PCAP, flow, metadata: Musím si vybrat?

OD DAT KE ZNALOSTEM II.

Analýza síťového toku

28.11.2019

Martin Půlpán
CEO net.pointers s.r.o.

martin.pulpan@pointers.cz

Trust ❖ Expertise ❖ Sensitive approach

POMÁHÁME VÁM BEZPEČNĚ FUNGOVAT

net pointers communications

Fidelis Cybersecurity

ARBOR NETWORKS

NETSCOUT

paloalto NETWORKS

splunk>

TOVEK

Detekce a reakce na bezpečnostní incidenty v síti i koncových bodech

- Advanced Persistent Threads protection (APT)
- Data/Network visibility and monitoring
- Automated Detection and Response (ADR)

Obrana proti DDoS útokům

- SP/ISP and top Enterprise
- Security Managed Service (MSSP)

Služby s přidanou hodnotou a technická podpora

- On-site cyber-attack simulation, STRESS/PEN testing
- Cloud Based Automated Detection and Metadata Correlation
- Security systems Evaluation and Integration





Security Assessment Vulnerability Testing | Detection Prevention | Technology Deployment Support | Incident Response Prediction | Training

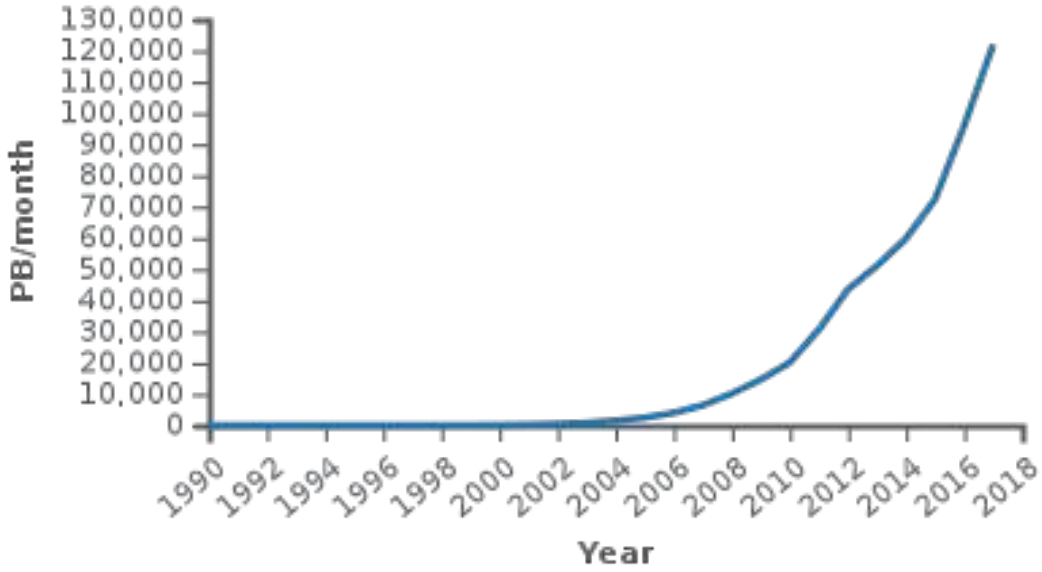
AGENDA

- Současné komunikační prostředí
- Možnosti identifikace
- Různé pohledy, různé zdroje
- Příklad na konec

KOMUNIKAČNÍ PROSTŘEDÍ

INTERNETOVÝ PROVOZ

- Za posledních 10 let se internetový provoz zdesetinásobil
- Kapacita koncových přípojek roste podobným tempem
- Růst kapacity cloudových uložišť je ještě intenzivnější



Trendy

- Rychlé připojení koncových stanic
- Rychlé a dostupné cloud služby nabízející virtuální systémy a aplikace
- Rostoucí množství "chytrých zařízení"

IDENTIFIKACE, MONITORING CHOVÁNÍ

Každá identifikace začíná popisem **komunikace**, která je definována **provozem v lokaní/WAN síti** a nějakým serverem, koncovou stanicí či místem v internetu…

Efektivní monitoring vychází z rychlé a přesné identifikace sledovaného provozu.

Provozu i komunikačních kanálů stále přibývá ….

