.5      | Zpracování výsledků.....                              | 40        |
| <b>6</b>   | <b>Rozbor a diskuze zpracovaných výsledků.....</b>    | <b>44</b> |
| <b>6.1</b> | <b>Efektivita .....</b>                               | <b>44</b> |
| <b>6.2</b> | <b>Porovnání výsledků .....</b>                       | <b>45</b> |
| <b>6.3</b> | <b>Rozbor výsledků .....</b>                          | <b>46</b> |
| <b>7</b>   | <b>Závěr .....</b>                                    | <b>48</b> |
|            | <b>Reference .....</b>                                | <b>50</b> |
|            | <b>Seznam příloh.....</b>                             | <b>53</b> |
| <b>A</b>   | <b>Obsah přiloženého DVD .....</b>                    | <b>54</b> |

## SEZNAM OBRÁZKŮ

|                                                                                                                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Obr. 4.1: Ukázka NNDS zobrazující: a) pozici 4 sond, 2 ref. bodů a cíle, b) vektor zpoždění $DV_1$ mezi $RB_1$ a sondami, c) vektor zpoždění $DV_2$ mezi $RB_2$ a sondami, d) vektor zpoždění $DV'$ mezi cílem a sondami. .... | 21 |
| Obr. 4.2: Scénář pro špatné určení pozice cíle metodou Shortest Ping.....                                                                                                                                                      | 25 |
| Obr. 4.3: Graf závislosti vzdálenosti ref. bodů na změřeném zpoždění, používaný pro určování <i>baseline</i> a <i>bestline</i> (převzato z [10]). ....                                                                         | 29 |
| Obr. 4.4: Hledání cíle pomocí multilaterace .....                                                                                                                                                                              | 30 |
| Obr. 4.5: Ukázkové případy při vytváření hranic vzdálenosti: a) nadhodnocení, b) podhodnocení c) mismatch .....                                                                                                                | 32 |
| Obr. 5.1: Mapa území, na kterém je prováděno měření (světle šedé oblasti). ....                                                                                                                                                | 35 |
| Obr. 5.2: Mapa znázorňující rozmístění sond .....                                                                                                                                                                              | 38 |
| Obr. 5.3: Mapa znázorňující rozmístění referenčních bodů.....                                                                                                                                                                  | 38 |
| Obr. 6.1: Graf výsledné efektivity určení pozice neznámé stanice metodou GeoPing .....                                                                                                                                         | 44 |
| Obr. 6.2: Graf porovnávající výsledky tří různých měření metodou GeoPing .....                                                                                                                                                 | 46 |

## SEZNAM TABULEK

|                                        |    |
|----------------------------------------|----|
| Tab. 4.1: Záznam v mapě zpoždění. .... | 22 |
| Tab. 5.1: Výsledky měření.....         | 41 |

## SEZNAM ZKRATEK

|      |                                     |                                                                                                               |
|------|-------------------------------------|---------------------------------------------------------------------------------------------------------------|
| AP   | - Address Prefix                    | - rozsah IP adres                                                                                             |
| AS   | - Autonomous System                 | - nezávislá, samostatně fungující část internetu (soubor IP sítí a směrovačů) s jednotnou směrovací politikou |
| CBG  | - Constraint-Based Geolocation      | - geolokační metoda                                                                                           |
| DNS  | - Domain Name Server                | - systém doménových jmen                                                                                      |
| DV   | - Delay Vector                      | - vektor zpoždění                                                                                             |
| GPS  | - Global Positioning System         | - globální polohový systém                                                                                    |
| ICMP | - Internet Control Message Protocol | - signalizační protokol v IP sítích                                                                           |
| IP   | - Internet Protocol                 | - protokol pro přenos dat přes paketové sítě                                                                  |
| IPTV | - Internet Protocol TV              | - služba distribuce digitální televize pomocí IP protokolu přenášená počítačovou sítí                         |
| ISP  | - Internet Service Provider         | - poskytovatel internetového připojení                                                                        |
| RB   | - Referenční bod                    | - stanice v roli referenčního bodu                                                                            |
| NNDS | - Nearest Neighbor in Delay Space   | - nejbližší soused v prostoru zpoždění                                                                        |
| P2P  | - Peer-to-peer                      | - rovný s rovným, typ architektury počítačových sítí                                                          |
| PoP  | - Point of Presence                 | - místo, odkud je k zákazníkům šířeno připojení k Internetu                                                   |
| RTD  | - Round Trip Delay                  | - obousměrné zpoždění                                                                                         |
| RTT  | - Round Trip Time                   | - alternativní výraz pro obousměrné zpoždění                                                                  |
| TBG  | - Topology Based Geolocation        | - geolokační metoda                                                                                           |
| URL  | - Unique Resource Locator           | - jednotný lokátor zdrojů informací, slouží ke specifikaci umístění zdrojů na Internetu                       |
| SGP  | - Single GeoPing                    | - geolokační metoda                                                                                           |
| SOI  | - Speed of Light                    | - geolokační metoda                                                                                           |
| SSH  | - Secure Shell                      | - zabezpečený komunikační protokol                                                                            |
| VoIP | - Voice over IP                     | - internetová telefonie fungující na TCP/IP                                                                   |

# ÚVOD

Způsobů hledání fyzické pozice stanice v Internetu je v dnešní době mnoho a mnoho je i důvodů, proč tak činit. Mezi nejpobulárnější způsoby patří vyhledávání pomocí IP adresy hledané stanice. V takovém případě mluvíme o tzv. IP geolokaci. Způsobů IP geolokace lze dnes nalézt velké množství, žádané jsou však zejména ty, jež jsou založeny na manuálně udržované databázi s IP adresami a jejich domnělými geografickými pozicemi. Přestože je tento způsob vyhledávání žádaný, nástup nových plně automatizovaných geolokačních metod s novými možnostmi překonává neautomatizované technologie a je jen otázkou času, kdy je přesune na okraj zájmu.

Ve své práci se věnuji metodám vyvinutým k automatickému určení neznámé pozice stanice, jejichž činnost se odvíjí od změřeného zpoždění. Pro plné porozumění těmto geolokačním metodám je nutné orientovat se v problematice zpoždění při přenosu dat mezi stanicemi v IP síti. Stejně tak je důležité znát způsoby jeho měření – dostateující informace k této problematice uvádím v první a druhé kapitole.

Ve třetí kapitole je prezentován krátký úvod k vymezení pojmu geolokace a zaměřuji se zde na její konkrétní typ – na IP geolokaci a následně její dělení. Pro splnění cíle mé práce bylo důležité obeznámit se s technologiemi pro vyhledávání neznámé polohy stanice v síti podle IP adresy. Zmíněny jsou široce rozšířené typy této metody, tedy geolokační metody založené na vyhledávání záznamu spjatého s IP adresou, který by dopomohl k určení pozice neznámé stanice, a také i sofistikovanější pracující metody.

Druhá polovina práce je zaměřena na rozbor metod určování fyzické pozice stanice na základě změřeného zpoždění. Tyto metody jsou podrobně popsány ve čtvrté kapitole. Po nabytí potřebných znalostí jsem vypracoval závěrečnou a stěžejní část práce – experiment, s cílem najít neznámé pozice stanic v síti Internet na základě znalosti jejich IP adres. K tomuto účelu jsem podle pokynů vedoucího práce použil metodu GeoPing, jejíž činnost je založena na hodnotách změřeného zpoždění, které byly naměřeny hlavně díky přístupu ke vzdáleným stanicím v síti PlanetLab.

# 1 ZPOŽDĚNÍ V SÍTI (LATENCE)

## 1.1 Pojem zpoždění

Zdrojem zpoždění v síti je každá její část. V telekomunikacích pojem zpoždění znamená čas, který uplyne od vyslání zprávy zdrojovou stanicí až po její přijetí stanicí cílovou. Konečná doba, kterou zpráva stráví na přenosové trase mezi těmito koncovými stanicemi (neboli jak velké bude koncové zpoždění), se skládá z několika dílčích částí a nehledě na tyto části můžeme zpoždění navíc rozdělit na deterministickou část a stochastickou část [2]. Deterministická část je neměnná a v podstatě se jedná o minimální hodnotu zpoždění. Naproti tomu, stochastická část bývá proměnlivá v závislosti na aktuálním stavu sítě [2], [15].

## 1.2 Dílčí faktory způsobující zpoždění

Těchto faktorů je celá řada, avšak rád bych zmínil jen natolik výrazné, které mohou ovlivňovat praktickou část mé práce, tedy měření prováděné za účelem geolokace.

### 1.2.1 Zpoždění na koncových zařízeních

Jako koncové zařízení považujeme zdroj a cíl přenášené informace. Zpoždění, které zde vzniká, je dáno časem potřebným ke zpracování paketu, přípravou k jeho k vysílání a následným vysíláním. Závisí mimo jiné na „výkonu“ prvku, který tuto činnost provádí. I přesto, že v podstatě není ovlivňováno stavem sítě, tak z důvodu, že čas pro zpracování paketu nebývá přesně stejný, je považováno za zpoždění stochastické [2].

### 1.2.2 Zpoždění na přenosových linkách

Je to čas, po který se signál šíří po přenosovém médiu, resp. doba, po kterou se elektromagnetická vlna šíří fyzickým kanálem přenosové trasy. Toto zpoždění tedy závisí na použitém médiu a délce fyzické trasy. Jelikož přenosové linky jsou vedeny zejména podél železničních tratí, silnic, dálkových rozvodů elektrické sítě a ke všemu může dojít ke směřování, které nevybere fyzicky nejkratší cestu, tak v naprosté většině případů je skutečná délka média větší než je přímá vzdálenost mezi vysílačem a přijímačem [2], [15].

Ideální rychlost šíření signálu médiem se blíží rychlosti šíření světla ve vakuu ( $c = 299\,792\,458\text{ m}\cdot\text{s}^{-1}$ ). V praxi je v transportních sítích použito především optické vlákno, které signálu umožňuje šířit se rychlostí přibližně  $\frac{2}{3}c$ , tedy přesněji cca  $194\,895\text{ km}\cdot\text{s}^{-1}$  [15].

Hodnota tohoto zpoždění je spodní hranice celkového zpoždění a v podstatě se jedná o deterministickou složku [2].

### **1.2.3 Zpoždění na mezilehlých zařízeních**

Jedná se o zpoždění vznikající na všech aktivních prvcích. Mezi hlavní úkoly těchto prvků patří směrování a přepínání datových jednotek.

Latenci na těchto prvcích tvoří zpoždění ve výstupní frontě, zpoždění při přesunu datové jednotky ze vstupu na výstup a zpoždění ve výstupní frontě.

Podle [15] uvažujeme, že různé aktivní prvky mají zpoždění obvykle v tomto rozmezí hodnot:

- přepínač (switch) – 1 až 10  $\mu$ s,
- směrovač (router) – 10 až 100  $\mu$ s,
- opakovač a aktivní hub – zanedbatelné.

Ke všemu je třeba brát na vědomí typickou vlastnost všech aktivních prvků, kterou představuje zvýšení zpoždění v okamžiku vysokého zatížení prvku. V takovém případě výše zmíněné hodnoty dosahují až několikanásobných hodnot prvku nezatíženého.

Z toho vyplývá, že zpoždění na mezilehlých uzlech je stochastická složka zpoždění [15].

## 2 MĚŘENÍ ZPOŽDĚNÍ

Tuto činnost lze provádět různými způsoby, které dělíme do několika kategorií. Zpoždění můžeme měřit pomocí softwarových aplikací, např. pomocí nejběžnější utility, programu Ping. Samozřejmě, alternativních programů existuje celá řada. Liší se prezentací naměřených veličin, metodami, které používají pro měření a množstvím nastavení. Využívá se i hardwarového řešení, např. pomocí test boxů.

Dále můžeme měřit zpoždění jednosměrné a obousměrné (RTT). Obě varianty mají své výhody, nevýhody a specifika. Volba způsobu měření se odvíjí od cílů, kterých potřebujeme tímto měřením dosáhnout, a zejména od prostředků, které máme k dispozici. Základním pravidlem pro přesné měření je provést měření několik (nejméně 10 až 15 [15]), přičemž nejmenší hodnotu bereme jako hodnotu vypovídající.

### 2.1 Protokol ICMP (Internet Control Message Protocol)

Dle RFC792 [25] je povinnou součástí IP protokolu a v sítích postavených na tomto (IP) protokolu slouží k signalizaci neobvyklých situací (ohlašování chybových zpráv, např. že požadovaná služba není dostupná nebo že požadovaný uzel není dosažitelný) a mimo jiné se dá využít i k měření zpoždění. ICMP zprávy se přenášejí v datové části IP datagramu, někdy jsou však tyto zprávy z bezpečnostních důvodů ignorovány, respektive zahazovány, což může právě při měření zpoždění vést k zásadním problémům. Program Ping využívající ICMP paketů [14] odesílá dotaz na vzdálenou stanici v podobě zprávy *Echo Request* (typ = 8, kód = 0)<sup>1</sup>, načež očekává odpověď, příjem zprávy *Echo Reply* (typ = 0, kód = 0). Čas mezi odesláním a přijetím této zprávy je roven zpoždění.

### 2.2 Program Ping (Packet InterNet Groper)

Aplikace vytvořená ke zjišťování doby zpoždění mezi dvěma stanicemi v IP síti, měřením RTT. Program pracuje na principu výměny ICMP zpráv (viz kapitola 2.1) a pokud je tato výměna úspěšná, vypíše hodnotu RTT zpoždění. Pokud úspěšná není, program vypíše chybovou hlášku zdůvodňující neúspěch. Program se v příkazové řádce spouští napsáním příkazu `ping` (+ další možné parametry), za který dopíšeme cílovou adresu. Cílová adresa může být zadána ve tvaru IP nebo URL adresy.

---

<sup>1</sup> Pozn.: Typ specifikuje typ zprávy, kód určuje hodnotu, jakou může daný typ zprávy nabývat

Program má také většinu svých parametrů nastavitelných (např. délku paketu, časový limit při čekání na odpověď, hodnotu TTL) a výstupní hodnoty odezvy jsou ve čtyřech variantách: minimální hodnota, průměrná hodnota, maximální hodnota a směrodatná odchylka.

## 2.3 RTT (Round Trip Time)

Nebo někdy také RTD (Round Trip Delay) je čas, během kterého je signál vyslán ze zdrojové stanice, doputuje k příjemci, kde je zpracován a odeslán jako odpověď k původnímu zdroji, kde je opět přijat. Z našeho pohledu je doba zpracování na koncových zařízeních zanedbatelná, avšak zanedbatelná není doba potřebná k samotnému šíření signálu skrz síť. Jak již bylo řečeno v části 1.2, na tuto dobu má velký vliv délka trasy a zatížení či výpadek sítě, který zapříčiní, že data jsou směrována jinou cestou – to často vede k navýšení doby RTT a zejména také k mystifikaci při odvozování jednocestného zpoždění.

K měření je využíváno paketů protokolu ICMP – ty mají ve směrovačích nízkou prioritu, což může vést k nežádoucímu navýšení RTT (pakety čekají, až jsou v mezilehlých uzlech zpracovány důležitější data). Jak bylo zmíněno v kapitole 2.1, některé uzly (či dokonce cílová stanice) mohou být nastaveny tak, že odpověď na dotaz Echo Request protokolu ICMP je zakázána, což má za následek, že měření RTT pomocí ICMP paketů není možné provést [16].

