Analýza nákazy v domácí síti

Robin Obůrka • robin.oburka@nic.cz • 3.12.2016

Turris – Aktuální analýzy

- „Kdo nám klepe na vrátka“
  - FW logy
  - Honeypot (SSH)
  - Minipot (Telnet)
- Jednoduché, ale máme výsledky
  - Nakažené CCTV kamery
  - Rezonance útoků na známé chyby v routerech
- Jak pomoci uživateli?
  - Vydání aktualizací FW
  - Medializace problému

Kdo jsou útočníci?

- Nakažené servery
- Zařízení v koncové síti
  - Počítače s Windows
  - Io(s)T
  - Domácí NAS
- Jak pomoci uživateli?
  - Nemá náhodou v síti nakažené zařízení?
  - Jak to zjistit?
    - Turris útočí na Turris
    - → analýza provozu sítě

Jaká data sbíráme

- Omezeno smlouvou
  - Hlavičky packetů
  - Pouze 10 denní okno
- Netflow-like data
  - (zdrojová adresa/klient, zdrojový port, protokol, cílový port, cílová adresa)
  - Pro každý směr: počet packetů, velikost přenesených dat, délka komunikace
- Další typy nebudou diskutovány v přednášce

Jaká data (ne)chceme sbírat

- Razíme přístup „ti hodní“
- Nesbíráme kompletní netflows
- Pouze tam, kde je podezřelá protistrana
- Vybraná sada portů často spojovaná s útoky
  - 25/TCP, 22/TCP, 23/TCP, ...
- Zakázali jsme si porty jako 80/TCP, 443/TCP

Analýza nákazy v domácí síti

- První experimenty
  - V duchu „Když chceme vědět co je divné, pojďme zjistit co je normální“
  - 25/TCP – velikost, počet komunikující protistran
  - První náznaky kudy se vydat
    - Velikost nezajímavá
      - Řádově 10-100MB útoky vs. mnoho GB legitimní aktivity
    - Počet serverů v řádu stovek nebo tisíců
      - Při použití velkých poskytovatelů běžně 100 serverů

Bad flows

- Flow classification
- Vychází z myšlenky „Jak vypadá skenování sítě“
- Netflow-like data rozdělena do 4 kategorií:
  - Communication – běžná komunikace
  - Inbound only – komunikace zahozená routerem
  - Outbound only – komunikace zahozená serverem
  - Answered – server odpověděl, ale router ji ihned ukončil
- Outbound only a answered jsou *bad flows*

Bad flows

- Fungující metoda
  - Téměř 2 roky studia chování útoků
  - Ve spolupráci s uživateli
    - Podezřelá aktivita předána CSIRT.CZ
    - Prosba o zjištění nákazy nebo vysvětlení divného chování
    - Cenná pozorování
  - Cca 15 odhalených nákaz (nepravidelné kontrolování)
- Nevýhody
  - Pouze 10 denní okno
  - Ruční spouštění
  - **Žádná historie**
    - Podezřelé chování dnes nebo dlouhodobě
    - Opakované divné chování
    - Malá, ale dlouhodobá aktivita

Analýza nákazy – stage 2

- Cíle
  - Posun od analýzy hrubých dat k analýze historie a historického vývoje chování klienta
  - Nutnost podivné chování hledat automatizovaně
- Prostředky
  - Přiřazování **skóre** jednotlivým klientům
  - Archivace výsledků
  - … smlouva