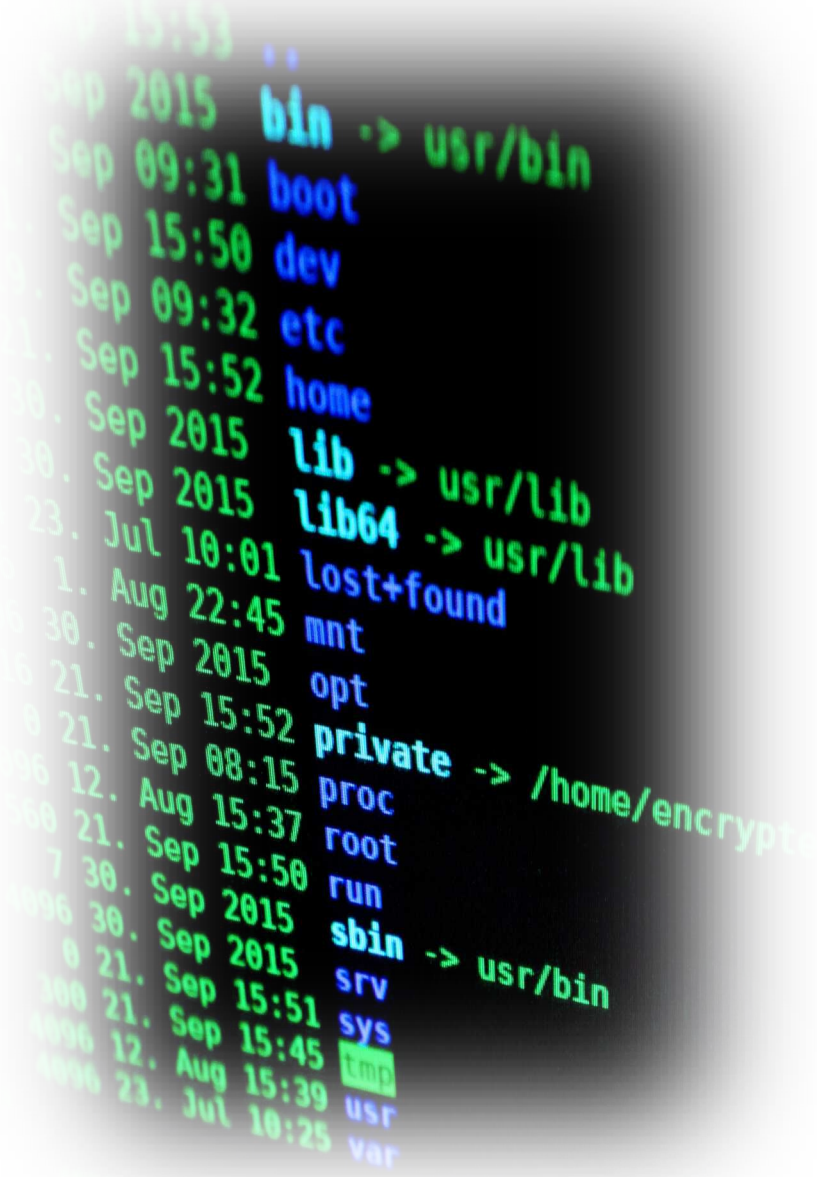
JAKÉ MÁME MOŽNOSTI

- **Data Flow**
- **PCAP**
- **Metadata**

Data Flow - vysoká rychlost zpracování, nižší přesnost identifikace a potvrzení problému

PCAP - nízká efektivita zpracování způsobena vysokými nároky na množství zpracovaných dat

Metadata - vyšší efektivita zpracovaných dat, akceptovatelná přesnost a flexibilita v zacílení zájmu



DATA FLOW

Flow-based standardy:
NetFlow, jFlow, sFlow, IPFIX a další

Zdroje data flow:
- Síťové prvky (routery, switche)
- Specializované sondy, které vzorkují provoz, poskytují statistiky, bihaviorální analýzy
- Další bezpečnostní nástroje, které pracují s NetFlow a jeho klony

Výhody:
- Velmi efektivní přehled o tom, co se v síti děje:
  - kdo s kým komunikuje
  - na jakých portech, druh komunikace
  - jak dlouho a jak často
  - kolik dat bylo přeneseno
- Nízké nároky na ukládání historických dat
- Rychlé hledání
- Široká vizibilita přes celou síť

Nevýhody:
- Poskytuje pouze trendová data a změny oproti historii
- Často se používá pouze vzorkování dat, ne plná analýza
- Kvalita NetFlow dat se velmi liší a je závislá na zdroji dat
- Nedostatečné pro Zero Day útoky, APT hrozby a moderní malwarové útoky (ransomware)
- Nedetekuje latentní šíření a skrytou komunikaci
- Nevidí do obsahu komunikace

PCAP

Packet capturing:
Libpcab, pcap-ng – binární formáty s časovým označením a blokovým dělením

