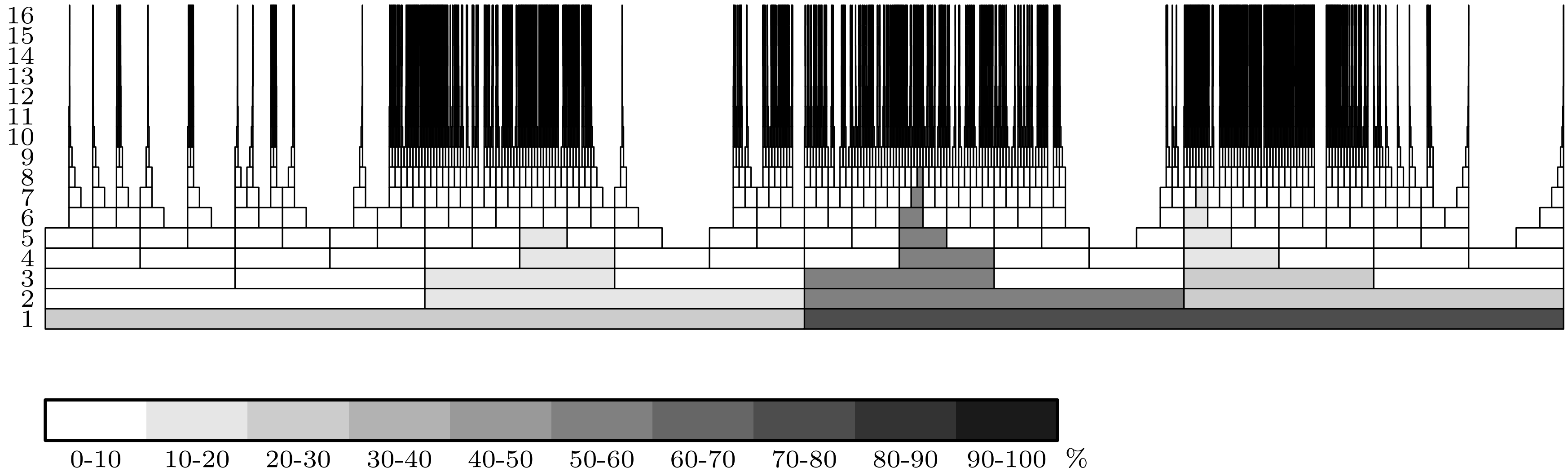
Kvantitativní analýza internetového provozu (7)

Ladislav Lhotka
⟨*lhotka@cesnet.cz*⟩

Osnova přednášky

- Fraktály a multifraktály
- Histogramová metoda
- Model
- Přehled otázek ke zkoušce
- Q & A

Fraktální struktura adres



Dělíme adresový rozsah na 2^p prefixů délky p, pro $p = 1, 2, \ldots, 32$. Počet prefixů obsahujících aspoň jednu adresu ze vzorku