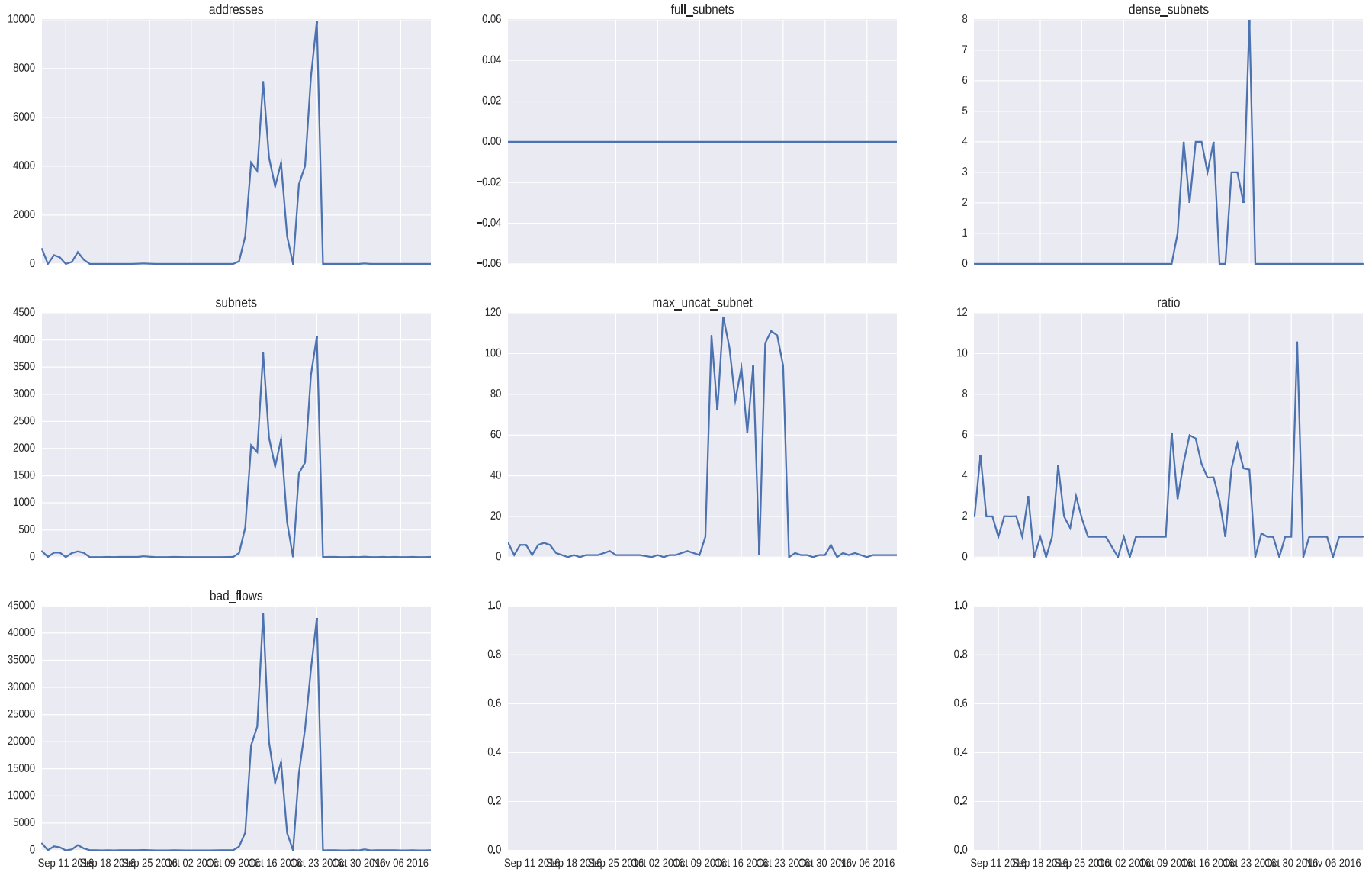
Jaká data můžeme archivovat

- Po 10 dnech pouze anonymizovaná data
- Nelze spojit klienta se serverem
- „Kdo klepe na vrátka“
  - (~~klient~~, zdrojový port, protokol, cílový port, cílová adresa)
  - „Je jedno jestli byl útok na Pepu nebo na Karla, server je zlý“
- Opačný směr
  - (klient, zdrojový port, protokol, cílový port, ~~cílová adresa~~)
  - „Pepa komunikoval“

Analýza historie

- Data
  - Předzpracovaná data z bad flows analýzy
  - (klient, ~~zdrojový port, protokol, cílový port, cílová adresa~~)
- Charakteristiky
  - Hodnoty, které popíší chování klienta
  - Vypovídající pro určení „podezřelosti“
    - S kolika adresami komunikoval
    - Kolik /24 subnetů bylo skenováno a jak „hustě“
    - Počet bad flows
- **Využití expertních poznatků**

