Rozšíření systému NEMEA pro nasazení v distribuovaném prostředí

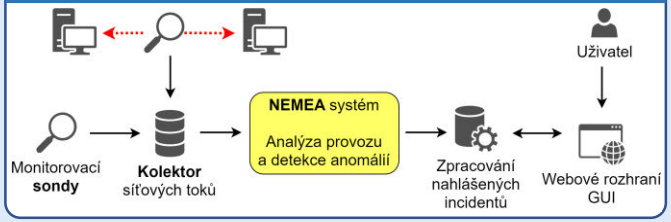
Marek Švepeš vedoucí práce: Tomáš Čejka

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Motivace

- Důležitost monitorovacích systémů v počítačových sítích – **detekce bezpečnostních incidentů**.
  1. Rostoucí počet zařízení a rychlost počítačových sítí – **více dat ke zpracování**.
  2. Přibývající bezpečnostní hrozby – **mnoho detekčních algoritmů najednou**.
- **Problém:** zpracování rostoucího množství dat na jednom výpočetním stroji mnoha algoritmy najednou ➡ neudržitelný stav!
- **Možné řešení:** paralelní zpracování dat.
- Chybějící řešení zaměřené na správné výsledky detekce pro systémy, které pracují se síťovými toky.

Monitorovací infrastruktura

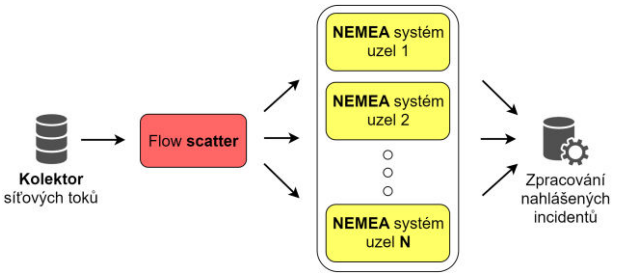


Systém NEMEA

- Provádí analýzu síťového provozu a detekci anomálií.
- Zpracovává **síťové toky** (agregované hlavičky paketů).
- **Modulární** – přirozeně distribuovatelný
- **Proudové** zpracování dat v **reálném čase**
- Open-source projekt

Navržené řešení

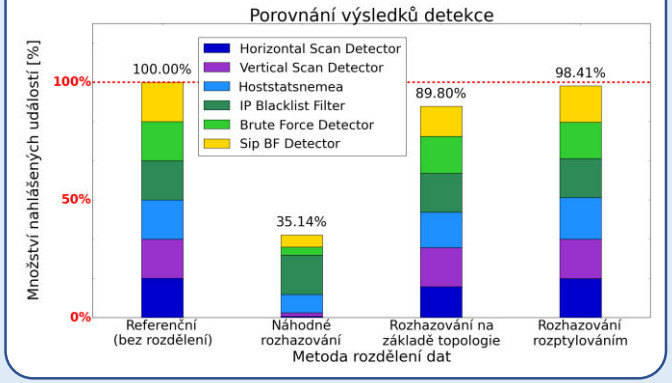
- Rozdělení proudu síťových toků mezi nezávislé stejné výpočetní uzly s následujícími požadavky:
  1. Rovnoměrnost rozdělení
  2. Efektivita rozdělení
  3. **Zachování správnosti výsledků detekce**



- Prvek **Flow scatter** rozděluje proud síťových toků pro paralelní zpracování. Experimentálně testované tři metody rozdělení:
  1. *Náhodné rozhazování*
     - Použito statistické rovnoměrné rozdělení.
  2. *Rozhazování na základě topologie*
     - Síťové toky rozděleny podle identifikátoru monitorované linky.
  3. *Rozhazování rozptylováním*
     - Podle vstupu rozptylovací funkce rozdělení toků na podmnožiny se stejnou charakteristikou.
     - Identifikovány 3 skupiny detekčních metod, které potřebují rozptylovat zvlášť podle zdrojové IP, cílové IP, uspořádané dvojice IP adres.
- Flow scatter **realizován** jako NEMEA modul.

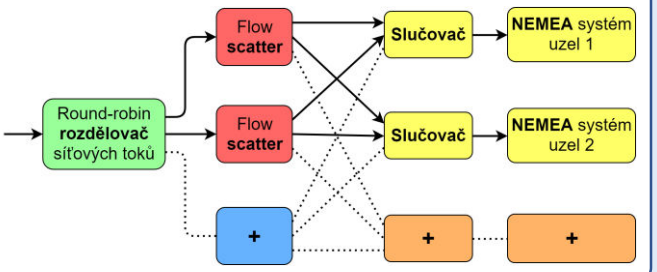
Experimentální ověření

- Použity pseudonymizované **reálné datové sady** z akademické sítě CESNET2 k otestování všech metod.
- Výsledky nejlepší metody *rozhazování rozptylováním:*
  - Rovnoměrnost rozdělení toků **se blíží optimu**.
  - V testovacím prostředí rozdělení **1,3 milionu toků za sekundu = zpracování 2,3 krát více dat**.
  - **Zachovává sémantické vazby** mezi toky – správné výsledky detekce.



Návrh škálovatelné architektury

- Celkovou propustnost systému lze zvýšit duplikací prvku Flow scatter a výpočetních uzlů.



Publikace: Švepeš M., Čejka T. "Making Flow-Based Security Detection Parallel." *IFIP International Conference on Autonomous Infrastructure, Management and Security,* AIMS2017. Springer, Cham, 2017.