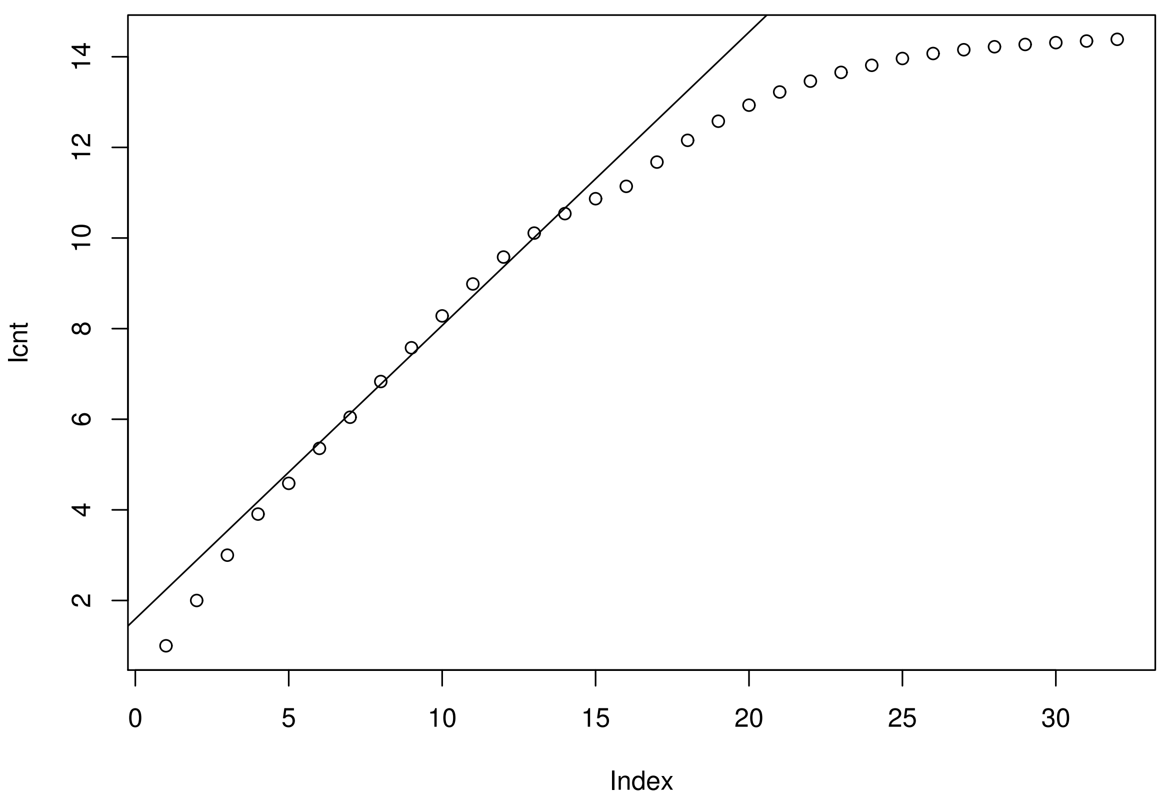
$$N_p \approx 2^{Dp}$$

Fraktální dimenze (box-counting dimension)

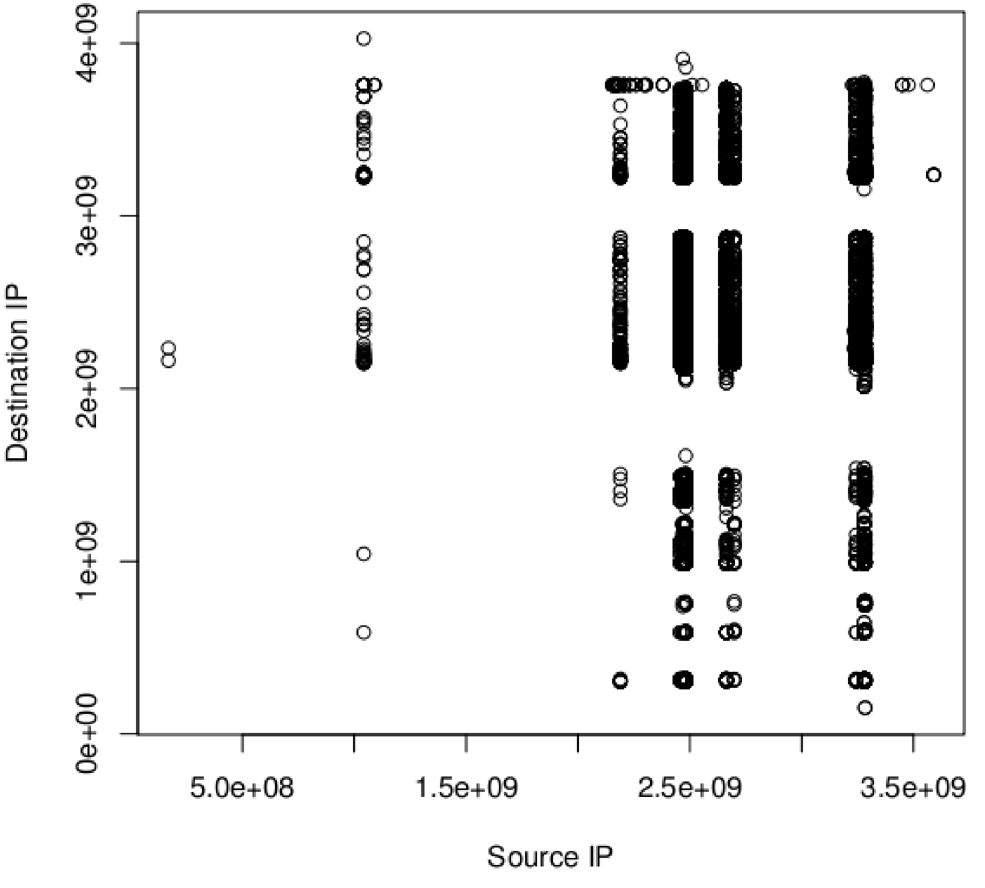
$$D = \frac{\log_2 N_p}{p}$$

Určení ze vzorku

Vzorek z okruhu GN2, $\mathrm{D} \approx 0,65$ jako směrnice tečny.