Hlavní výhodou měření RTT je jeho proveditelnost využitím pouze jedné stanice. Naopak, jak již bylo naznačeno, pokud chceme z hodnot RTT odvodit čas potřebný k dosažení cílové stanice, nezbyvá nám nic jiného, než výslednou hodnotu RTT podělit dvěma, což je z důvodů nesymetrickosti mnohdy nepřesné [2].

### Odvození vzdálenosti vzhledem k RTT

Autoři publikace „An Investigation of Geographic Mapping Techniques for Internet Hosts“ [19] empiricky zjistili, že při RTT do 5 ms se stanice ve více než 90 % případů nachází do 50 km od sebe a při RTT menším než 10 ms je to nanejvýš 300 km.

## 2.4 Jednosměrné zpoždění (one-way delay)

Toto zpoždění je měřeno ze dvou od sebe vzdálených koncových stanic, což je v podstatě hlavní nevýhodou této metody. Základem pro precizní měření je, aby tyto stanice byly velmi přesně časově synchronizovány (řádově v mikrosekundách [2]). Výsledkem měření je vysoce směrodatná hodnota zpoždění.

## 3 GEOLOKACE

Geolokace je v souvislosti s internetovými aplikacemi čím dál žádanější službou. Tento pojem vystihuje činnost, kdy se snažíme určit geografickou pozici objektu (počítače připojeného k Internetu, mobilního telefonu...) na základě znalosti a následné analýzy jeho určité vlastnosti či známých údajů. Mezi tyto údaje patří např. informace z GPS modulu, informace z přístupového bodu Wi-Fi sítě, nebo IP adresa počítače [31].

### 3.1 IP geolokace

U tohoto typu geolokace je vlastností, respektive údajem, ze kterého vycházíme, IP adresa cílové stanice. Avšak vycházet při geolokaci pouze z IP adresy není snadný úkol, protože ta v sobě nenese žádné informace o geografické poloze vztahující se k dané stanici – mezi IP adresou a geografickým umístěním stanice neexistuje žádná přímá souvislost [16].

Pokud se v mé práci budu dále zmiňovat o „geolokaci“, budu tím mít na mysli konkrétně IP geolokaci.

#### Užití

V dnešní době se odvětví geolokačních metod a aplikací silně rozvíjí a je středem zájmu jak vědecké sféry, tak i sféry veřejné. Proč tomu tak je? Jistě si uvědomujeme fakt, že Internet se neustále rozvíjí – počet uživatelů neustále roste, počet služeb, které lze nalézt na Internetu neustále roste, ale zejména i objem dat přenášených v Internetu roste [29]. Vzhledem k tomuto lze nalézt spoustu odvětví a činností, kde se dají geolokační aplikace využít. Představím z nich jen některé vybrané s tím, že dalších možností využití je mnoho [11], [16].

**Online reklama:** S neustále se zvětšujícím se počtem uživatelů Internetu se Internet stává ideální „reklamní plochou“ – nicméně, poněkud rozsáhlou a je žádané reklamu zaměřit na správnou cílovou skupinu. Je výhodnější nabízet produkty a služby z určité geografické oblasti právě potencionálním zákazníkům v okolí.

**Aktuálnost z hlediska polohy:** Taktéž i samotní uživatelé internetu, kteří například často cestují, si přejí mít informace vztahující se zejména k jejich okolí (např. počasí, kulturní akce, zprávy) a je žádané, aby tyto informace byly cílené automaticky.

**VoIP telefonie:** Spolu s rozvojem Internetu se rozvíjí i odvětví VoIP telefonie. Zde bývá přáním uživatelů zjistit místo, kde se nachází volající, či volaný. Mohou nastat výjimečné případy tísňových hovorů, kdy vznikne krizová situace a volající není schopen odpovídat (je zmatený, ve stresu nebo upadne do bezvědomí) – v tu chvíli může přesné určení lokace volajícího zachránit život.

**Ochrana proti podvodům:** další užitek geolokačních praktik lze spatřit v pomoci odhalování podvodů s kreditními kartami. Například, pokud si zákazník zvolí oblast pro obvyklé používání jeho karty, může se lehce přijít na případné zneužití provedené z místa, které do této oblasti nepatří.

**Hledání nejbližšího zdroje dat:** jak jsem již zmínil, objem dat přenášených v globální síti dramaticky roste každým dnem. Přispívá k tomu například i rozvoj IPTV, používání P2P (peer-to-peer) sítí či masivní nárůst uživatelů hledajících zábavu v hraní on-line her, které jsou čím dál více náročné na objem dat potřebných k přenosu a závislé na rychlosti jejich přenosu. Aby se tato data dostala k uživatelům co nejrychleji a zatížila síť co nejméně, je potřeba zvolit jejich zdroj tak, aby byl co nejbližší jak z pohledu relativní pozice v síti, tak i z pohledu fyzické vzdálenosti, od které se odvíjí, kudy data pocestují a jak dlouho jim cesta potrvá. Z toho plyne, že důležitým vodítkem ke správné volbě zdroje dat – buď uživatele (v P2P síti) nebo serveru (případně jeho mirroru) – je znalost jeho geografické polohy, ze které jde následně vyvodit vhodnost tohoto zdroje.

## **3.2 Základní rozdělení metod IP geolokace**

Metody geolokace dělíme podle toho, zda ke své činnosti využívají měření zpoždění anebo se bez jakýchkoliv měření obejdou a informace hledají „někde“ v síti.

### **3.2.1 Metody využívající měření v síti**

Pro lepší pochopení těchto metod je z jejich podstaty žádoucí orientovat se v problematice zpoždění a jeho měření a v problematice topologií sítí. Z toho plyne, že tyto metody můžeme rozdělit na dvě odvětví – do prvního patří ty, jež jsou založené zejména na měření zpoždění a do druhého takové, které využívají znalosti topologie sítě. Existují i metody, které kombinují obě tato odvětví a nutno říct, že patří mezi metody s nejlepšími výsledky.

### 3.2.2 Metody využívající informace o IP adrese

Patří sem všechny metody, které mají geolokaci založenou na záznamech z databází s IP adresami a jejich pozicemi (soukromé, např. [13] nebo veřejně dostupné databáze, např. WhoIs databáze [26]) a dále metoda GeoTrack [19], která pozici odhaduje na základě kódu obsaženého v DNS směrovači patřícího cíli, nebo ve směrovačích poblíž cíle. Tyto metody jsou dnes stále velmi rozšířené a oblíbené [18] i přes nutnost údržby databází, ze kterých čerpají.

### 3.2.3 Názvosloví

Ve většině prací zabývajících se IP geolokací můžeme narazit na výrazy s touto činností úzce spjaté. Tyto výrazy se samozřejmě vyskytují i v mé práci a proto uvádím ty nejběžnější společně s jejich popisem, který vystihuje jejich podstatu:

**Cíl** (T; *target*): za tímto výrazem se skrývá stanice, jejíž geografickou polohu neznáme, ale snažíme se ji zjistit.

**Referenční bod** (RB; *landmark*): jedná se o stanici, která slouží jako referenční geografický bod, čili geografické souřadnice této stanice - bodu jsou nám známy. Některé geolokační metody potřebují, aby byl RB schopný měřit RTT (viz kapitola 2.3) a pokud tato metoda používá ref. body schopné měření v kombinaci s ref. body měření neschopnými, rozlišujeme referenční body na aktivní (schopné měřit) a pasivní (bez této schopnosti).

**Sonda** (S; *probe*): opět se jedná o stanici, o které víme, kde se nachází. Jejím hlavním úkolem je sondování – měřením zjišťovat zpoždění k dalším stanicím (referenčním bodům či cílům). Rozdíl mezi sondou a aktivním ref. bodem bývá takový, že u sondy nepotřebujeme tak akutně znát její geografickou polohu.

**Dataset**: je to soubor všech stanic, které využíváme k nalezení cíle. Dataset obsahuje například 10 sond a 100 ref. bodů. Cíl se obvykle nepovažuje za součást datasetu, ovšem získáme-li o jeho poloze důvěryhodné informace, dá se využít i jako referenční bod.

## 4 GEOLOKAČNÍ METODY VYUŽÍVAJÍCÍ MĚŘENÍ ZPOŽDĚNÍ

Tato kapitola je věnována výstižnému popisu pěti geolokačních technik, založených na soustavě referenčních bodů (které musejí být vybírány s rozmyslem, pokud chceme dosáhnout co nejlepších výsledků [10], [32]) a vycházejících ze změřeného zpoždění. Nutno říci, že ne všechny metody informací o zpoždění využívají stejným způsobem a i jejich přesnost a náročnost na uvedení do praxe se vzájemně značně liší.

Metody tohoto typu slouží k plně automatickému zjišťování polohy cíle, což znamená, že mezi jediná vstupní data, která jsou zpracovávána manuálně, patří geografické souřadnice a IP adresy známých stanic. Jistě stojí za zmínku, že abychom se dostali k těmto informacím, je mnohdy používáno různých technik, založených i na jiných principech než měření zpoždění. Poté, co uvedeme takovéto metody v činnost, není obvykle nutná další spolupráce s lidskými zdroji a jejich funkce nezávisí na žádných rozsáhlých databázích, které je nutné denně aktualizovat. Údržba vyžaduje jen minimální úsilí a tyto metody teoreticky dovedou fungovat neomezeně dlouho se stále stejnou přesností, kdy jediné omezení představuje stav sítě.

Základní princip těchto metod skýtá oproti většině ostatních velkou výhodu, která se zřejmě projeví až s plným nástupem IPv6<sup>2</sup>. Jakmile bude tento protokol rozšířen, poskytovatelé internetu budou moci díky zvětšení adresového prostoru rozsahy IP adres přeorganizovat v podstatě podle potřeby [28]. To je však pro tyto geolokační metody nepodstatný fakt a svou činnost budou moci vykonávat i navzdory přesouvání adresových bloků a „neustále“ se měnícím IP adresám.

### 4.1 Současné omezení geolokačních technik

V dnešní době je spousta uživatelů k Internetu připojena přes proxy server, firewall, nebo bránu (*gateway*) patřící poskytovateli internetu, takže je častým jevem, že při pohledu na IP adresu tohoto uživatele je jeho adresa shodná s adresou „uzlu“ (může se jednat i o softwarové řešení těchto „bran“) který slouží pro „zprostředkování“ připojení. Podle [12] a [19] je to jedna z překážek všech přesných geolokačních technik založených na znalosti IP adresy cíle. V jistých případech tento jev nemá příliš velký dopad na určení polohy, a to když je lokace uživatele nedaleko samotného proxy serveru (např. univerzitní kampus, uživatel který sídlí ve stejném městě, nebo blízkém okolí jako jeho poskytovatel internetu (ISP) apod.). Jsou však případy, kdy je uživatel od „brány“ vzdálen až stovky kilometrů. V této situaci nám geolokační aplikace pravděpodobně udá

---

<sup>2</sup> IPv6 je nová verze internetového protokolu, jenž nahrazuje IPv4

výslednou cílovou pozici naprosto nepřesně, což je ve výsledku mnohem horší, než kdyby cílová pozice zůstala neurčena. Tomuto problému některé geolokační aplikace umí předcházet a když zjistí, že pozice cíle, kterou odhadly, je ve skutečnosti pozice „brány“ a samotný cíl je naprosto na jiném místě, raději odhad pozice cíle neuvedou vůbec (např. metoda Topology Based Geolocation [16]). Tyto metody jsou většinou založeny na prozkoumávání topologii sítí, resp. na rozpoznávání lokace podle AP (adresového prefixu) či AS (autonomního systému), do kterého stanice patří. Metody, kterým se věnuje tato práce, jsou takovýmito jevem bohužel omezeny.

## 4.2 GeoPing

Metoda GeoPing, kterou světu představili její autoři Padmanabhan a Subramanian (a z jejichž díla [19] vycházím) svou činnost staví na znalosti zpoždění mezi sondami, referenčními body a cíly, a znalosti geografických pozic referenčních bodů. Přestože mezi geografickou pozicí a zpožděním nemusí být ve všech případech a za všech okolností vzájemný vztah, dosahuje díky svému přístupu zajímavých výsledků.

Tato metoda své měření provádí v datasetu, kde několik stanic má roli sondy (*probe*), čili jsou schopny provádět měření zpoždění a jejich poloha nám je (většinou) známa a několik (podstatně více) jich slouží jako referenční body. U této metody jsou geografické pozice ref. bodů jediné možné výsledky, které dostaneme jako odhad pozice požadovaného cíle. Jinak řečeno, pozice cíle se může shodovat jen s pozicí jednoho ref. bodu. Toto vede ke konečnému počtu odpovědí, které může GeoPing nabídnout a tento počet odpovědí se logicky rovná počtu ref. bodů, což je jedna z příčin způsobujících nepřesnost této metody (pojem nepřesnost je myšlena skutečnost, že i nejvhodnější ref. bod, jehož pozice byla vybrána za nejvíce se shodující s cílem, může být stále velmi vzdálený od reálné pozice cíle a tato vzdálenost vytváří neoddělitelnou část chyby určení pozice cíle). Tento druh chyby nemůže být v podstatě zcela eliminován – vždy se totiž najde cíl, jehož pozice se neshoduje s pozicí ref. bodu v datasetu. Ve všeobecném měřítku může být pouze snížen vhodným rozmístěním sond a vhodným počtem ref. bodů [12], [32]. Přesného odhadu (myšleno s přesností shody alespoň v rámci oblasti, regionu) se dá docílit za předpokladu, že alespoň jeden referenční bod má velmi podobnou geografickou polohu jako cíl. Bohužel ani v tom případě není vždy pravidlem, že tento ref. bod bude vybrán a dochází k nepřesné geolokaci.

### 4.2.1 NNDS – Nearest Neighbor in Delay Space („Nejbližší soused v prostoru zpoždění“)

Tento pojem je nosnou myšlenkou celé metody GeoPing, resp. díky NNDS vznikají základní stavební kameny pro algoritmus, jehož pomocí je nakonec vypočítáno geografické umístění cíle,

aniž by bylo potřeba jakéhokoliv přepočítávání zpoždění na vzdálenost. Úvod tohoto principu pro názornější rozbor formuluji do příkladu:

Představme si několik stanic v síti: stanici ① a ② a  $m$  dalších stanic, které zastávají roli ref. bodů. Stanice ① má k těmto ref.bodům soubor určitých zpoždění (o počtu  $m$ ), dohromady tato zpoždění vytvoří *vektor zpoždění (delay vector)  $DV$* . Pokud se vektor zpoždění  $DV$  vytvořený měřeními ze stanice ① shoduje (nebo alespoň velmi podobá) s tím ze stanice ②, tak můžeme vyvodit, že tyto dvě stanice mají stejnou (nebo alespoň velmi podobnou) geografickou pozici.

