Detekce anomálií v síťovém provozu

Práce se zabývá se optimalizací metody navržené v článku

Lakhina, A. et al.: Mining Anomalies Using Traffic Feature Distributions. *ACM SIGCOMM Computer Communication Review*, r. 35, č. 4, 2005.

Metoda detekuje anomálie jako změny v histogramech zdrojových a cílových adres a portů.
Z histogramů je vypočítávána entropie. Odchylka entropie od normálu v několika dimenzích současně – anomálie.

Optimalizace rychlosti

Problém

- Jak uložit histogram?
- Obyčejné pole
  - Rychlé
  - Ale pro některé dimenze příliš velké (např. pro IP adresy 2^{32} položek)
- Asociativní pole
  - Optimální využití paměti
  - Pomalé

Navržené řešení

- Pomocí libovolné hash funkce převést vstupní hodnoty na menší šířku
- Stačí malé obyčejné pole
- Kolize nevadí – výsledky jsou ovlivněny minimálně

Výsledky

- Až 20-násobné zrychlení
  - V porovnání s použitím asociativního pole
- Spotřeba paměti pod 1 MB

Zvýšení citlivosti

Problém

- Metoda je málo citlivá
- Aby byla anomálie detekována, musí tvořit alespoň 5-10% objemu provozu.
  - To na vysokorychlostních sítích znamená opravdu masivní útoky

Navržené řešení

- Přidány další dimenze
  - Kromě adres a portů i počet paketů a bajtů v toku a TCP flagy
- Použity i kombinace dimenzí
  - Sledujeme např. četnost výskytů daného portu na dané adrese
- Navrženo 5 nových metrik místo entropie

Výsledky

- Některé z navržených kombinací dimenzí a metrik poskytují mnohonásobně vyšší citlivost na různé typy útoků.