

## Metodika: Mapa rychlosti internetu v ČR

Tento dokument obsahuje popis postupu měření rychlosti internetu vycházející z provozních dat společnosti Seznam.cz. V České republice chybí dostatečné a spolehlivé statistiky zachycující stav rychlosti internetu, proto se Seznam.cz pokusil přispět do diskuze vlastním měřením. Jedná se o první verzi metodiky, která je otevřená pro další úpravy na základě budoucích podnětů.

### Sběr dat

Data vstupující do tohoto měření jsou sbírána na hraničním routeru datacentra Nagano, kde společnost Seznam.cz sleduje veškerou komunikaci se službami. Pro každé spojení se například zaznamenává, kolik dat se přeneslo, na jakých portech, jak dlouho komunikace probíhala, z jaké byla IP adresy a řada dalších ukazatelů. Záznam pro jedno spojení se nazývá tok dat (tzv. flow). Jeden tok dat je na hraničním routeru zachytáván maximálně 5 minut, pokud je komunikace delší, záznam se za 5 minut odešle a začne se měřit další záznam. Všechny toky dat se posílají na sběrné místo, kde se uchovává čtrnácti denní historie.

Sběr dat probíhal v období 16.2.-22.2.2015. Pro každou IP adresu byly vybrány 4 hodnoty z každé hodiny v časovém rozmezí 17h až 21h. Z těchto 28 hodnot pro jednotlivé IP adresy byla vybrána ta nejvyšší.

### Příprava a zpracování dat

Pro potřeby měření byly ve sledovaném období vyhodnocovány komunikace (datové toky) mezi IP adresami a CDN (Content Delivery Network) servery. Volba využití CDN byla dána tím, že video obsah je z pohledu přenášené velikosti dat náročnější, než tomu je u jiných služeb Seznamu a bývá tak využita maximální rychlost uživatelských přípojek.

Dále byly vyfiltrovány bloky přenášených dat s pakety většími než 1 000 B, komunikující déle než 5 sekund a s komunikující rychlostí více než 20 paketů za sekundu (tzn. pouze přenosy větších dat). Takto vznikly záznamy z více než 450 000 IP adres z České republiky vstupující do závěrečné analýzy.

### Výsledky měření

Výstupem z výše uvedených dat jsou grafy pro jednotlivé kraje zobrazující procentuální zastoupení komunikujících IP adres s CDN servery v jednotlivých rychlostních intervalech. Intervaly jsou zleva uzavřeny (tj. započítává se i první hodnota intervalu). Pro pohodlnější orientaci byly grafy zaneseny k jednotlivým krajům do mapy. Určení lokality IP adresy bylo provedeno napojením GPS souřadnic ke každé IP adrese pomocí údajů z GeoIP databáze serveru maxmind.com. Přesnost GeoIP je velice diskutovaným tématem v odborných kruzích a i společnost Seznam.cz si uvědomuje limity této metody a bude v blízké budoucnosti podnikat kroky k lepšímu určení polohy uživatelů.

Další pohled na data poskytují histogramy za jednotlivé kraje a okresy zachycující počet IP adres v detailnějších rychlostních intervalech. Histogramy a další výstupy jsou k dispozici ke stažení na webové stránce [www.seznam.cz/mapa-rychlosti-internetu](http://www.seznam.cz/mapa-rychlosti-internetu).

Společnost Seznam.cz dlouhodobě podporuje snahy o otevřená data, proto v tomto případě dává k dispozici zájemcům soubor obsahující zdrojová data sesbírána pro potřeby tohoto měření. Z legislativních důvodů byly ze souboru vymazány konkrétní IP adresy.