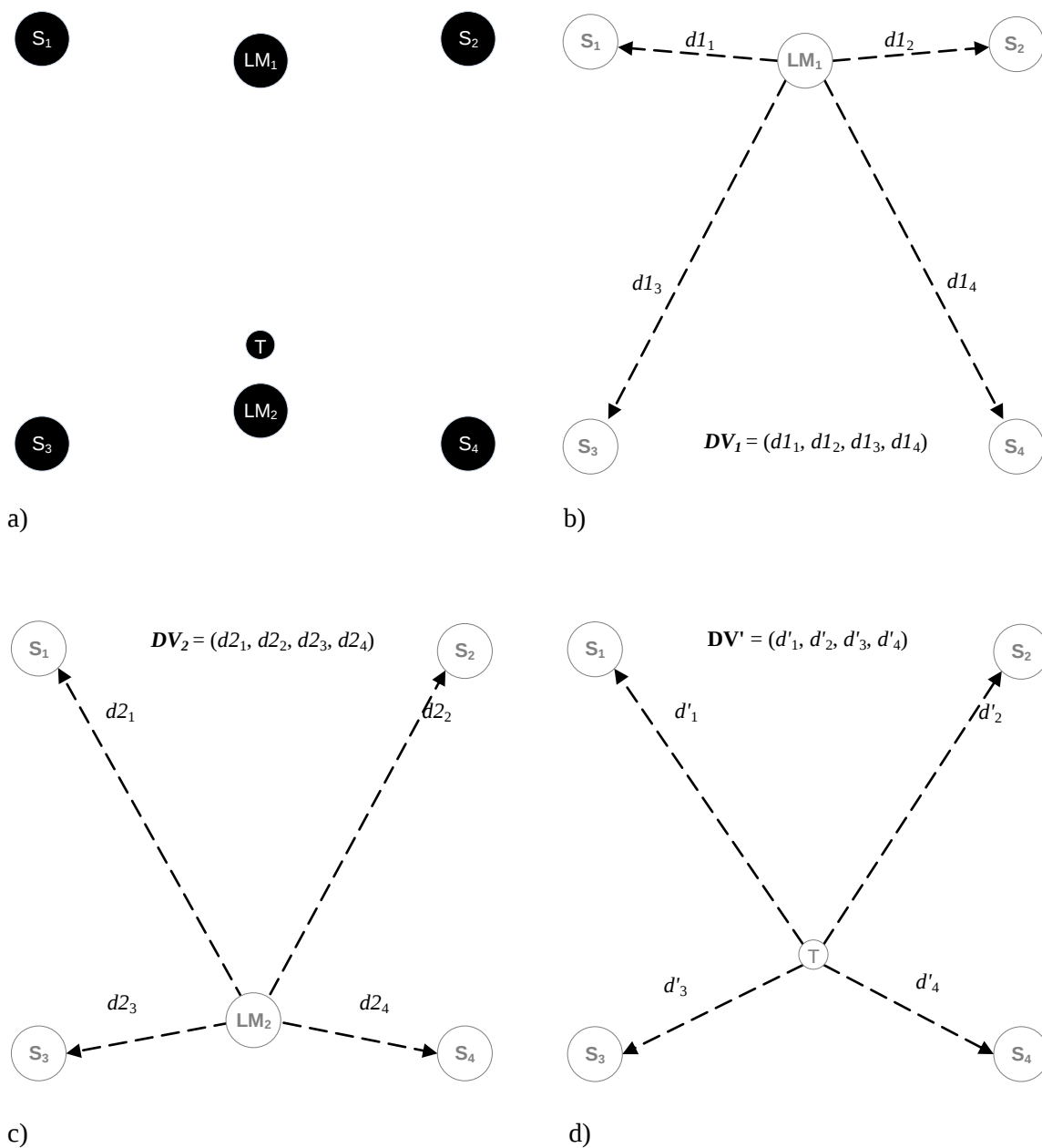
Z tohoto příkladu přejdeme do všeobecné terminologie a můžeme říci, že GeoPing geolokaci provádí porovnáváním *vektoru zpoždění cíle  $DV'$*  (vektor mezi sondami a cílem);  $DV' = (d'_1, \dots, d'_N)$ , s vektorem zpoždění  $DV$ ;  $DV = (d_1, \dots, d_N)^3$  každého ref. bodu (tedy mezi sondami a ref. bodem), což je níže názorně předvedeno na Obr. 4.1. Důležité je vybrat z vektorů zpoždění připadajících jednotlivým ref. bodům (tyto jsou nazvány  $DV_{(1-x)}$ ,  $x$  = počet ref. bodů v datasetu) takový, který se co možná nejvíce shoduje s vektorem zpoždění cíle  $DV'$ . Geografická pozice ref. bodu, jehož vektor  $DV$  (takže např. vektor  $DV_{23}$ ) byl určen jako nejvíce odpovídající s vektorem cíle  $DV'$ , je pak odhadnutou geografickou polohou cíle. Kompletací vektorů zpoždění do jednoho souboru vytváříme *mapu zpoždění*.

Následuje další jednoduchý a výstižný příklad, jak si lze představit nejbližšího souseda v prostoru zpoždění:

Na obrázku 4.1 v části a) vidíme rozmístění 4 sond ( $S_1-S_4$ ), 2 ref. bodů ( $RB_1, RB_2$ ) a cíle T, který leží „poblíž“  $RB_2$ . Vše je bráno v prostoru zpoždění (*delay space*). V části b) vidíme, jak vypadá vektor  $DV_1 = (d1_1, d1_2, d1_3, d1_4)$  patřící  $RB_1$ . V části c) vidíme, jak vypadají jednotlivá zpoždění šířící se od  $RB_2$  a tvořící  $DV_2$ ;  $DV_2 = (d2_1, d2_2, d2_3, d2_4)$ . Nakonec, poslední část d) zobrazuje zpoždění, která tvoří  $DV'$ . Už na první pohled vidíme, který ref. bod má všechna zpoždění podobná jako cíl T. Můžeme tedy prohlásit, že ref. bod  $RB_2$  je *nejbližší soused v prostoru zpoždění* (NNDS) k cíli T, a proto za pozici cíle prohlásíme oblast, ve které se nachází ref. bod  $RB_2$ .

---

<sup>3</sup>  $d'$  a  $d$  je jedna ze změřených hodnot RTT (vztahující se buď k cíli, respektive referenčnímu bodu),  $N$  = počet sond



Obr. 4.1: Ukázka NNDS zobrazující: a) pozici 4 sond, 2 ref. bodů a cíle, b) vektor zpoždění  $DV_1$  mezi  $RB_1$  a sondami, c) vektor zpoždění  $DV_2$  mezi  $RB_2$  a sondami, d) vektor zpoždění  $DV'$  mezi cílem a sondami.

#### 4.2.2 Mapa zpoždění (delay map)

Tento výraz vystihuje kompletní soubor vektorů zpoždění. Každý záznam v mapě obsahuje:

- souřadnice ref. bodu, mezi nímž a sondami jsou měřena zpoždění,

- vektor zpoždění  $DV$ , který je tvořen zpožděními od ref. bodu k sondám;  $DV = (d_1, \dots, d_N)$ ,  $N$  = počet sond.  $DV$  tvoří minimální hodnoty změřených zpoždění (zpoždění je třeba měřit vícekrát, viz kapitola 2).

Zde uvádím ukázkou, jak může vypadat jeden záznam v mapě zpoždění (tento záznam se vztahuje k datasetu tvořeného 9 sondami a nejmenšími změřenými zpožděními 32 ms, 14 ms, 82 ms, 61 ms, 10 ms, 53 ms, 79 ms, 39 ms a 28 ms)

Tab. 4.1: Záznam v mapě zpoždění.

| $DV\#^4$ | $d_1$<br>[ms] | $d_2$<br>[ms] | $d_3$<br>[ms] | $d_4$<br>[ms] | $d_5$<br>[ms] | $d_6$<br>[ms] | $d_7$<br>[ms] | $d_8$<br>[ms] | $d_9$<br>[ms] | Souřadnice |
|----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|------------|
| $DV_1$   | 32            | 14            | 82            | 61            | 10            | 53            | 79            | 39            | 28            | „Praha“    |

Při vytváření „mapy zpoždění“ je důležité dodržet několik základních pravidel:

- při vytváření mapy zpoždění je třeba použít co nejvíce ref. bodů,
- měření zpoždění provádíme vícekrát, minimální počet opakování je 10–15 [19].

Uvědomíme-li si možnosti této metody, pak vyplývá, že výsledky odhadu pozice je třeba brát spíš jako přibližnou polohu cíle, která se nedá považovat za přesné souřadnice.

#### 4.2.3 Hledání vektoru zpoždění co nejvíce shodného s vektorem zpoždění cíle

Jak je uvedeno výše, pro určení pozice cíle je třeba najít takový  $DV$ , který se nejlépe shoduje s  $DV'$ . Hodnoty  $DV'$  zjistíme podobně jako hodnoty  $DV$ , měřením zpoždění ze sond k cíli,  $DV' = (d'_1, \dots, d'_N)$ . Následně můžeme prozkoumat mapu zpoždění. Ke srovnání mezi  $DV' = (d'_1, \dots, d'_N)$  a jednotlivými  $DV = (d_1, \dots, d_N)$  použijeme princip Euklidovské vzdálenosti (podle [19]). Euklidovská vzdálenost mezi vektorem  $DV$  a  $DV'$  je

$$\sqrt{(d_1 - d'_1)^2 + \dots + (d_N - d'_N)^2}, \quad (4.1)$$

kde  $d$  a  $d'$  představuje hodnoty RTT.

---

<sup>4</sup> Dále ve studii s těmito  $DV$  pracuji a je i třeba rozlišit, o jaký  $DV$  se jedná, ve smyslu ke kterému landmarku je měření prováděno. Rozhodl jsem se používat logické značení – pro vektor  $DV$  vůči  $RB_1$  používám značení  $DV_1$ .)

Čím více se hodnota výsledku blíží nule, tím je Euklidovská vzdálenost menší a tím více se daná dvojice **DV** a **DV'** shoduje.

#### 4.2.4 Teoretické předpoklady pro přesné měření

Při vytváření datasetu pro GeoPing je třeba držet se následujících pravidel (viz [19]):

- ideální počet sond je 7 až 9. Nižší, ale i vyšší počet obvykle způsobuje méně efektivní geolokaci,
- sondy by měly z geografického hlediska dobře rozmístěny – tzn. s rozumným rozestupem, resp. by neměly tvořit „shluk/y“,
- počet ref. bodů by měl být co nejvyšší – čím více jich máme, tím více máme oblastí, do kterých můžeme směřovat odhad umístění cíle a tím je větší pravděpodobnost odstranění chyby (tzv. „jemná zrnitost“),
- jako ref. bod je vhodné použít takové stanice, o kterých můžeme říct, že jsou z hlediska měření RTT dobře dostupné (odezva není navýšena zbytečným směřováním, nepřímostí trasy či pomalým připojením),
- každé zpoždění naměřit 10 až 15 krát a za směrodatnou hodnotu brát tu nejmenší.

#### 4.2.5 Poslední míle

*Poslední míle* na trase ke stanici může neúměrně navýšit zpoždění oproti zbytku celé trasy a celé měření RTT tak může být dosti zkreslené. Proto pokud zjistíme, že poslední míle je tvořena nevhodným připojením (tj. připojením které způsobuje velké zpoždění), např. satelitní či dial-up linkou, je vhodnější nahradit tuto stanici posledním směrovačem na cestě k cíli, s tím, že pokud neznáme polohu tohoto mezilehlého prvku, odvodíme ji z polohy těžko dosažitelné stanice (viz [19]).

#### 4.2.6 Efektivita

Při splnění všech praktických předpokladů, se autorům metody GeoPing [19] podařilo najít cíle s přesností pohybující se v těchto číslech:

- 90% provedených pokusů odhadu bylo s chybou menších než zhruba 1350 km,
- 75% pokusů mělo chybu menší než přibližně 750 km,
- 50% pokusů mělo chybu odhadu pod 400 km,
- 25% pokusů bylo úspěšných natolik, že chyba se pohybovala pod hodnotou zhruba 150 km.

Autoři metody CBG [10], při srovnávání úspěšnosti těchto dvou metod dospěli k poněkud lepším výsledkům:

- medián chyby pro U.S dataset byl 150 km,
- medián chyby pro Evropský dataset byl přibližně 100 km.

(autoři však neuvádějí, jestli se jednalo o srovnání pouze s metodou GeoPing a nebo byla-li tato metoda doplněna ještě nějakým dalším typem geolokace, např. metodou Shortest Ping [16], což by mohlo být důvodem dosažení lepších výsledků)

Největší vliv na přesnost má počet sond – jakmile k měření není použito alespoň 7 sond, chyba začne dle měření prováděných ve [19] stoupat a roste nepřímou úměrou k počtu sond. A naopak, při použití více než 9 sond začne chyba odhadu pozice cíle růst také. Dalším důvodem k růstu chyby přesnosti jsou špatně rozmístěné sondy – tento nárůst však není tak znatelný, jako při nedodržení pravidla o optimálním počtu sond.

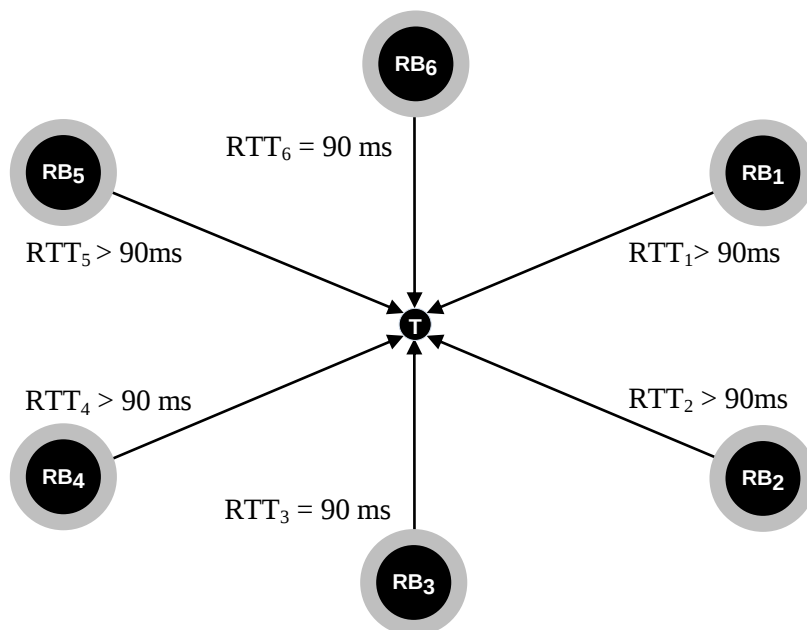
### 4.3 Shortest Ping

Metoda Shortest Ping [16] je asi nejjednodušší z metod založených na měření zpoždění (RTT). Tento způsob geolokace vychází z principu použití nejmenší hodnoty RTT. Základem je opět vytvoření datasetu, který tvoří soubor ref. bodů. Rozmístění referenčních bodů je nám známé a od nich měříme zpoždění k cíli.

Pozice referenčního bodu, ze kterého jsme k cíli naměřili nejmenší RTT, je považována za pozici cíle (takže obdobně jako u metody GeoPing máme diskrétní množství možných pozic, kde se cíl může nacházet). I přesto, že je metoda Shortest Ping principiálně jednodušší než GeoPing, její autoři tvrdí, že je tato metoda úspěšnější, což podložili i experimentálními výsledky obou těchto metod.

Shortest Ping má však oproti GeoPing jednu značnou nevýhodu. Pokud totiž nastane případ, který demonstruji níže, je velká pravděpodobnost, že určení pozice cíle bude nesprávné [11].