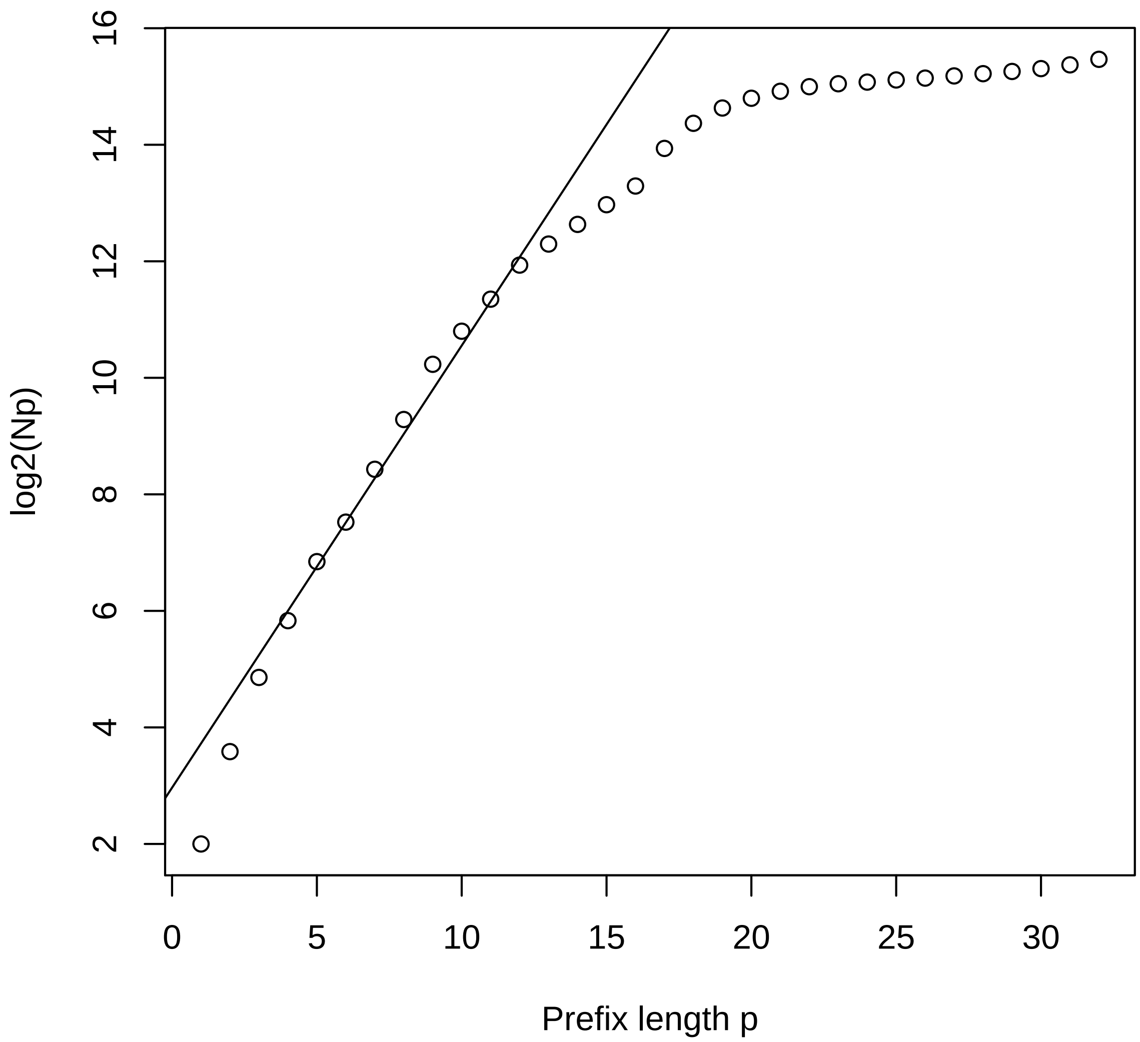


Dvourozměrná data



Odhad dimenze pro obě adresy

```
flows <- read.csv("../4-081103/gn2-flows-10ge.data",
    col.names=c("source","dest","proto","sip","dip",
    "sport","dport","packets","bytes"))
adresy <- flows[c("sip","dip")]
lcnt <- c()
for (p in 1:32) {
    lcnt <- c(log2(length(unique(adresy)$dip)),lcnt)
    adresy <- adresy %/% 2
}
plot(lcnt, xlab="Prefix length p", ylab="log2(Np)")
p <- 3:13
np <- lcnt[p]
fit <- lm(np ~ p)
abline(coef(fit))
```