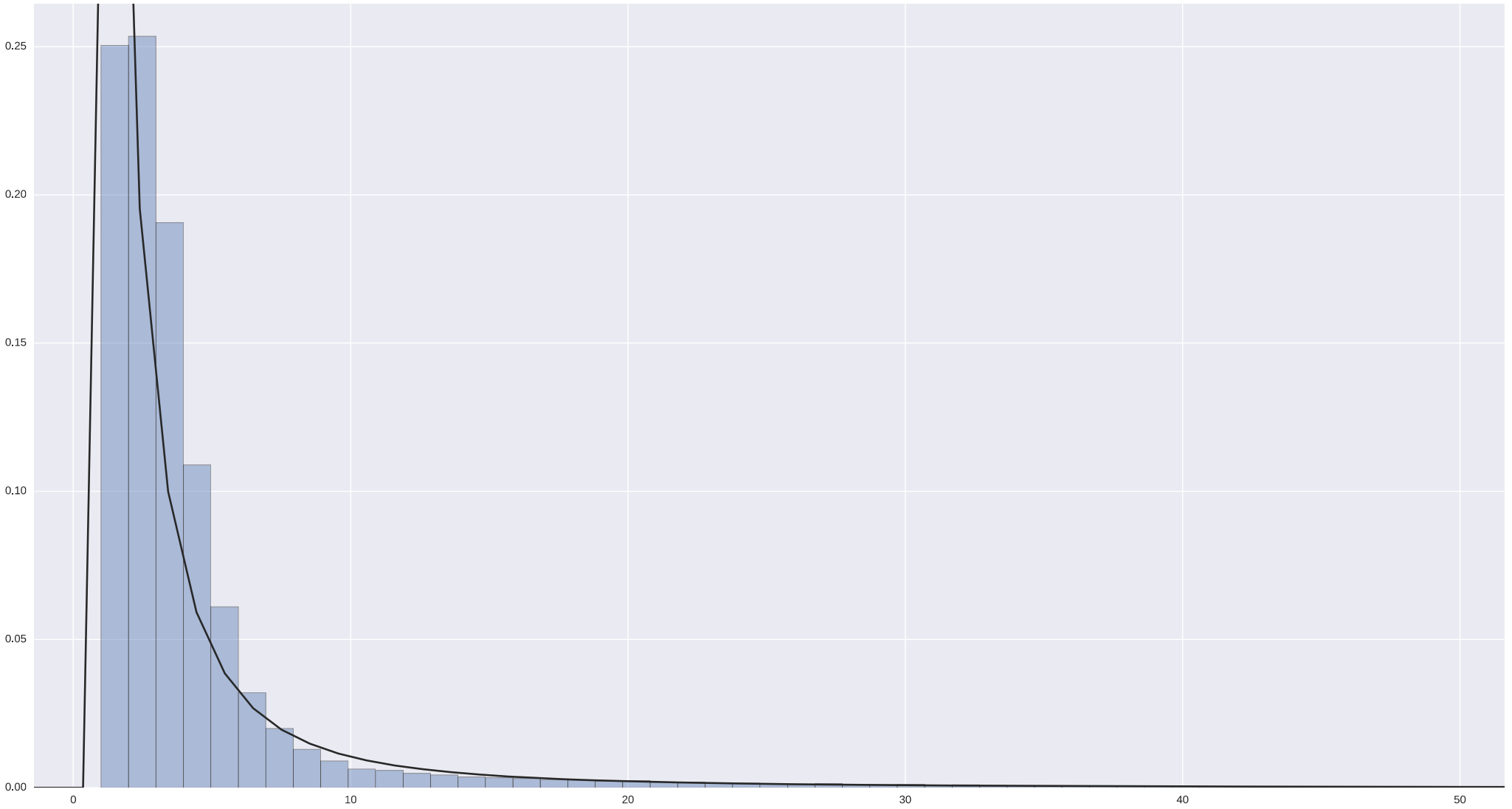
Ukázka útoku



Skóre klienta

- Jednoduché techniky ML – statistické modely
- Modelování normálního (většinového?) chování
- Pro každou charakteristiku vlastní model
  - Normální je neútočit
  - Histogram
  - Pareto distribuce (Heavy-tail distribuce) – podobná exponenciálnímu rozdělení
  - Model: fitování odpovídající distribuce na data
- Výpočet skóre
  - Z modelu získáme pravděpodobnost výskytu dané hodnoty – tzv. likelihood
  - Prozatím záměrně nepřesný předpoklad: charakteristiky jsou nezávislé
  - Skóre je tedy součet „odchylek“ od běžného chování

Model pro počet adres



Vize do budoucna

- Analýza historie
  - Dobrý nápad jak se posunout z mrtvého bodu
  - Funguje velmi dobře
  - Stále začátek dlouhé cesty
- Naše situace
  - Nedostatek dat pro hledání komunikace s C&C centry
  - Můžeme zachytit pouze útoky
  - Jaké typy útoků můžeme snadno zachytit?

Vize do budoucna

- Nejbližší cíle
  - Silnější propagace expertních znalostí do modelů
  - Přidání dalších analýz a zdrojů dat
- Co nejvíce se zaměřit na automatizaci
  - Automatické varování CSIRT
  - Automatické varování uživatele (?)
- Získávání další expertních znalostí

Děkuji za pozornost

Robin Obůrka • robin.oburka@nic.cz