Podívejme se na situaci znázorněnou na Obr. 4.2. Vidíme zde 6 ref. bodů ( $RB_1$  až  $RB_6$ ).  $RB_3$  a  $RB_6$  mají společně nejnižší RTT vůči cíli T ( $RTT = 90$  ms), ale přesto jsou od sebe oba dva ref. body velmi vzdálené. Z toho už pomalu vyplývá problém, který nastane – a to, že poloha cíle nebude moct být určena správně a místo toho bude cíl odhadován v oblastech okolo  $RB_3$  a zároveň i  $RB_6$ .



Obr. 4.2: Scénář pro špatné určení pozice cíle metodou Shortest Ping.

## Efektivita

Podle výsledků dosažených autory této metody [16] je Shortest Ping efektivní hlavně v případech kdy RTT je nízké a ref. bod je blízko, čili oblast kde se nachází cíl je malá. Při experimentu s touto metodou její autoři dosáhli mediánu chyby pro 53 cílů (se zpožděním menším než 4 ms) pouhých 15 km; pro 75 cílů se (zpožděním větším než 4 ms) byl medián chyby 266 km. Založeno na těchto výsledcích, je efektivita této velmi jednoduché metody v průměru lepší než u metody GeoPing a blíží se k úspěchům metod CBG a SOI.

## 4.4 Simple GeoPing (SGP)

Princip, který je u této metody geolokace používán vychází z metody GeoPing. Jak název napovídá, postup při geolokaci bude ještě jednodušší než metodou GeoPing, avšak stojí na podobných základech [12].

Nejprve je vytvořen dataset skládající se z několika ref. bodů a na rozdíl od metody GeoPing pouze z jediné měřicí stanice, ze které měříme zpoždění ke všem ref. bodům a cílům. RTT směrem

k cíli srovnáme s jednotlivými hodnotami RTT k ref. bodům, přičemž pozice ref. bodu s nejvíce se shodujícím zpožděním je považována za výsledné umístění cíle. Aby se zvýšila přesnost techniky Simple GeoPing, měření hodnot zpoždění opakujeme  $n$ -krát a výslednou oblast určíme průnikem oblastí, které byly určeny jako výsledky těchto  $n$  pokusů. Jestliže při měřeních dosáhneme různých geografických pozic a nedokážeme u nich provést průnik, odhad pozice považujeme za neúspěšný.

## **Efektivita**

Autoři SGP v [12] došli k závěru, že tato metoda je bohužel poněkud nepřesná a dosáhli s ní nevalných úspěchů. Pouze 67% cílů bylo odhadnuto a z toho jen 40% bylo odhadnuto správně. Na základě těchto výsledků se dá říci, že tento způsob se svou 26,8% úspěšností poněkud selhává.

## **4.5 Constraint Based Geolocation (CBG)**

Tato geolokační metoda překonává omezení, které přinášejí metody s diskrétním počtem možností pozice cíle. I u této metody se pro určení pozice cílů používá sada referenčních bodů, avšak metoda CBG byla vytvořena s myšlenkou, že k zaměření cíle bude nejvhodnější použit multilateraci, což je způsob relativně srovnatelný se způsobem využívající technologie GPS [6].

Multilaterace se všeobecně používá jako způsob výpočtu pozice neznámého bodu (cíle), kdy cíl vysílá signál, který se šíří konstantní rychlostí ke třem a více přijímačům se známou polohou (referenční bod). Každý přijímač tento signál zaznamená v jiný okamžik než ostatní přijímače (za předpokladu že jsou od cíle v různých vzdálenostech) a vyhodnocením vzájemných rozdílů těchto odezev se vypočítá pozice cíle [20]. U CBG soustavu referenčních bodů tvoří stanice, které jsou schopny provádět měření zpoždění k cíli T.

Zaměřováním metodou CBG dostáváme výsledek ve formě spojitě oblasti, resp. spojitěho množství pozic, kde je předpokládán výskyt cíle T [10], [12]. Jen v naprosto ideálním případě by nám tato metoda byla schopna pozici cíle určit přesně na konkrétní souřadnice, což bude uvedeno dále.

### **4.5.1 Hranice vzdálenosti (distance constraint)**

Klíčovým prvkem CBG je transformace změřeného zpoždění na vzdálenost, která se co nejvíce podobá té reálné (tedy přímé geografické vzdálenosti mezi dvěma koncovými stanicemi) a tuto vzdálenost použít pro vymezení *hranice vzdálenosti (distance constraint)*, uvnitř které se nachází cíl. Vymezování hranic se provádí pro každý ref. bod zvlášť a to pomocí tzv. kalibrace. Při kalibrování se vychází z poznatku, že signál v optickém vlákne putuje rychlostí téměř přesně dvoutřetinovou rychlostí světla ve vakuu ( $\frac{2}{3} c$ ) [15] a tento poznatek se ještě dál vhodně upravuje.

Postup tvorby „hranice vzdálenosti“ bych rozvrhl do několika celků z důvodu, že samotná tato hranice se skládá z několika částí.

### **Dolní hranice (*lower bound*)**

Začneme pojmem *dolní hranice (lower bound)* – což je vzdálenost odvozena z úvahy, že změřené zpoždění je čistě ideální, tzn. je způsobeno pouze zdržením dat cestou v optickém vlákne, bez započítání jakýchkoliv dalších mimořádných zpoždění, způsobených například směřováním dat, zatížením linky zkrátkou objektem, který na trase tvoří „úzké hrdlo“. Takový převod je v praxi však nepřesný a proto je tato hranice tvořena nejzazší vzdáleností od referenčního bodu (stanice, k níž je prováděno měření, nemůže být od ref. bodů více vzdálená). Jelikož v podstatě neexistují žádné ideální trasy, musíme se na zpoždění dívat tak, že je vždy navýšeno. Souhrn těchto nežádoucích navýšení nazýváme *přídavek zpoždění (Additive distortion)* a vytváří nám nežádoucí zkreslení vzdálenosti (*Additive distance distortion*) [10].

Každý takový nárůst zpoždění, který vzniká nedokonalou metodou měření, vytváří *nežádoucí zkreslení vzdálenosti (Additive distance distortion)*, tzv. *přídavek zpoždění (Additive distortion)*.

### **Horní hranice (*upper bound*)**

Tato hranice je tvořena hodnotou zpoždění očištěného od přídavku zpoždění. Hodnota zpoždění tvořící tuto hranici tedy bude menší, než původní zpoždění (to tvořilo dolní hranici) a v ideálním případě půjde pouze o zpoždění způsobené propagací dat přenosovým médiem. Horní hranice tedy bude menší než dolní hranice a ideálně její hodnota bude shodná s reálnou geografickou vzdáleností mezi referenčním bodem a stanicí, ke které bylo prováděno měření.

Rozdíl mezi dolní a horní hranicí je tedy tvořen nežádoucím zkreslením vzdálenosti, které je úměrné hodnotě přídavku zpoždění [10].

## **4.5.2 Kalibrování**

Zjednodušeně řečeno, pojem kalibrace vyjadřuje převod zpoždění na realitu odpovídající vzdálenost.

Velikost nežádoucího zkreslení je potřeba pro efektivní funkci metody CBG odhadnout a odstranit z tak velké části, jak je to jen možné, což provádíme pomocí kalibrace. CBG navíc disponuje schopností tzv. autokalibrace – automatickým odstraněním nežádoucího zkreslení, čímž reaguje na aktuální stav sítě.

Kalibraci založenou na následujícím postupu provádí každý RB:

- Ref. bod změří zpoždění ke všem ostatním ref. bodům.
- Dostaneme hodnoty zpoždění, ke kterým však dokážeme přiřadit přesnou vzdálenost, a z těchto hodnot vytvoříme graf poměru zpoždění a vzdáleností (na osu x vyneseme vzdálenost a na osu y zpoždění (viz Obr. 4.3).
- Do tohoto grafu vyneseme přímku pojmenovanou baseline – tato přímka odpovídá teoretickým hodnotám poměru zpoždění ke vzdálenosti (podle [10] 1 ms zpoždění je rovna přibližně 100 km trasy). Jedná se o dolní hranici daného ref. bodu, která je vyjádřena směrnicovou rovnicí přímky

$$y = mx + b, \quad (4.2)$$

kde  $m$  je vztaženo k rychlosti putování světla v optickém vlákne,  $b = 0$ , protože neuvažujeme žádné zkreslující zpoždění [10].

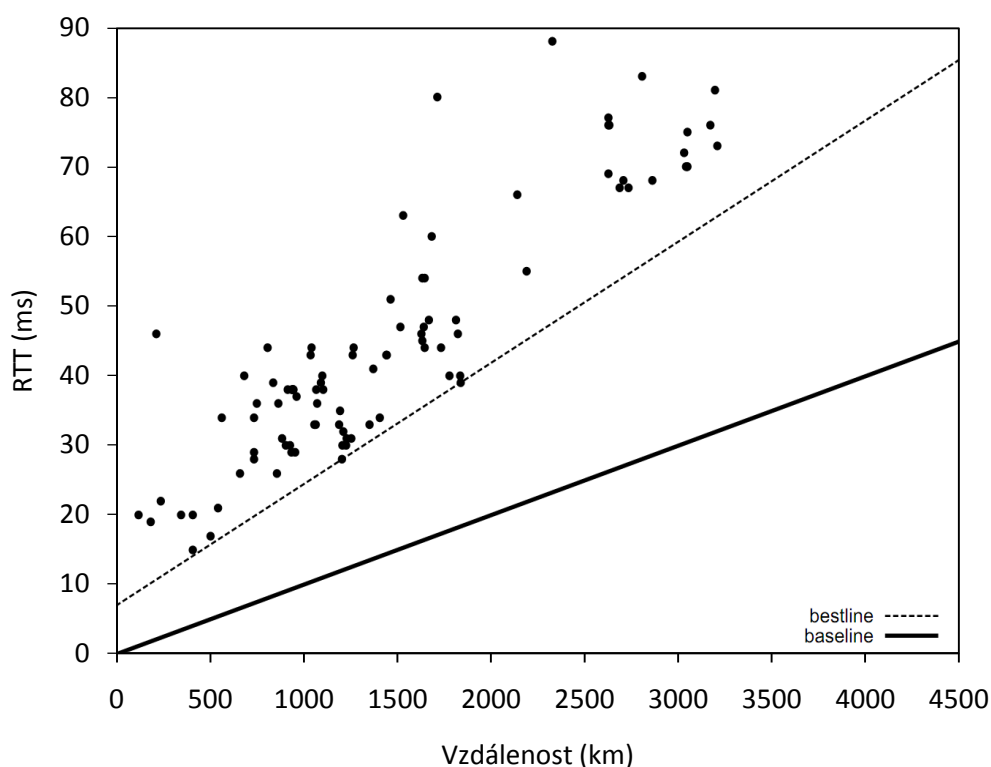
- Z grafu vyčteme nejlepší možný poměr hodnot zpoždění a vzdálenosti akceptovatelný vzhledem ke všem ref. bodům – tento poměr opět vyjádříme směrnicovou rovnicí přímky a nazveme ho bestline. Bestline je definován jako přímka

$$y = m_i x + b_i, \quad (4.3)$$

která je co nejbližší, avšak pod všemi body v grafu, z čehož je následně i odvozena hodnota  $m$ . Hodnota  $b_i$  odráží přítomnost přídatku zpoždění.

Pro lepší představu o vytváření baseline a bestline uvádím graf (viz Obr. 4.1), ze kterého je princip tvorby těchto „hodnot“ dobře čitelný.

Výše zmíněný postup provádí každý referenční bod, čímž zjistí jaký poměr zpoždění a vzdálenosti má uvažovat při určování své (finální) hranice vzdálenosti. Jinak řečeno, pro zjištěné zpoždění vůči cíli určí správný koeficient pro převedení tohoto zpoždění na vzdálenost. Každý ref. bod má tedy svou hranici vzdálenosti odvíjející se od tohoto jedinečného poměru a provádějí-li referenční body své měření v jistých časových intervalech, zjišťují tím, jak se mění jejich bestline v závislosti na stavu sítě (jaká reálná vzdálenost odpovídá aktuálnímu zpoždění v síti). Jinými slovy provádějí autokalibrace, čímž si upřesňují hranice vzdálenosti podle aktuálního stavu sítě.



Obr. 4.3: Graf závislosti vzdálenosti ref. bodů na změřeném zpoždění, používaný pro určování *baseline* a *bestline* (převzato z [10]).

### 4.5.3 Multilaterace

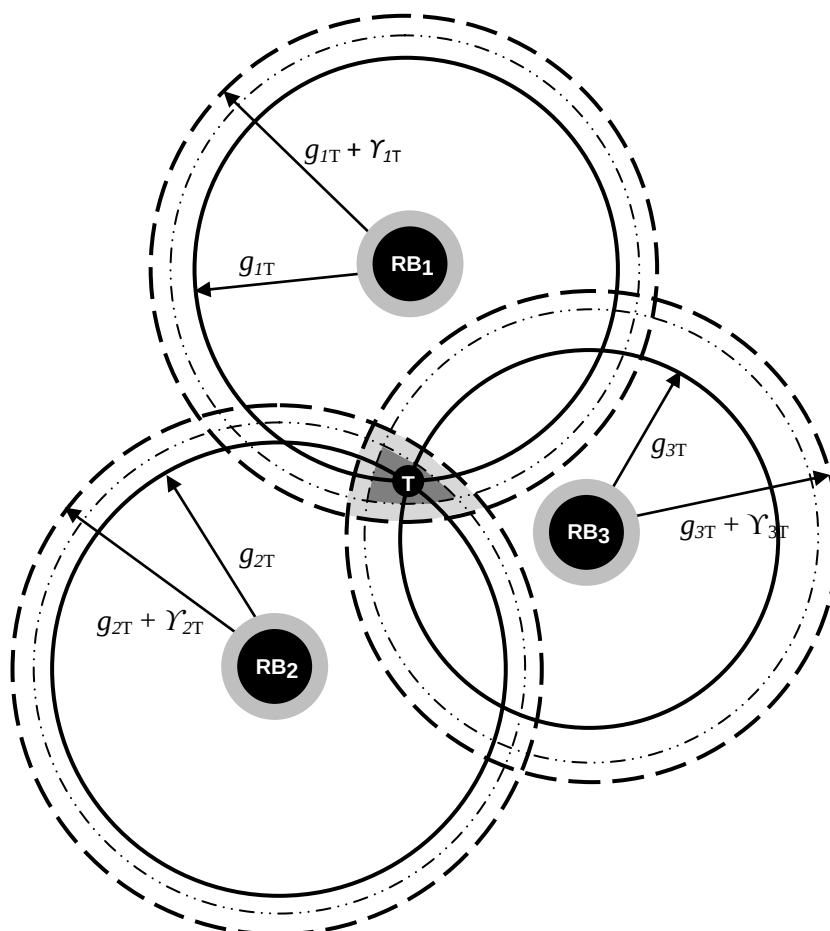
Má-li každý ref. bod stanovenou svou hranici vzdálenosti, přistoupíme k určování polohy cíle pomocí multilaterace. Tento způsob geolokace ilustruji na názorném příkladu, kde popořádku shrnu i doposud zmíněná fakta.