$D \approx 0,76$

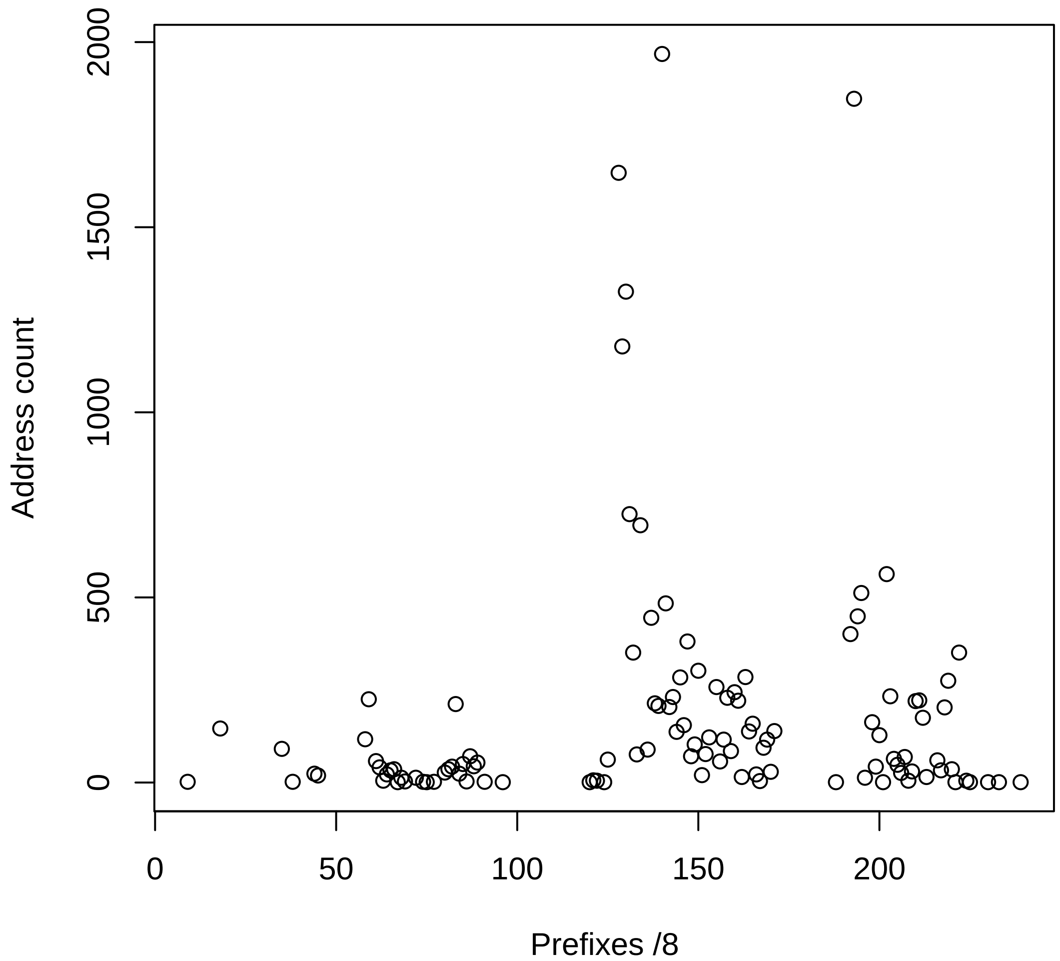
Váha prefixů

Výpočet fraktální dimenze bere v úvahu pouze, zda je daný prefix $/p$ ve vzorku vůbec zastoupen, tj. zda obsahuje aspoň jednu cílovou adresu patřící do prefixu $/p$.

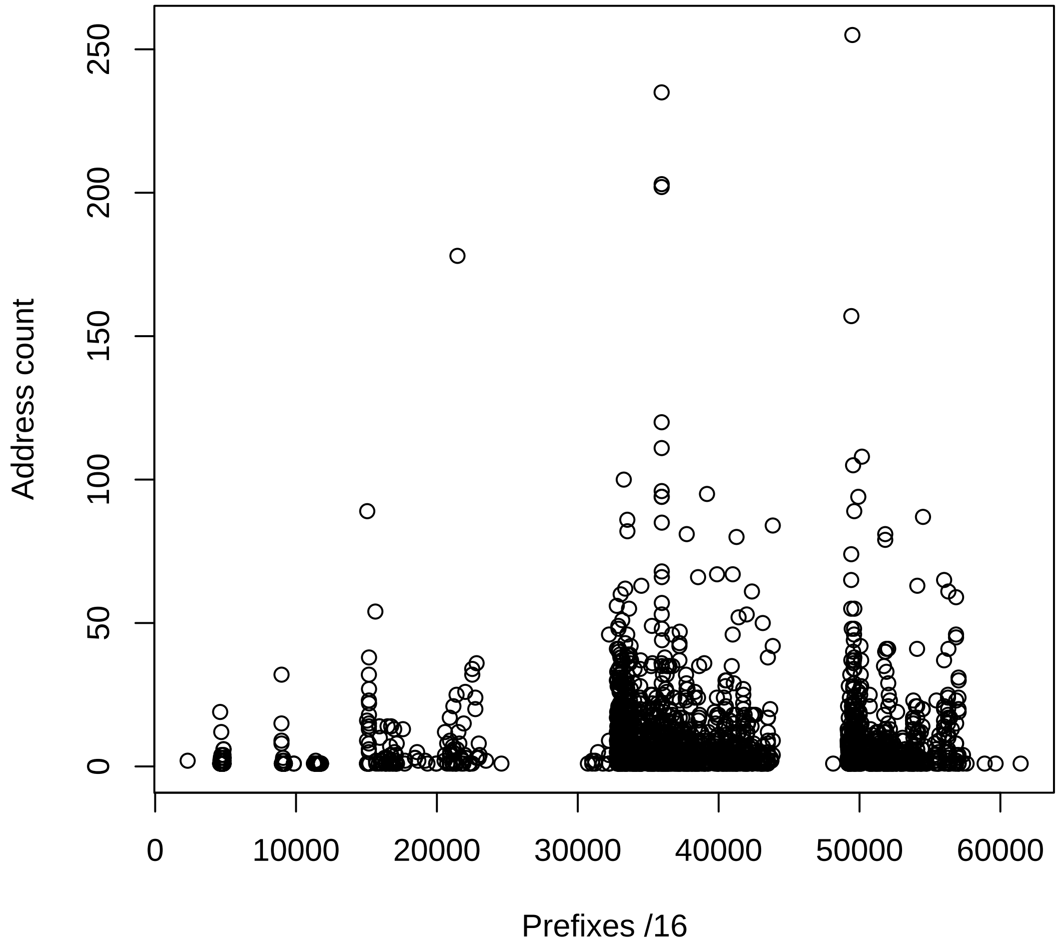
Různé prefixy ale zdaleka nemají stejnou váhu, pro detailnější analýzu je tedy potřeba brát v úvahu např. kolik (i) adres, (ii) paketů nebo (iii) bajtů patří do daného prefixu.