Na obrázku Obr. 4.4 máme multilateraci pomocí 3 ref. bodů (RB1, RB 2, RB 3). Uvažujeme reálnou situaci kdy zpoždění je zkresleno, tudíž bereme v potaz nežádoucí zkreslení. Všechny RB se pomocí vytvoření hranic vzdálenosti snaží odhadnout neznámou pozici cíle T (tj. odpovídající odvozenému *bestline*). Hranice vzdálenosti bude u každého RB vytvořena individuálně (ovšem mohou být i shodné).

Prvně si každý ref. bod vytvoří dolní hranici (odvozenou z *baseline*)  $g_{DiT}$ , kterou v daný okamžik tvoří  $g_{DiT} = g_{UiT} + \gamma_{iT}$ , kde  $g_{UiT}$  odpovídá reálné geografické vzdálenosti (horní hranici) a  $\gamma_{iT}$  je nežádoucí zkreslení vzdálenosti. Pokud bychom určovali pozici cíle jen pomocí horních hranic (vytvořených na základě reálné vzdálenosti, resp. na základě takového zpoždění, které by odpovídalo reálné vzdálenosti), cíl by se nacházel v bodě, kde se tyto hranice kříží. Avšak horní hranice neznáme a jako důsledek nežádoucího zkreslení vzdálenosti se nám výsledek odhadu

pozice cíle rozšiřuje na prostor  $\mathbf{R}$ , který je na obrázku znázorněn jako světle šedá oblast vymezená dolními hranicemi.

Nyní přichází na řadu kalibrace. Díky ní se v ideálním případě (a ve většině případů tomu tak je) vytvoří hranice vzdálenosti (čerchovaná čára) mezi výše zmíněnými hranicemi (dolní hranice – přerušovaná čára a horní hranice – plná čára). Čím přesněji bude určen bestline, tím více se „finální“ hranice vzdálenosti bude oproti dolní hranici přibližovat ke svému středu, referenčnímu bodu (hranice vzdálenosti bude více odpovídat skutečné vzdálenosti, tedy horní hranici), oblast  $\mathbf{R}$  se bude zmenšovat (viz tmavě šedá oblast, Obr. 4.4) a odhad bude přesnější.



Obr. 4.4: Hledání cíle pomocí multilaterace

*V nejideálnějším případě by hranice vzdálenosti odpovídala horní hranici ( $g_{iT}$ ), avšak v reálných případech se setkáváme se třemi typy výsledku stanovení hranice vzdálenosti: nadhodnocením, podhodnocením a současnému nadhodnocení i podhodnocení (mismatch).*

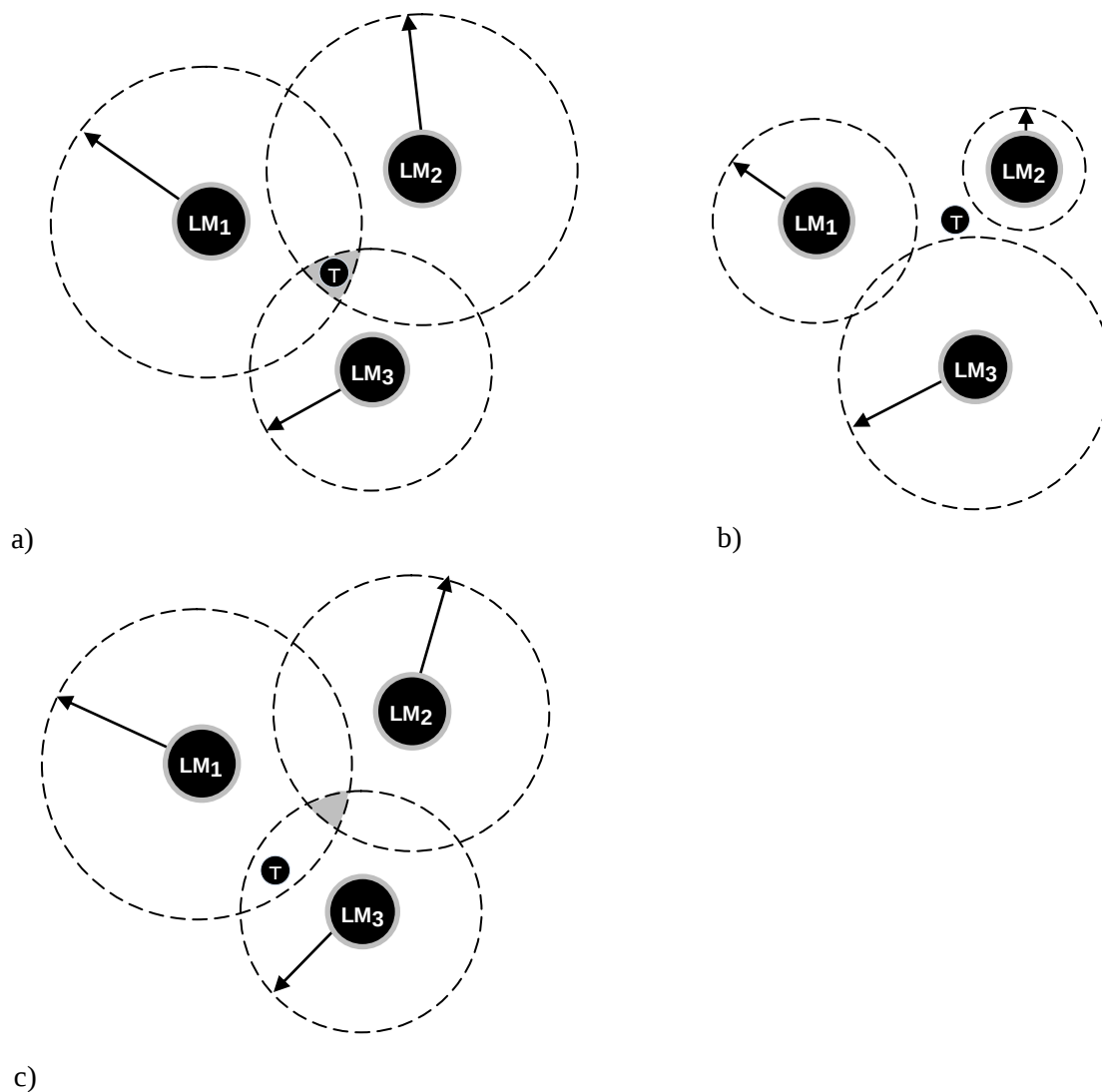
#### 4.5.4 Možné případy při vytváření hranice vzdálenosti

Následují ukázkové situace jednotlivých případů popsanych v [10] a [12], kterým jako doprovod slouží Obr. 4.5.

**Nadhodnocení hranic:** v podstatě se jedná o běžnou praxi. Hranice vzdálenosti zúčastněných ref. bodů jsou poněkud vzdálenější než horní hranice, ale důležité je, že dokážou vytvořit oblast **R**. Pokud nadhodnocení není příliš velké (tím pádem i oblast **R** není příliš velká), CBG pozici cíle odhaduje přesně.

**Podhodnocení hranic:** tento případ je chybným výsledkem kalibrace a jeho důsledkem je nedosažení geolokace cíle. Tato situace nastává, pokud cíl T má lepší vztah mezi zpožděním a vzdáleností vůči jednomu z ref. bodu než oproti bestline, tedy zbytku bodů. Z toho plyne, že referenční body nevytvoří region **R** a prohlásí, že nemohou určit polohu cíle. Toto je důležitá schopnost CBG, protože vždy je lepší pozici cíle neurčit vůbec, než ji naslepo určit naprosto špatně.

**Současné nadhodnocení i podhodnocení (mismatch):** na Obr. 4.5c) je znázorněna situace, kdy  $RB_1$  a  $RB_3$  nadhodnotili své hranice vzdálenosti vůči cíli T, zatímco  $RB_2$  svou hranici vzdálenosti podhodnotil. Z neshody v hranicích vzdálenosti mezi ref. body vyplývá, že oblast **R** nezahrnuje pozici cíle T. Naštěstí, tento stav nebývá příliš častý.



Obr. 4.5: Ukázkové případy při vytváření hranic vzdálenosti: a) nadhodnocení, b) podhodnocení  
c) mismatch

#### 4.5.5 Efektivita

Co se týče přesnosti, autoři publikace [16] při aplikaci CBG dospěli k takovýmto výsledkům – při testování za použití datasetu stanic ze Spojených států byl medián chyby odhadu pozice cíle v 80 % případů pod 100 km, respektive pod 25 km pro západoevropský dataset. Oproti tomu, autoři na stejném datasetu použitím metod na bázi GeoPing dosáhli takových mediánů chyby: 150 km pro americký dataset a 100 km při měření v západoevropském datasetu.

## 4.6 Speed of Light (SOI)

SOI je velmi podobná metodě CBG. Je odvozena z jejího základního principu – určování polohy pozice pomocí referenčních bodů ve smyslu multilaterace pomocí jejich hranic vzdálenosti. Nová myšlenka, se kterou SOI přichází, spočívá v tom, že vzdálenost mezi stanicemi je v drtivé většině případů menší, než by se dalo odvodit ze zpoždění, resp. z rychlosti šíření světla ve vakuu [16]. Jak již bylo zmíněno, data v optickém vlákne cestují  $\frac{2}{3}$  rychlostí světla. Kvůli zpoždění, které u CBG způsobovalo nežádoucí zkreslení (snižovalo efektivitu přenosu na trase) je výsledné zpoždění mezi stanicemi zvětšeno. Z měření, která provedli autoři [16] vyplynulo, že zpoždění mezi stanicemi odpovídá takové vzdálenosti, jakoby data cestovala rychlostí  $\frac{4}{9}$  rychlosti světla. Tento jev se autorům potvrdil v 80 % měření a v 20 % měření došli k závěru, že by mohl být použit faktor  $\frac{2}{9} c$ .

Toto zjištění je využito k přepočtu mezi zpožděním a vzdáleností a následně jednoduchému vytvoření hranice vzdálenosti u každého ref bodu. Metoda SOI tedy vytváří hranice vzdálenosti použitím faktoru  $\frac{4}{9} c$  k přepočtu mezi zpožděním a vzdáleností, což je mnohem jednodušší způsob, oproti tomu který používá metoda CBG. Odpadá také potřeba měření odezvy mezi jednotlivými ref. body, což je jistě také výhoda.

Myšlenka takového přepočtu se dá samozřejmě uplatnit u dalších odvětví, jejichž činnost je založena na měření zpoždění, respektive na přepočtu hodnoty latence na vzdálenost.

### 4.6.1 Podhodnocení hranic

I u metody SOI hrozí podhodnocení hranic. To obdobně jako u CBG znamená, že se hranice mohou setkat buď na „nesprávném místě“ a vytvořit tak oblast, ve které se cíl nenachází, anebo se nemusí setkat vůbec. V publikaci [16] je uvedeno, že při experimentu toto podhodnocení hranic při použití metody CBG zapříčinilo špatný odhad 27 cílů ze 128 a při aplikování myšlenky SOI to bylo příčinou pouze u 3 cílů.

Bohužel toto podhodnocení se napravuje nesnadno. Při podhodnocení hranice, kdy se nevytvoří žádná oblast průniku **R**, se v podstatě nedá určit, u kterého ref. bodu jsme hranice podhodnotili. Možná náprava (i když ne moc užitečná) se nabízí ve formě vytvoření nové hranice, s použitím faktoru  $\frac{2}{3} c$ . Tento způsob nápravy není vhodné aplikovat v situaci, kdy se region **R** vytvořil, avšak „pouze“ neobsahuje cíl. Výsledek bude pravděpodobně velmi nepřesný (např. hranice dvou ze tří ref. bodů by byly určeny správně a přesto by byly zvětšeny, čímž by došlo k efektu rozšiřování plochy prostoru **R**).

#### **4.6.2 Efektivita**

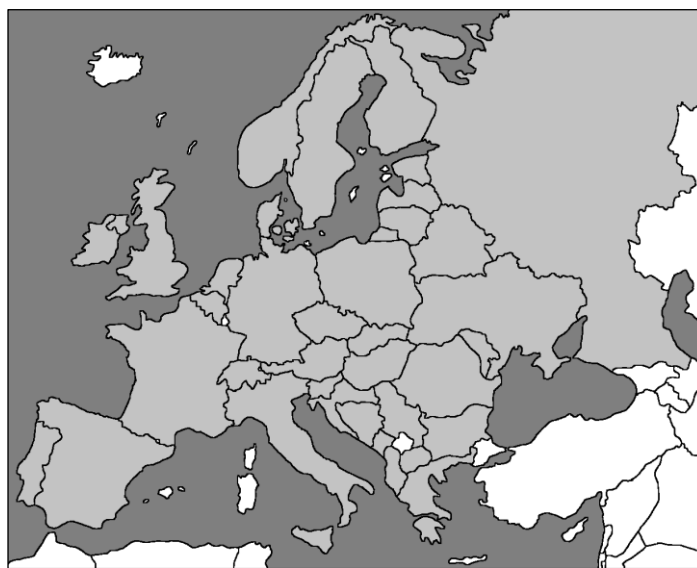
Účinnost této metody potvrzuje, jak je koeficient pro převod zpoždění na vzdálenost přesný. Aniž by technika SOI byla zatížena nutnostmi, které vyžaduje CBG (složitého vytváření bestline, kalibrace RB), tak dosahuje v podstatě stejných výsledků jako CBG [16].

## 5 MĚŘENÍ POMOCÍ METODY GEOPING

K experimentálnímu měření jsem podle zadání využil metodu GeoPing. Způsob, kterým se dopracovává ke zjištění pozice cíle, není triviální ale pro úvod do problematiky geolokace je tato technika dostatečně transparentní. Také si od vybrané techniky slibuji, že podklady vytvořené pro měření touto metodou budou moci být použity i při použití jiných metod. Dále bych rád zmínil, že technika GeoPing ze své podstaty nepatří mezi nejpřesnější, avšak z jejího otestování mohou vyplynout nejen u této metody použitelné poznatky.

### 5.1 Geografické umístění

Celá má práce je z geografického hlediska vymezena hranicemi většiny evropských zemí (světle šedé oblasti na Obr. 5.1). Tento prostor zahrnuje 36 států a síťová infrastruktura ve většině těchto míst může být považována za vyspělou [29], tzn. s dostatkem vysokorychlostních spojů s dobrou šířkou pásma a sítí ISP s dobře rozprostřenými PoP (Point of Presence). Z toho usuzuji, že k účelu zjišťování neznáme pozice stanice v Internetu pomocí metody Geoping, je tato oblast vyhovující.



Obr. 5.1: Mapa území, na kterém je prováděno měření (světle šedé oblasti).

## 5.2 PlanetLab

K praktické části této bakalářské práce jsem využil služeb experimentální vědecké sítě PlanetLab. Pomocí stanic v této síti jsem mohl provést měření, která se zakládají na reálných skutečnostech, omezeních a podmínkách.

PlanetLab je stabilní a dobře udržovanou globální výzkumnou sítí, kterou tvoří seskupení univerzit a firem z odvětví informačních technologií za účelem podpory výzkumu síťových aplikací a experimentů, které mají pracovat v globálním měřítku a k jejichž vývoji je třeba testování v reálných podmínkách, na které můžeme narazit při používání Internetu. PlanetLab byl použit při vývoji a testování nových technologií v odvětví distribuovaného uchovávání dat, P2P systémů, mapování sítě a dalších síťových aplikací [21].

PlanetLab podle svých internetových stránek poskytoval k únoru tohoto roku přístup k 1064 stanicím (*node* – dedikovaný server, na kterém běží prováděné aplikace), spadající pod 550 organizací (*site* – de facto „sídlo“ dané organizace, která nabízí k dispozici své stanice) a to všem svým členům. Česká republika je v síti PlanetLab zastoupena připojením 3 stanic sdružených pod organizaci Cesnet.

### Měření v síti PlanetLab

Z mého pohledu využitelnost sítě PlanetLab spočívá v umožnění přístupu ke vzdáleným stanicím, které se nacházejí po celém světě. K dosažení vzdáleného přístupu je nutné provést tyto kroky (jako vydatná pomoc mi posloužil návod [22]):

- registrace v systému PlanetLab (uživatel musí být pod záštitou členské společnosti – university),
- vytvoření SSH klíče a následný upload veřejné části klíče do databáze PlanetLab, odkud se rozšíří na stanice spadající pod projekt uživatele
- stanice, ke které požadujeme přístup, se musí nacházet v konkrétním projektu (*slice*) – projekty obsahují určitý počet stanic (kdykoliv si do projektu můžeme přidat požadovanou stanici) a tvoří tak pracovní skupinu, ve které se mohou šířit testované aplikace.

Název projektu je pak využíván zejména jako přihlašovací jméno (*login*) na vzdálených stanicích. Osobně jsem k vytvoření SSH klíče a následovnému připojování ke vzdáleným stanicím použil program Putty [24]. Moje práce spadá do projektu *cesnet\_s2*, ke kterému bylo potřeba přidat i několik vhodných stanic.

## 5.3 Vytváření mapy zpoždění

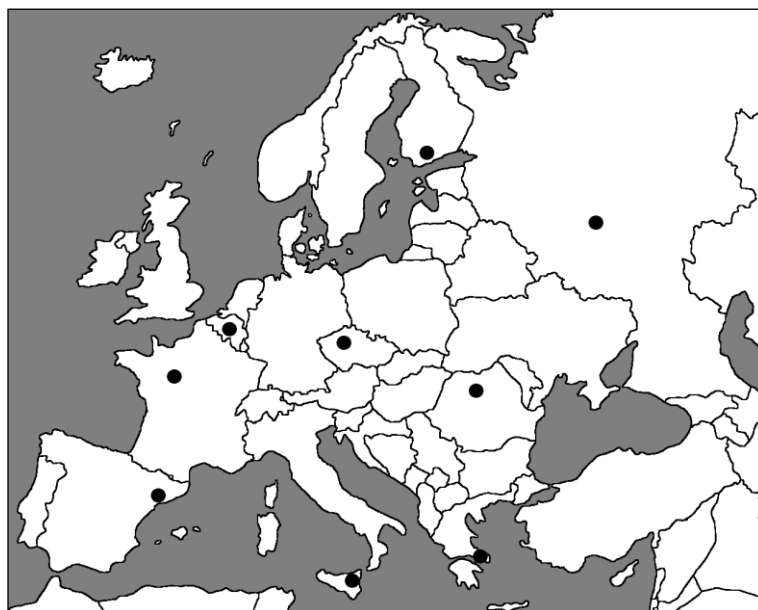
Celé měření je založeno na datasetu, který tvoří 9 sond a 200 referenčních bodů. Dále, aby technika GeoPing mohla být řádně prověřena v praxi, bylo potřeba najít stanice, které zastávají roli cílů – těch je použito 60.

### 5.3.1 Sondy

Pro měření RTT jsem použil stanice výhradně ze sítě PlanetLab. Podařilo se mi najít množství použitelných terminálů spadajících pod evropskou oblast, u kterých jsem se rozhodoval o jejich použití. Vhodnost stanic byla posuzována podle toho, zda splňovala tyto požadavky:

- Ke stanici musí být povolen vzdálený přístup a musí být on-line (určité procento stanic prochází údržbou, či je naopak nedostatečnou údržbou mimo provoz).
- Stanice musí být tímto způsobem také dosažitelná (u několika terminálů se nezdařilo připojení z důvodů „connection timeout“ (podle [22] se pravděpodobně jedná o chybné nastavení operačního systému Windows, ze kterého byl prováděn přístup), odmítnutí autentizačního klíče nebo nekompatibility autentizačních metod terminálu s programem Putty).
- Stanice musí být schopna provádět základní typy měření RTT (ping či traceroute) „ven ze své sítě“ do Internetu.
- Stanice musí mít vhodnou geografickou polohu vzhledem k ostatním sondám, to znamená, že musí být mezi sebou rozumně vzdáleny, aby vytvářely rozprostřenou síť a ne shluky – z tohoto důvodu je třeba znát jednotlivé polohy stanic.
- Stanice by neměla spadat pod společnou organizaci (*site*) s jinou stanicí, využívanou pro stejný účel (a to z důvodu, že u většiny terminálů spadajících pod stejnou organizaci je v záznamech PlanetLab-u či WhoIs databázi [26] uváděna stejná geografická poloha, z čehož plyne, že takto geograficky shodně umístěné terminály by nesplňovaly podmínku o geografickém rozprostření).

Použito tedy bylo 9 stanic umístěných na universitách v 9 státech, které těmto požadavkům vyhovovaly. Jejich rozptýlení můžeme vidět na Obr. 5.2 a je následující: Namur (BEL), Praha (CZE), Barcelona (ESP), Helsinky (FIN), Rennes (FRA), Atény (GRC), Catania (ITA), Turgu Mures (ROM) a Moskva (RUS).



Obr. 5.2: Mapa znázorňující rozmístění sond

### 5.3.2 Referenční body

Roli ref. bodů zastává 200 stanic umístěných v 36 zemích (viz Obr. 5.3). Jsou zastoupena téměř všechna hlavní města států, na jejichž území provádím měření.



Obr. 5.3: Mapa znázorňující rozmístění referenčních bodů

Na každou stanici sloužící jako referenční bod jsou kladeny tyto požadavky:

- Stanice musí být schopná odpovídat na dotaz „Echo Request“ protokolu ICMP.

- Poloha ref. bodu musí být důvěryhodně zjištělná.
- Z hlediska geografické polohy by většina stanic vůči ostatním referenčním bodům měla mít rozumný odstup (nemělo by docházet k vytváření shluků referenčních bodů).

Při výběru podle těchto kritérií byly nejčastěji jako vhodné ref. body voleny různé univerzitní servery, servery poskytovatelů internetu ve větších městech, či servery, na kterých hostují internetové stránky vybraných měst, nebo stránky, které jsou s vybranými městy nějak spjaté, a samozřejmě i několik uzlů ze sítě PlanetLab.

Některé referenční body byly záměrně vybrány tak, že se jejich poloha shoduje s pozicí, kde se vyskytuje i nějaký cíl. To z důvodu, aby GeoPing ukázal, jestli to je pro takový cíl výhodou a jeho poloha bude určena k onomu geograficky blízkému ref. bodu, anebo i přes tento fakt bude jeho poloha určena ke geograficky vzdálenějšímu bodu.

### 5.3.3 Zjišťování polohy stanic

Každá položka v mapě zpoždění musí obsahovat kromě změřených zpoždění i souřadnice uzlu, ke kterému se položka vztahuje. K získávání těchto informací bylo použito databáze WhoIs [4] a proto se nedá definitivně říci, že jsou tyto informace skutečně pravdivé.

Pozice stanic v mapě zpoždění, která je vytvořena pro tuto bakalářskou práci, je uvedena ve dvou formátech – ve formátu souřadnic a ve formátu „název města“. K extrakci souřadnic daných referenčních bodů jsem využil internetovou aplikaci GeoLocator [8].

### Získání informací o poloze sond

Každá sonda je tvořena uzlem v síti PlanetLab, který náleží jisté organizaci (*site*). U každé *site* je kromě jiných informací zveřejněna i geografická poloha v podobě geografických souřadnic Světového geodetického systému 1984. Tyto informace jsou vkládány správcem dané *site*, nicméně je možné, že pokud pod danou organizaci spadá více uzlů, tak všechny tyto uzly nemusí geografickou polohou odpovídat sídlu této organizace. K uzlům samotným bohužel informace o jejich poloze nejsou uvedeny. Z toho důvodu byly internetové adresy uzlů prověřeny ve WhoIs databázi pomocí aplikace DomainTools [4]. Zde se s trochou štěstí daly najít informace o geografické poloze, o kterých se domnívám, že jsou zadány pravdivě – jejich přesnost je na úrovni měst či dokonce ulic. Ve všech případech kontroly jsem našel shodu informací z databáze WhoIs s informacemi uvedenými v síti PlanetLab.

## **Získání informací o poloze referenčních bodů a cílů**

Zde byl úkol už poněkud složitější a výsledky musí být bohužel považovány za méně důvěryhodné. U ref. bodů a cílů, které byly vybrány ze sítě PlanetLab jsem postupoval stejně jako při zjišťování umístění sond. V naprosté většině srovnání jsem mezi záznamy z PlanetLab a WhoIs databáze dospěl ke stejným informacím. Pouze u jediného uzlu (planetlab1.fit.vutbr.cz) se záznamy neshodovaly – jedná se o uzel spadající pod *site* Cesnet, kde jsou udány souřadnice odpovídající Praze. Tento uzel se, jak už sám název napovídá, nachází v Brně.

Ostatní stanice pochází z „civilního“ sektoru, kde mi jako bohatý zdroj IP adres přiřazených ke geografické poloze posloužil vyhledavač internetových herních serverů [7]. Informace vztahující se k poloze těchto uzlů jsem hledal pomocí dotazování se do WhoIs databáze, kde byly k nalezení pouze ve slovní formě a s různou přesností (stát, město, ulice). Přesnost takovýchto záznamů je samozřejmě zpochybnitelná a pravděpodobnost mylného či zavádějícího záznamu je oproti záznamům dosažitelných ze stránek Planetlab poněkud vyšší.

### **5.3.4 Měření RTT**

K měření zpoždění byl podle zadání vytvořen bash skript (za vydatné pomoci tutoriálů a návodů [17] a [27]), který je uveden a podrobně popsán na přiloženém DVD v příloze A. Ve stručnosti uvedu, že skript je koncipován tak, že z daného souboru načte seznam adres uzlů, na které následně provede ping. Výsledky změřených zpoždění uloží do textového souboru, odkud poté proběhla jejich extrakce a zpracování do mapy zpoždění.

V kapitole 2.3 je uvedeno, že při měření je třeba získat 10–15 vzorků. Obecně se dá říct, že při tomto počtu opakování dospějeme na dané trase k nejnižší možné hodnotě zpoždění. Avšak ani tato zásada nezaručí, že bude dosaženo nejmenší možné hodnoty zpoždění mezi danými stanicemi. V jistých případech, kdy odhad pozice cíle byl velmi nepřesný, bylo měření zopakováno. Některá tato opakování vedla k překvapivým výsledkům – hodnoty nově naměřených zpoždění byly jak menší tak i větší. Větší hodnoty, jež byly získány přeměřením, jsou ignorovány (z důvodu, že nejvýhodnější cesta odpovídá nejmenší latenci) a lepší hodnoty (s rozdílem až 8 ms) nahrazují ony původní. U žádného z cílů toto opakované měření nemělo za následek zvětšení chyby odhadu pozice a u následujících cílů bylo dosaženo zmenšení chyby: Brno, Olomouc, Paříž, Leeds, Hanover, St. Gallen.

### **5.3.5 Zpracování výsledků**

Změřené hodnoty jsem zpracoval pomocí programu Microsoft Office Excel 2007. Veškeré záznamy lze najít v souboru Mapa\_zpozdeni.xlsx, který tvoří přílohu A. Databáze změřených

hodnot RTT (která je nejpodstatnější částí mapy zpoždění) obsahuje 200×9 záznamů jen pro referenční body a dalších 60×9 záznamů pro cíle. U všech referenčních bodů a cílů je uvedena jejich poloha ve formě stát, město a geografické souřadnice.

Použitím funkcí programu Excel bylo zautomatizováno vypočítání vektoru **DV** všech ref. bodů a vektoru **DV'** každého cíle. Výsledné vyhledávání nejshodnějšího **DV** s **DV'** je samozřejmě také automatizováno. Lokace, kde se nachází ref. bod odpovídající vybranému **DV**, je také uvedena automaticky.

Pro měření chyby, tj. vzdálenosti mezi opravdovou polohou cíle a polohou zjištěnou pomocí měření, jsem použil nástroj Google Maps Distance Calculator<sup>5</sup> [5]. Tento nástroj výslednou vzdálenost mezi zadanými body udává v podobě „vzdušné čáry“. Řečeno jinými slovy, při výpočtu je aplikována sférická trigonometrie (měření na kulové ploše).

Výpis velikosti chyby (v km) vzniklé při určování cíle a uvedení geograficky nejbližšího RB pro konkrétní cíl (s jejich vzájemnou vzdáleností) zautomatizováno není; všechny vzdálenosti jsou zaokrouhleny na celých 5 km.

Následující tabulka je výtahem toho nejdůležitějšího ze souboru Mapa\_zpozdeni.xlsx a shrnuje výsledky mého experimentálního měření – výsledky určování polohy neznámých stanic v síti Internet:

Tab. 5.1: Výsledky měření

| Skutečná pozice cíle | Pozice přisouzena k ref. bodu | Chyba odhadu (km) | Geograficky nejbližší ref. bod | Vzdálenost mezi geograficky nejbližším LM a cílem (km) |
|----------------------|-------------------------------|-------------------|--------------------------------|--------------------------------------------------------|
| Graz                 | Krakow                        | 305               | Maribor                        | 60                                                     |
| Leibnitz             | Podgorica                     | 565               | Maribor                        | 25                                                     |
| Mechelen             | Fareham                       | 395               | Brugy                          | 90                                                     |
| Nyiregyhaza          | Jesenice                      | 600               | Košice                         | 95                                                     |
| Modrica              | Podgorica                     | 290               | Tuzla                          | 55                                                     |
| Brno                 | Magdeburg                     | 295               | Brno                           | 0                                                      |
| Olomouc              | Olomouc                       | 0                 | Olomouc                        | 0                                                      |
| Skanderborg          | Esbjerg                       | 110               | Århus                          | 25                                                     |
| Zaragoza             | Madrid                        | 270               | Pau                            | 185                                                    |
| Maardu               | Kristiansand                  | 985               | Talin                          | 10                                                     |
| Kotka                | Varberg                       | 920               | Laahti                         | 90                                                     |

<sup>5</sup> Autoři tohoto nástroje za jeho přesnost neručí, avšak podle dosažených výsledků (na 400 m dlouhé běžecké dráze byla chyba měření 6 metrů) ho považují vzhledem k potřebě přesnosti a hlavně k následnému zaokrouhlování vzdálenosti za dostatečně přesný.