Pro tento účel se definuje *míra* μ, která může, ale nemusí být normována na 1.

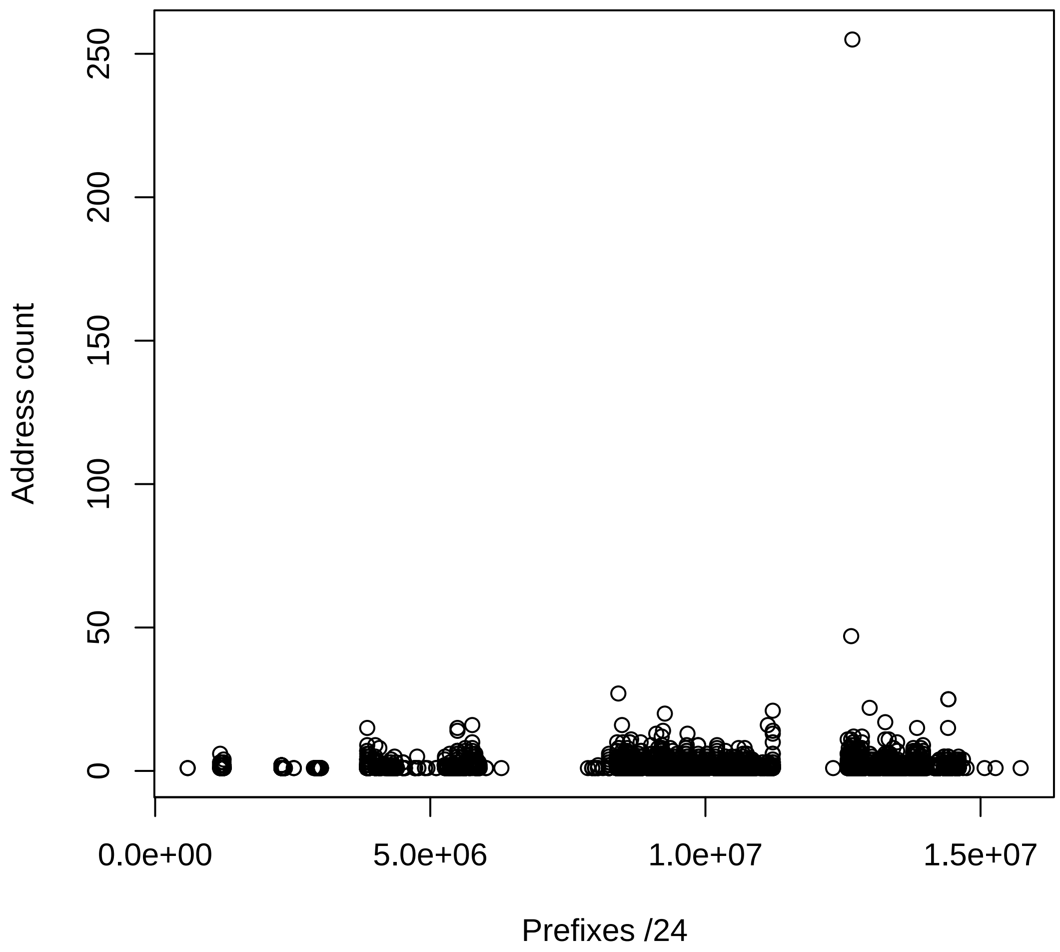
Adresy v prefixech /8



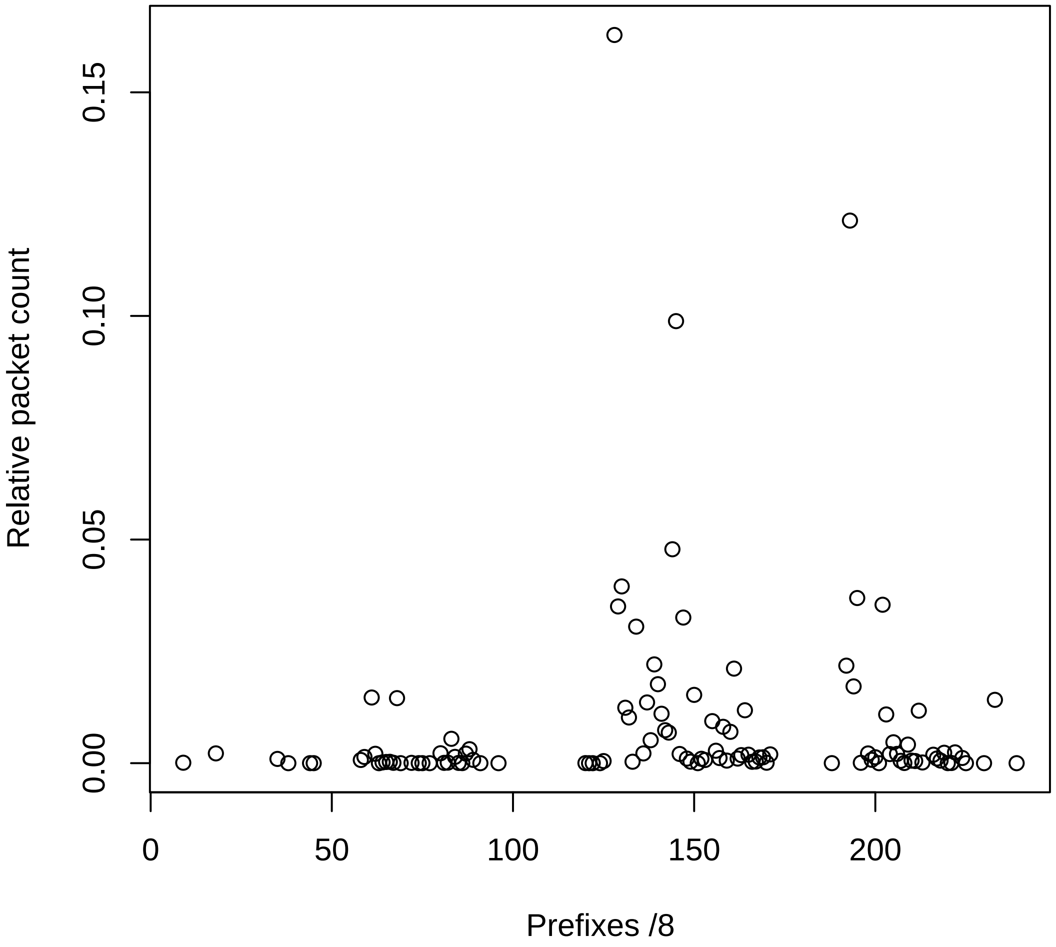
Adresy v prefixech /16