| Skutečná pozice cíle | Pozice přisouzena k ref. bodu | Chyba odhadu (km) | Geograficky nejbližší ref. bod | Vzdálenost mezi geograficky nejbližším LM a cílem (km) |
|----------------------|-------------------------------|-------------------|--------------------------------|--------------------------------------------------------|
| Espoo                | Lidköping                     | 675               | Helsinki                       | 15                                                     |
| Paříž                | Orleans                       | 110               | Paříž                          | 0                                                      |
| Grenoble             | Nice                          | 205               | Lyon                           | 70                                                     |
| Marseille            | Bologna                       | 495               | Valbonne                       | 135                                                    |
| Vern                 | Limoges                       | <b>205</b>        | Limoges                        | <b>205</b>                                             |
| Roubaix              | Bielefeld                     | 400               | Brugy                          | 60                                                     |
| Camberley            | Leeds                         | 95                | Londýn                         | 45                                                     |
| Glasgow              | Liverpool                     | <b>285</b>        | Liverpool                      | <b>285</b>                                             |
| Manchester           | Bielefeld                     | 740               | Liverpool                      | 50                                                     |
| Maidenhead           | Liverpool                     | 260               | London                         | 40                                                     |
| Worcester            | Groningen                     | 600               | Birmigham                      | 40                                                     |
| Höhenkirchen         | Štětín                        | 390               | Mnichov                        | 20                                                     |
| Drážd'any            | Hanover                       | 195               | Praha                          | 115                                                    |
| Bremen               | Magdeburg                     | 215               | Hamburg                        | 95                                                     |
| Gunzenhausen         | Mönchengladbach               | 380               | Norimberg                      | 45                                                     |
| Berlin               | Hanover                       | 250               | Berlin                         | 0                                                      |
| Mnichov              | Essen                         | 490               | Mnichov                        | 0                                                      |
| Soluň                | Patras                        | 285               | Thermi                         | 15                                                     |
| St. Gallen           | Zurich                        | <b>150</b>        | Ženeva                         | <b>150</b>                                             |
| Lausanne             | Ženeva                        | <b>30</b>         | Ženeva                         | <b>30</b>                                              |
| Schindellegi         | Jesenice                      | 415               | Bern                           | 100                                                    |
| Limerick             | Dublin                        | 175               | Galway                         | 70                                                     |
| Waterford            | Plymouth                      | 300               | Cork                           | 100                                                    |
| Galway               | Cork                          | 155               | Galway                         | 0                                                      |
| Neapol               | Chieti                        | 165               | Casoria                        | 10                                                     |
| Rimini (San Marino)  | Řím                           | 240               | Ravenna                        | 70                                                     |
| Lecce                | Chieti                        | 400               | Bari                           | 140                                                    |
| Tukums               | Rēzekne                       | 260               | Riga                           | 55                                                     |
| Amersfoort           | Groningen                     | 140               | Utrecht                        | 20                                                     |
| Drammen              | Oslo                          | <b>35</b>         | Oslo                           | <b>35</b>                                              |
| Wroclaw              | Štětín                        | 305               | Poznaň                         | 145                                                    |
| Lodž                 | Krakow                        | 195               | Varšawa                        | 120                                                    |
| Żywiec               | Klaipeda                      | 680               | Krakow                         | 65                                                     |
| Torres Vedras        | Řím                           | 1865              | Lisabon                        | 40                                                     |
| Faro                 | Cromarty                      | 1715              | Faro                           | 0                                                      |

| Skutečná pozice cíle | Pozice přisouzena k ref. bodu | Chyba odhadu (km) | Geograficky nejbližší ref. bod | Vzdálenost mezi geograficky nejbližším LM a cílem (km) |
|----------------------|-------------------------------|-------------------|--------------------------------|--------------------------------------------------------|
| Linda-a-Velha        | Lisabon                       | 10                | Lisabon                        | 10                                                     |
| Bryansk              | Novgorod                      | 610               | Kaluga                         | 185                                                    |
| Jarfalla             | Stockholm                     | 10                | Stockholm                      | 10                                                     |
| Mölnadal             | Oslo                          | 260               | Gothenburg                     | 5                                                      |
| Karlstad             | Gothenburg                    | 205               | Örebro                         | 95                                                     |
| Norrköping           | Oslo                          | 340               | Örebro                         | 90                                                     |
| Sollentuna           | Örebro                        | 155               | Stockholm                      | 15                                                     |
| Haninge              | Lidköping                     | 295               | Stockholm                      | 20                                                     |
| Karlskrona           | Varberg                       | 230               | Kalmar                         | 75                                                     |
| Stenungsund          | Oslo                          | 210               | Gothenburg                     | 40                                                     |
| Dněpropetrovsk       | Odessa                        | 245               | Zaporozhye                     | 70                                                     |
| Chernigov            | Visaginas                     | 555               | Homyel                         | 105                                                    |
| Ternopil             | Visaginas                     | 675               | Khmelnitskiy                   | 105                                                    |
| Bělehrad             | Záhřeb                        | 365               | Bělehrad                       | 0                                                      |
| Průměrná hodnota:    |                               | 373,3             |                                | 66,2                                                   |
| Medián:              |                               | 285               |                                | 52,5                                                   |
| Minimální odnota:    |                               | 0                 |                                | 0                                                      |
| Maximální odnota:    |                               | 1865              |                                | 285                                                    |

Nejen to, co tyto výsledky vypovídají o měření, respektive o metodě GeoPing, diskutují a shrnují v následující kapitole.

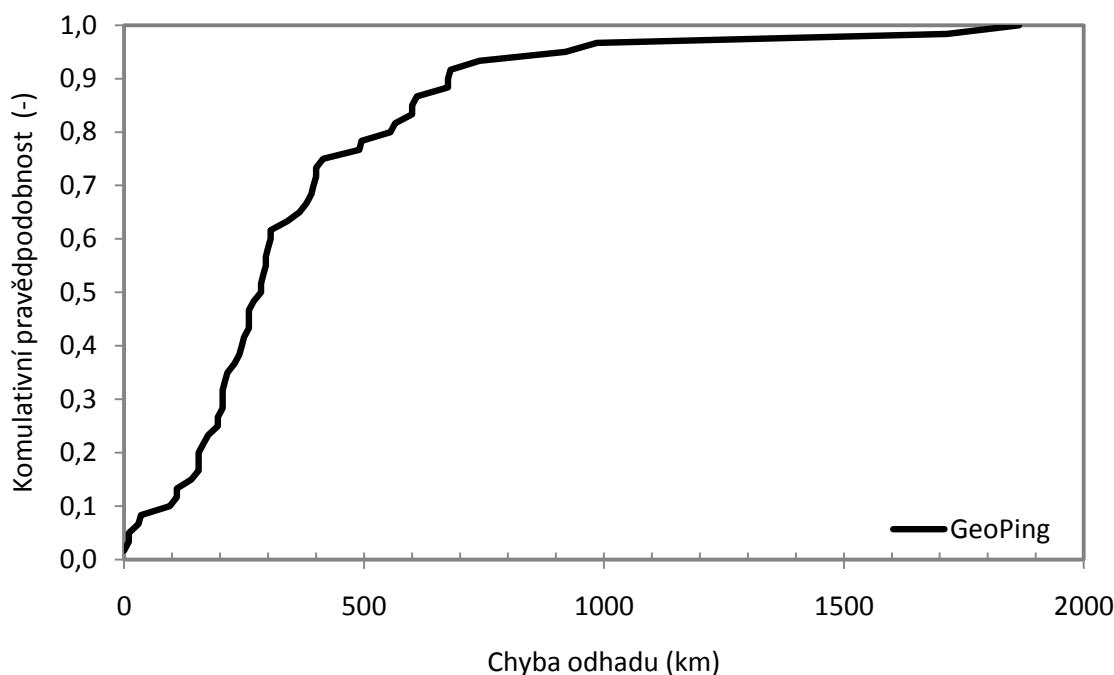
## 6 ROZBOR A DISKUZE ZPRACOVANÝCH VÝSLEDKŮ

### 6.1 Efektivita

Z tabulky Tab. 5.1 je patrná hrubá představa, jakého úspěchu v geolokaci bylo pomocí metody GeoPing dosaženo. Netroufám si definovat rozmezí, do kterého by se měla chyba určení pozice cíle vměstnat, aby se jednalo o dostatečně přesný odhad. Důvody určování totiž mohou být různé a různé jsou poté i nároky na přesnost. Efektivitu tedy sumarizuji podobným způsobem, jakým to provedli samotní autoři metody GeoPing. Výsledná efektivita měření v datasetu o 200 referenčních bodech a při určování 60 cílů je následující:

- 20% cílů byla určena geografická pozice s chybou přibližně 150 km a méně,
- 72% cílů bylo určeno s chybou 400 km a méně,
- 90% cílů bylo určeno s odchylkou menší než 750 km,
- pouze u 3% stanic bylo určení polohy provedeno s nepřesností 1350 km a vyšší.

Tuto efektivitu vystihuje i následující graf:



Obr. 6.1: Graf výsledné efektivity určení pozice neznámé stanice metodou GeoPing

Průměrná hodnota chyby odhadu je 373,3 km, medián této chyby je 285 km. Určení pozice cíle je nejpřesněji provedeno s chybou 0 km (pozice cíle byla určena k místu, kde se cíl opravdu nacházel) a naopak největší chyby bylo dosaženo při určení cíle na lokaci vzdálenou 1865 km od jeho opravdového umístění.

Brát v úvahu přesnost určení polohy cíle ve vztahu k území států není příliš objektivní. Je to z důvodu, že GeoPing se snaží najít nejvhodnější pozici cíle jeho přiřazením ke geograficky nejbližšímu RB (což odvozuje ze zpoždění), takže např. neznámá stanice v blízkosti hranic států může mít nejbližší ref. bod ve státě vedlejším (což se samozřejmě stalo i v tomto experimentálním měření) a i takové přiřazení je považováno za správné.

Nicméně, i přesto uvádím přesnost odhadu tohoto typu; v 56 % případů (33 cílů) byla pozice cíle určena do správného státu

Úspěchů na nejvyšší úrovni (neznámá pozice stanice byla přiřazena z geografického hlediska nejbližšímu referenčnímu bodu) se povedlo provést z necelých 14% (8 cílů).

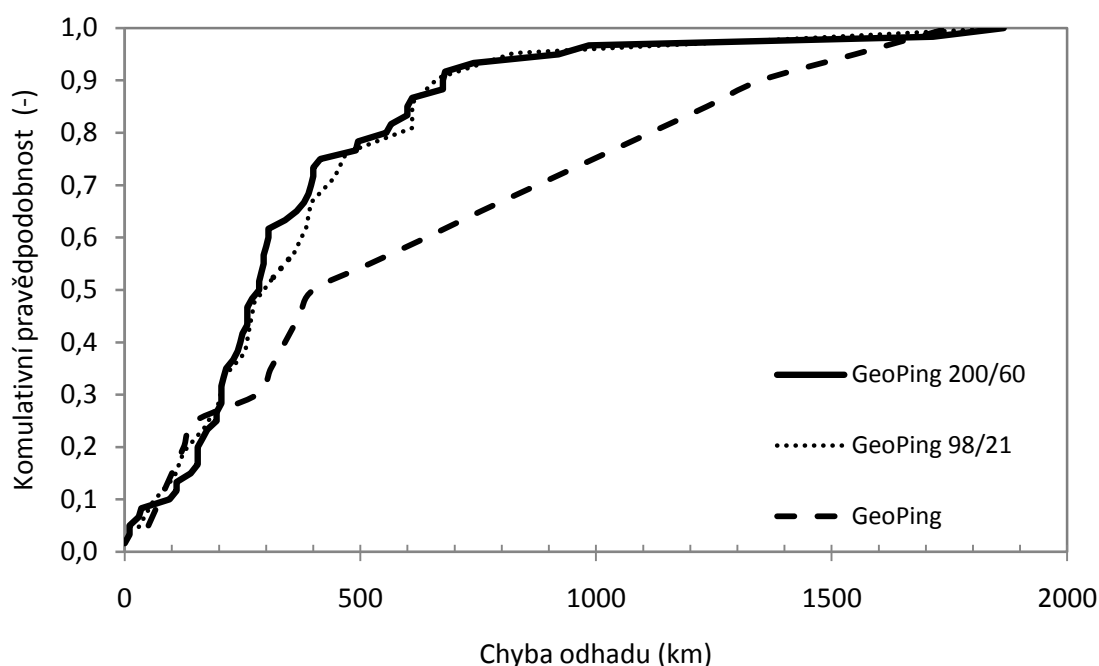
## 6.2 Porovnání výsledků

Pokud porovnáím výsledky mé bakalářské práce s výsledky, ke kterým dospěli autoři metody GeoPing, trůfám si říci, že jsem při experimentu dospěl přinejmenším ke stejně přesnému odhadu cílů. Nějaké hlubší srovnání s jejich prací není bohužel možné, protože jejich výsledky jsou prezentovány celkem stroze. Srovnání s výsledky, kterých dosáhli autoři CBG (viz kapitola 4.5.5) neuvádím z důvodu zavádějící formulace způsobu dosažení jejich výsledků. Podle jejich slov, bylo měření provedeno „GeoPing-like“ metodou, pod čímž si můžeme představit například použití metody Shortest Ping nebo její kombinaci s metodou GeoPing.

Mnohem lépe mohu tuto práci srovnávat s výsledky, kterých jsem dosáhl v mé semestrální práci, kdy bylo použito 98 referenčních bodů a určováno bylo 21 cílů. Přiznám se, že oproti semestrální práci jsem si od použití aktuálního (v podstatě dvojnásobného) datasetu sliboval zlepšení na vyšší úrovni, nicméně i tak došlo k zajímavému pokroku – tehdy byl proveden odhad pozice s průměrem chyby 421 km a mediánem 315 km, což činí rozdíl 48 km respektive 30 km (dodávám, že při tomto měření byla z důvodu nefunkčnosti vyměněna sonda z Budapešti (HUN) za sondu v Targu Mures (ROM)).

Níže je pro lepší představu o rozdílech mezi měřeními uveden graf (Obr. 6.2) porovnávající výsledky z této práce (*GeoPing 200/60* – dataset 200 RB, odhad 60 cílů) s výsledky z mé semestrální práce (*GeoPing 98/21* – dataset 98 RB, odhad 21 cílů) a výsledky autorů metody GeoPing (*GeoPing*), které jsou však z důvodu absence přesných hodnot pro účel tohoto porovnání odvozeny z procentuální úspěšnosti (viz kapitola 4.2.6). V grafu vidíme, že rozdíl mezi výsledky

odhadu pozice za použití datasetu 98/21 a 200/60 jsou relativně malé a odpovídající rozdílům v průměru a mediánu chyby odhadu. Tento závěr mě přivádí k úvahám, které rozvádím dále.



Obr. 6.2: Graf porovnávající výsledky tří různých měření metodou GeoPing

### 6.3 Rozbor výsledků

Při zamyšlení se nad vzniklou chybou určení pozice je nezbytné si uvědomit, co všechno ji způsobuje: do úvahy na první pohled můžeme brát rozprostření referenčních bodů. Ovšem zaměříme-li se kupříkladu na rozložení RB na území severských států Finska a Švédska, zjistíme, že mají hustou síť referenčních bodů a přesto cíle nacházející se zde s velkou přesností odhadnuty nebyly. Další hustě pokryté státy jako Německo a Itálie tento trend také potvrzují, resp. jen minimum států tento trend vyvrací. Nicméně, zvýšením počtu RB bylo v celkovém měřítku dosaženo zlepšení, bohužel v poměru, který není moc příznivý – dvojnásobný dataset přinesl zlepšení přibližně o 10 %.

Druhým činitelem je zajisté dostupnost RB a cílů z hlediska internetového spojení. Dataset vytvořený v této práci obsahuje vysoké procento dobře dostupných stanic, ovšem nelze vyloučit, že na cestě dlouhé například přes celou Evropu v okamžiku měření nedošlo k nějakému výpadku spojení na nejkratší možné trase nebo nevhodnému směřování, což mohlo způsobit nárůst zpoždění (při ověřování změřených hodnot zpoždění bylo u některých cílů naměřeno lepších výsledků). Další věcí je sama dostupnost stanic: dvě teoreticky dobře dostupné stanice, nepříliš geograficky

vzájemně vzdálené (např. 150 km) mohou mít oproti sobě latenci lišící se jen např. 5–15 ms a z principu nejbližšího souseda v mapě zpoždění tyto stanice mohou vypadat jako sousedé opravdu vzdálení. Toto vede k myšlence, že pokud je zpoždění k některým stanicím navýšeno i jen o malou hodnotu, vykompenzovat takovéto zkreslení je těžce proveditelné.

Po rozebrání jednotlivých příčin způsobujících chybu určení polohy cíle bohužel docházím k závěru, že v dnešní době kdy se geolokační technologie na bázi měření zpoždění ubírají jiným (a lepším) směrem, by jakékoliv úsilí věnované ke zlepšení této metody nevedlo ke konkurenceschopnosti těmto novým metodám (CBG, SOI, Octant [30], Geoweight [1]).

## 7 ZÁVĚR

Ve shodě se zadáním jsem se v této bakalářské práci zaměřil na problematiku geolokačních metod využívajících ke své činnosti informaci o zpoždění dat v síti Internet a proto je většina práce věnována jejich profilu, zejména pak metodě GeoPing, kterou jsem využil při experimentálním měření. Moji snahou bylo vystihnout především princip konkrétních metod, postupy a techniky, kterých využívají při dopracovávání se k výsledkům, jejich výhody a nedostatky a také informovat o jejich přesnosti, respektive úspěšnosti v určování neznámé pozice cíle, nežli se pouštět do detailních rozborů a výpočtů (výjimku z pochopitelných důvodů tvoří metoda GeoPing, té byla věnována mnohem větší pozornost). Geolokační techniky uvedeny v této práci tedy jsou: GeoPing, Shortest Ping, Simple GeoPing, Constraint-Based Geolocation (CBG) a Speed of Light (SOI). O prvních třech jmenovaných metodách lze prohlásit, že ke geolokaci změřené zpoždění nijak nepřevádí na vzdálenost; pozici cíle vyvozují na základě shody zpoždění s jedním ze svých referenčních bodů. CBG a SOI oproti těmto metodám přichází se zajímavým přístupem vytváření „přesných konstant“ pro transformaci konkrétního zpoždění na konkrétní vzdálenost a díky tomuto přístupu dosahují velmi zajímavých výsledků.

Informace, které jsem načerpal při vytváření této práce, mě přivedly k názoru, že ať s budoucím vývojem Internetu přijde cokoli, metody založené na měření zpoždění budou z důvodů jejich jednoduchosti a robustnosti i nadále využitelné, protože ke své plné funkčnosti vystačí pouze se znalostí IP adresy cíle a s k ní vztaženou latencí. Také si troufám předpovědět, že s budoucím rozsáhlým pokrokem ve zkvalitňování síťové infrastruktury (zejména v Evropské unii), perspektiva těchto metod bude jen stoupat.

Za přínosnou součást této práce považuji sestavení datasetu, který jsem použil při geolokaci několika desítek vybraných stanic. Nejnáročnější částí celé práce bylo nalezení dostatečného počtu stanic, které v datasetu zastávají roli referenčních bodů a cílů. Naopak, jednodušším, ale stejně důležitým krokem bylo opatření si přístupu k 9 stanicím, sloužících jako sondy. K tomuto účelu bylo nutno získat přístup do sítě PlanetLab, kde jsem mohl využít vzdálených stanic – nenahraditelného prvku celého experimentu. Ve finální podobě dataset čítá 269 stanic, jejichž poloha byla ověřována zejména pomocí dotazováním se do WhoIs databáze.

Tento dataset je použit jako jádro celé praktické části – určování polohy stanice metodou GeoPing, jež svou činnost staví na myšlence nejbližšího souseda v mapě zpoždění. Vydatné usnadnění při vytváření mapy zpoždění jsem dosáhl díky jednoduchému (bash) shell skriptu provádějícího měření zpoždění, který jsem pro tento účel naprogramoval. Při experimentu nastalo

několik komplikací různé závažnosti. Ty nejzásadnější byly ve formě výpadku provozu většiny sond a 4 sondy musely být dokonce nahrazeny jinými (naštěstí se našly 3 sondy umístěné na stejných místech jako ty původní). Další typ komplikace byl v podobě nedostupnosti některých referenčních bodů z hlediska měření zpoždění. Toto bylo ve všech případech vyřešeno nalezením uzlu, který ležel na trase a byl danému ref. bodu z hlediska polohy nejbližší, respektive se s ním z geografického hlediska přibližně shodoval. Ve finále byla vytvořena kompletní mapa zpoždění, znázorňující prostor zpoždění mezi sondami a referenčními body. Pomocí aplikace Euklidovské vzdálenosti na tento prostor zpoždění bylo možno dokončit experiment a vyvodit závěry.

Nakonec tedy bylo ověřeno, za jakých podmínek je metodou GeoPing možno docílit výsledků, kterých dosáhli samotní autoři této metody a dokonce je mírně předčit. Experiment však také potvrdil poměrně nízkou kvalitu této metody a prokázal fakt, že její zlepšování by bylo náročné a neperspektivní. Proto na základě získaných výsledků doporučuji v praxi použít modernějších a přesnějších geolokačních metod, nebo je metodou GeoPing vhodně doplnit.

## REFERENCE

- [1] ARIF, M. J. ; KARUNASEKERA, S. ; KULKARNI, S. . GeoWeight : Internet Host Geolocation Based on a Probability Model for Latency Measurements. In *Research and Practice in Information Technology (CRPIT)*. Brisbane, Australia : , 2010. s. 90-98.
- [2] BOVY, C. J., et al. *Analysis of End-to-end Delay Measurement in Internet* [online]. PAM 2002, 2002. s. 8. Dostupný z WWW: <[http://www.pamconf.net/2002/Analysis\\_of\\_End\\_to\\_end\\_Delay\\_Measurements\\_in\\_Internet.pdf](http://www.pamconf.net/2002/Analysis_of_End_to_end_Delay_Measurements_in_Internet.pdf)>.
- [3] Broadband coverage in rural areas. In *Europe's Digital Competitiveness Report* [online]. Brusel : 17.5.2010 [cit. 2011-05-28]. Dostupné z WWW: <<http://ec.europa.eu/ceskarepublika/pdf/press/ks7ec2.pdf>>.
- [4] *Domaintools* [online]. 2010 [cit. 2010-11-20]. WHOIS Lookup at DomainTools.com. Dostupné z WWW: <<http://whois.domaintools.com/>>.
- [5] *Daftlogic.com* [online]. 2005-12-29 [cit. 2011-05-29]. Google Maps Distance Calculator. Dostupné z WWW: <<http://www.daftlogic.com/projects-google-maps-distance-calculator.htm>>.
- [6] *Eso.org* [online]. 2008 [cit. 2011-05-28]. How does GPS work. Dostupné z WWW: <<http://www.eso.org/public/outreach/eduoff/seaspace/docs/navigation/navgps/navgps-3.html>>.
- [7] *Game Servers* [online]. 2011 [cit. 2011-05-30]. Dostupné z WWW: <[www.game-monitor.com/](http://www.game-monitor.com/)>.
- [8] *Geolocator* [online]. V0.28. 2008, 31.3.2010 [cit. 2011-05-20]. Dostupné z WWW: <<http://tools.freeside.sk/geolocator/geolocator.html>>.
- [9] *Google Maps* [online]. [cit. 2011-05-20]. Dostupný z WWW: < <http://maps.google.com/> >.
- [10] GUEYE, Bamba, et al *Constraint-Based Geolocation of Internet Hosts*. In *Internet Measurement Conference*. 2004. [s.l.] : ACM/SIGCOMM, 2004. s. 14. Dostupné z WWW: <<http://www.cs.bu.edu/fac/crovella/paper-archive/imc04-geolocation-full.pdf>>. E00-108534666
- [11] HOLZHAUER, Florian. *IP Geolocation*. Berlin, 2007. 11 s. Seminární práce. Technische Universität Berlin. Dostupné z WWW: <[http://fholzhauer.de/wp-content/uploads/2006/10/ip\\_geolocation\\_handout.pdf](http://fholzhauer.de/wp-content/uploads/2006/10/ip_geolocation_handout.pdf)>.
- [12] CHATZOPOULOU, Doxa; KOKKODIS, Marios. *IP Geolocation* [online]. [s.l.] , . 6 s. . Computer Science and Engineering Department, UC Riverside. Dostupné z WWW: <<https://www.cs.ucr.edu/~mak/grad/geolocation.pdf> >.

- [13] *IPinfoDB* [online]. 2010 [cit. 2010-12-13]. IP address geolocation database. Dostupné z WWW: <[http://ipinfodb.com/ip\\_database.php](http://ipinfodb.com/ip_database.php)>.
- [14] JAROŠOVÁ, Lucie. *Měření vzdáleností stanic prostřednictvím ICMP protokolu v IP sítích*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací 2008. 63 s. Vedoucí bakalářské práce Ing. Radim Burget.
- [15] KOMOSNÝ, Dan; BALEJ, Jiří. Zdroje zpoždění při komunikaci v Internetu. *Elektrorevue* [online]. 2010, 2010, 42, [cit. 2010-10-15]. Dostupný z WWW: <[www.elektrorevue.cz](http://www.elektrorevue.cz)>. ISSN 1213-1539.
- [16] KATZ-BASSETT, Ethan, et al Towards IP Geolocation Using Delay and Topology Measurements. In *Internet Measurement Conference, 2006*. [s.l.] : ACM/SIGCOMM, 2006. s. 13. Dostupné z WWW: <<http://conferences.sigcomm.org/imc/2006/papers/p7-bassett.pdf>>. ISBN 1-59593-561-4.
- [17] MILAR, Bohdan. První část zajímavého a čtivého seriálu o BASHi. *Linuxexpres* [online]. 2005-6-30, [cit. 2011-04-27]. Dostupný z WWW: <<http://www.linuxexpres.cz/praxe/bash-1-dil-1?highlightWords=Bash%2C+d%C3%ADl>>.
- [18] Release of IP2Location IPv6 Country Free Edition. *PRLog* [online]. 2011-05-03, , [cit. 2011-05-20]. Dostupný z WWW: <<http://www.prlog.org/11471235-may-2011-release-of-ip2location-ipv6-country-free-edition.html>>
- [19] PADMANABHAN , Venkata N. ; SUBRAMANIAN, Lakshminarayanan. An Investigation of Geographic Mapping Techniques for Internet Hosts. In *Internet Measurement Conference, 2001*. [s.l.] : ACM/SIGCOMM, 2001. s. 13. Dostupné z WWW: <<http://conferences.sigcomm.org/sigcomm/2001/p14-pabmanabhan.pdf>>. ISBN 1-58113-411-8.
- [20] PÉK, Vojtěch. *Využití informace z ADS-B v pasivních multilateračních systémech*. Plzeň, 2008. 51s. Diplomová práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra matematiky.
- [21] *PlanetLab* [online]. 2007 [cit. 2011-05-30]. Dostupné z WWW: <<http://planet-lab.org/>>.
- [22] PlanetLab European Tutorial : An open, shared platform for developing, deploying, and accessing planetary scale applications. *FIREworks* [online]. 2008, , [cit. 2010-10-15]. Dostupný z WWW: <[http://www.ict-fireworks.eu/fileadmin/events/FIREweek/PL\\_1.pdf](http://www.ict-fireworks.eu/fileadmin/events/FIREweek/PL_1.pdf)>.
- [23] *PuTTY FAQ* [online]. 2010-10-20 [cit. 2010-11-18]. PuTTY: A Free Telnet/SSH Client. Dostupné z WWW: <<http://www.chiark.greenend.org.uk/~sgtatham/putty/faq.html#faq-timeout>>.
- [24] *Putty.org* [online]. PuTTY - a free SSH and telnet client for Windows. Dostupné z WWW: <<http://www.putty.org/>>.

- [25] RFC 792. *Internet Control Message Protocol : DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION*. [s.l.] : USC/Information Sciences Institute, 1981. 20 s. Dostupné z WWW: <<http://www.faqs.org/rfcs/rfc792.html>>.
- [26] RFC 1834. *Whois and Network Information Lookup Service Whois++*. University of California, Davis : J. Gargano, K. Weiss, 1995. 7 s. Dostupné z WWW: <<http://www.faqs.org/rfcs/rfc1834.html>>.
- [27] Root.cz *Programování v (Bash) Shellu*. 2005-6-30, [cit. 2011-04-27]. Dostupný z WWW: <<http://www.root.cz/clanky/programovani-v-bash-shellu/>>
- [28] SVANTESSON, Dan. *Geo-location Technologies*. [s.l.], 2004. 10 s. . Faculty of Law, Bond University (Australia). Dostupné z WWW: <[www.svantesson.org/svantesson20040906.doc](http://www.svantesson.org/svantesson20040906.doc)>.
- [29] Studie Cisco Visual Networking Index očekává vysoký nárůst objemu mobilních dat. *IT Point* [online]. 2011-2-10, , [cit. 2011-05-28]. Dostupný z WWW: <<http://www.itpoint.cz/zprava/?i=studie-cisco-visual-networking-index-ocekava-vysoky-narust-objemu-mobilnich-dat-6572>>.
- [30] WONG, Bernard ; STOYANOV, Ivan ; SIRER, Emin Gün . Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. In *USENIX Symposium on Networked Systems Design & Implementation* [online]. Cambridge, Massachusetts : [s.n.], 2007 [cit. 2011-05-24]. Dostupné z WWW: <[https://www.usenix.org/events/nsdi07/tech/full\\_papers/wong/wong.pdf](https://www.usenix.org/events/nsdi07/tech/full_papers/wong/wong.pdf)>.
- [31] What is the meaning of Geolocation?. *Security-Science* [online]. [cit. 2011-05-30]. Dostupný z WWW: <<http://www.security-science.com/security-encyclopedia/item/geolocation>>.
- [32] ZIVIANI, Artur, et al. Demographic Placement for Internet Host Location. In *Global Telecommunications Conference, 2003*. [online]. San Francisco, USA : GLOBECOM, 2003-12, 2004-12-14 [cit. 2011-05-28]. Dostupné z WWW: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.7487>>. ISBN 0-7803-7974-8 , 10.1109/GLOCOM.2003.1258945 .

## SEZNAM PŘÍLOH

|   |                            |    |
|---|----------------------------|----|
| A | Obsah přiloženého DVD..... | 54 |
|---|----------------------------|----|

## **A OBSAH PŘILOŽENÉHO DVD**

Součástí práce je přiložené DVD obsahující materiály, jenž byly použity a zmíněny v práci.

- Elektronická verze práce.
- Použité obrázky.
- Dokument `Mapa_zpozdeni.xlsx`, vytvořený v programu Microsoft Office Excel 2007, obsahující veškeré tabulky a výpočty vztahující se k mapě zpoždění a určování pozice cíle.
- Skript `ping.sh` použitý pro měření zpoždění.