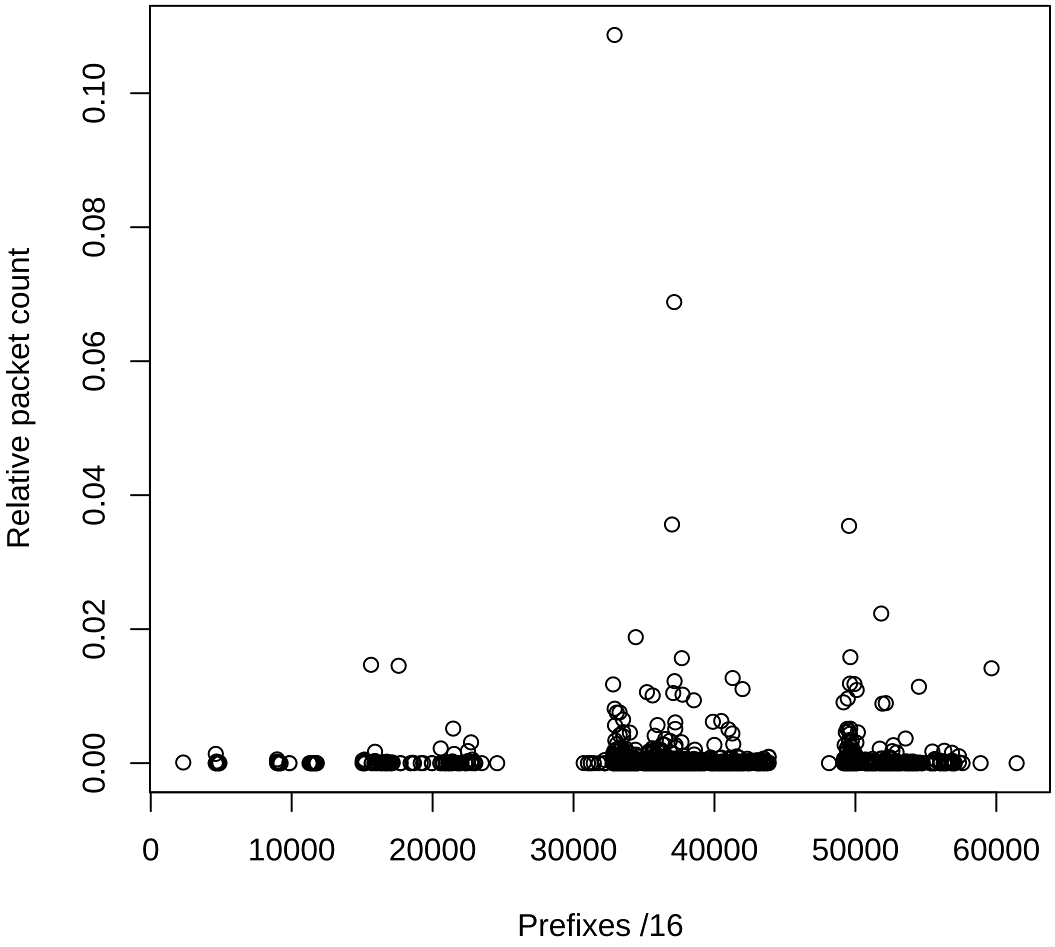
Adresy v prefixech /24



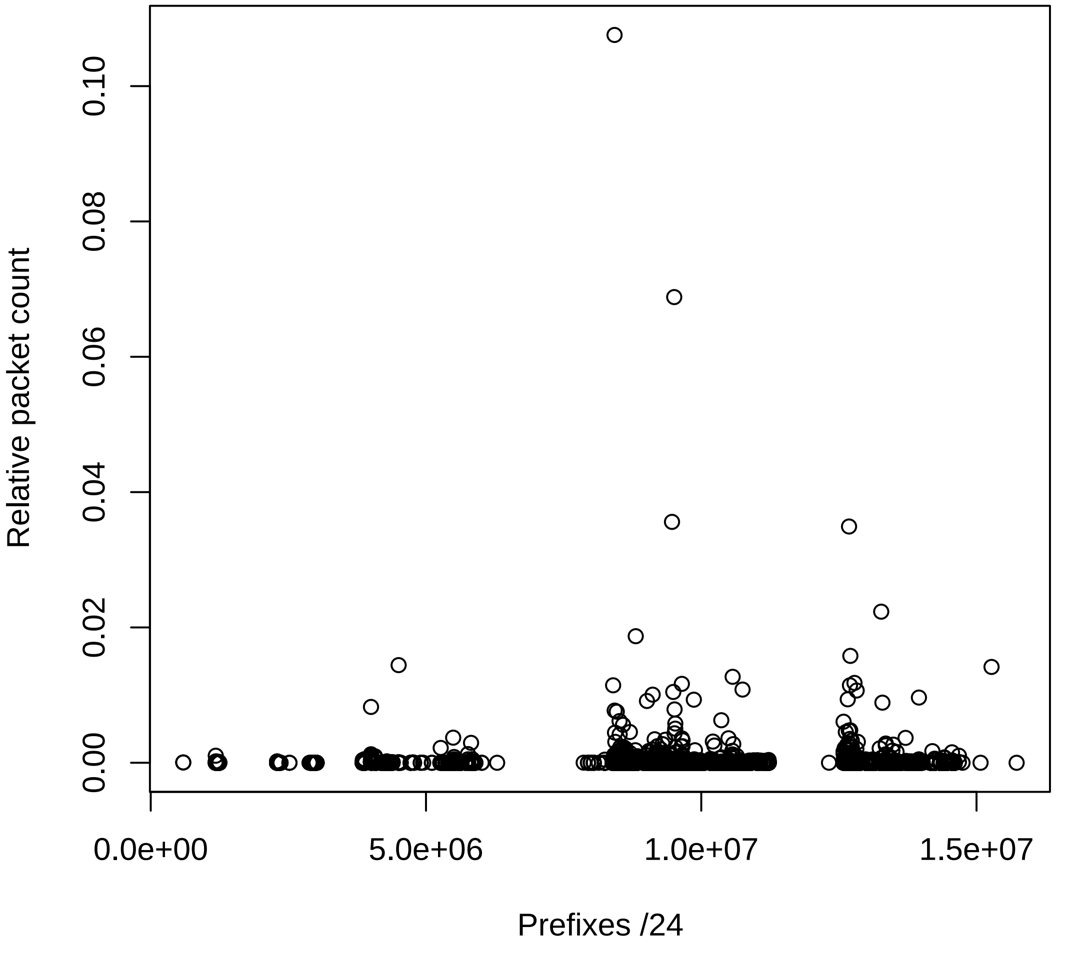
Rel. počty paketů v prefixech /8



Rel. počty paketů v prefixech /16



Rel. počty paketů v prefixech /24



Multifraktální míra

Podobně jako u fraktální dimenze dělíme adresový rozsah na prefixy délky $/p$ a sledujeme, jaká je míra každého prefixu a jak rychle klesá s rostoucím p. Tato charakteristika je zřejmě závislá na pozici v adresovém rozsahu:

$$\mu(a/p) \approx 2^{-p\alpha(a)}$$

Hölderův exponent:

$$\alpha(a) = -\frac{\log_2(\mu(a/p))}{p}$$

Zajímá nás relativní zastoupení jednotlivých hodnot α:

$$N_p(\alpha) \approx 2^{pf(\alpha)}$$

Charakteristika míry:

$$f(\alpha) = \frac{\log_2 N_p(\alpha)}{p}$$

Histogramová metoda

1. Při rozdělení adresového prostoru na prefixy délky p spočítej Hölderův exponent pro každý prefix P_i ($i = 1, 2, \ldots, 2^p$):

$$\alpha_i = -\frac{\log_2(\mu(P_i))}{p}$$

2. Vytvoř histogram hodnot α_i, tj. rozděl celý rozsah hodnot α na intervaly a zjisti počet hodnot $N_p(\alpha)$ v každém intervalu.
3. Nakresli graf (aproximace) funkce $f(\alpha) = \log_2 N_p(\alpha)/p$.
4. Postup opakuj pro různá p.

Algoritmus v R

```
flows <- read.csv("../4-081103/gn2-flows-10ge.data",
    col.names=c("source","dest","proto","sip","dip",
    "sport","dport","packets","bytes"))
udip <- unique(flows$dip)
totad <- length(udip)
p <- 16
pref <- rle(sort(udip %/% 2**p))
mu <- pref$length / totad
alfa <- -log2(mu) / p
hist <- hist(alfa, breaks=1000)
plot(hist$mids, log2(hist$counts)/p)
```

Výsledky Kohler et al.

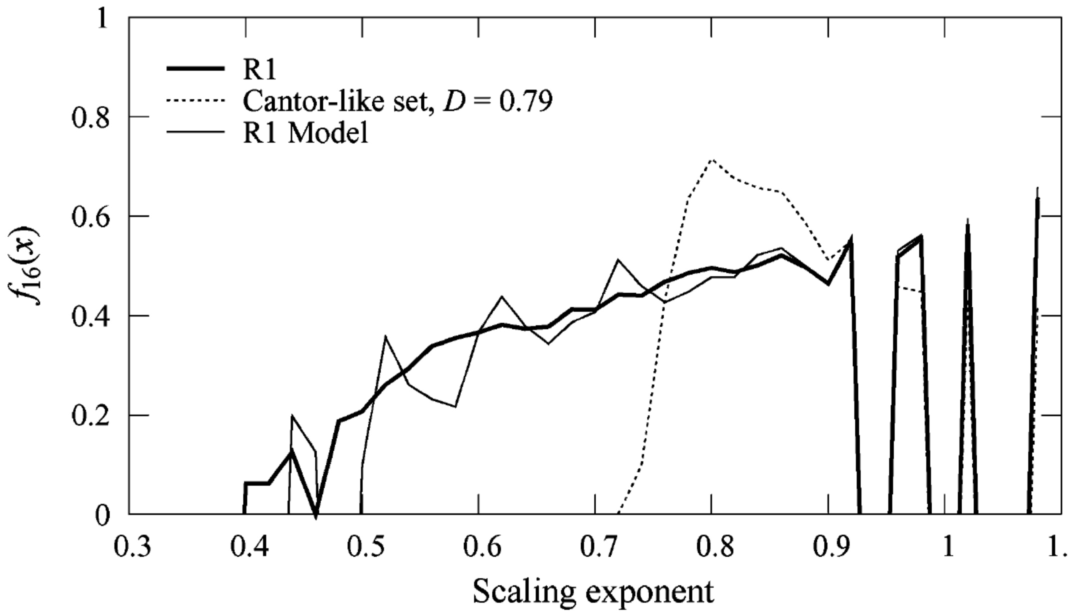
V konstrukci Cantorovy množiny lze délkou h vyjmutého prostředního intervalu regulovat fraktální dimenzi:

$$D = -\frac{\log 2}{\log \frac{1}{2}(1 - h)}$$

Fraktální model: “zředit” adresový prostor na Cantorovu množinu s fraktální dimenzí shodnou s dimenzí cílových adres ve vzorku.

Spekulativní konstrukce, není žádný důvod k tomu, aby adresy ve vzorku měly kvalitativně strukturu Cantorovy množiny.

Multifraktální spektrum (Kohler)



Multifraktální model

Stejná konstrukce “Cantorova ředění” adresového prostoru, v každém kroku se ale oběma zbývajícím částem přiřadí míry (= pravděpodobnost výskytu adresy) m_0 a $1 - m_0$. Parametr m_0 je odhadnut tak, aby jeho multifraktální spektrum co nejlépe souhlasilo s modelovaným vzorkem.