Zdroje PCAP:
- Aplikační programy (Wireshark, snort, TCPdump)
- Specializované sondy, které zaznamenávají síťový provoz a poskytují vyhodnocovací nástroje
- Open-source knihovny

Výhody:
- Zaznamenává vše, co projde sítí pro pozdější kontrolu a forenzní vyšetřování
- Plný nebo filtrovaný záznam (úspora místa)
- Účinné proti APT a Zero Day útokům
- Dostatečné pro forenzní vyšetřování – lze přesně vysledovat vektor útoku, zmapovat chovaní uživatelů, detailně analyzovat vše, co prošlo sítí

Nevýhody:
- Extrémní nároky na ukládání dat (100MB ETH – 1TB – 22 hodin, 10GB ETH – 14 minut)
- Složité je něco najít – vyžaduje to bigdata přístup – hledání jehly v kupce sena
- Nelze použít na realtime monitoring
- Problém je šifrovaný provoz
- Ukládání delšího časového úseku je velmi finančně náročné

METADATA

Metadata
Obsahují informace o přenášených datech, popisují co bylo přeneseno, jakým způsobem, jaký byl charakter informací, co bylo obsahem, odkud, kam a kdy probíhala komunikace a další atributy získané z přenášených relací.

Zdroje METADA:
- Specializované sondy, které vytvářejí vlastní metadata (hw nebo virtuální)
- Další síťové bezpečnostní systémy
- End-point agenti
- Log management systémy
- Open-source nástroje
- a další zdroje z komunikačního prostředí

Výhody:
- Výrazně menší nároky na ukládání historie komunikace
- Detekce v reálném čase – lze použít pro aktivní mitigaci a ADR
- Rychlé hledání a analýzy
- Účinné proti APT a Zero Day útokům
- Dostatečné pro rychlé forenzní vyšetřování – lze kombinovat se selektivním PCAPem
- Lze ukládat i delší historii (i několik let)

Nevýhody:
- Neexistuje standard – kvalita metadat z různých zdrojů se může výrazně lišit
- Omezená práce s obsahem komunikace
- Zašifrovaná data bez dekryptoru nelze tagovat
- Pro složitější a bohatší tagování je třeba využít externí inteligenci a sendboxing, což může zvýšit náklady

VYTVOŘME SI METADATA ...

A) můžeme je vytvářet z provozu na sítích, ke kterým máme fyzický nebo rádiový přístup

B) máme k dispozici API rozhraní k datům, která jsou zpracovávána jinými systémy

C) Prostřednictvím API rozhraní cloudových služeb

D) Na základě informací z monitorovacích systémů (face recognition kamery a jiné systémy)

AGREGOVANÁ METADATA

Agregační platformy

- Propojují jednotlivá metadata
- Selektivní analýza
- Vytvářejí vlastní agregovaná metadata
- Používají AI pro detekci určitých paternů
- Modelace vzorců chování
- Vytváření archivů pro další možné zpracování

RŮZNÉ POHLEDY, RŮZNÉ ZDROJE

Flow data z IP sítě
IP1 -> IP2 sport 123, dport 234
- snifování provozu

IMEI -> IP1 , location A
- info od telecom providera

Location A s provozem portu 123

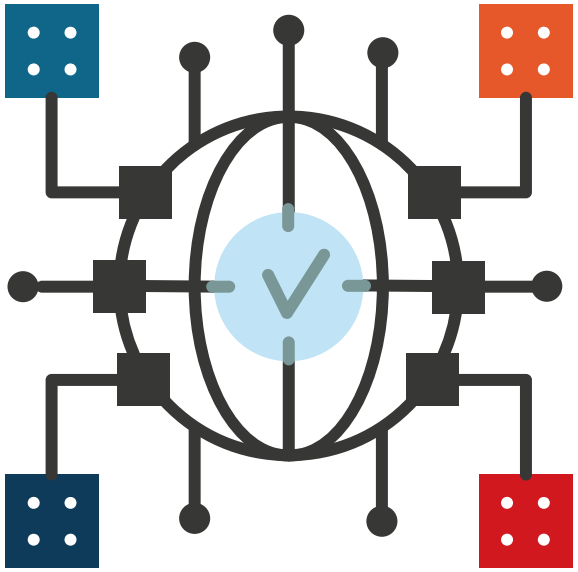
Uloženo v DB pro vyhledávání

Metadata popisují sledovaný provoz ze dvou nezávislých zdrojů

DATA SCIENCE

Co je to?

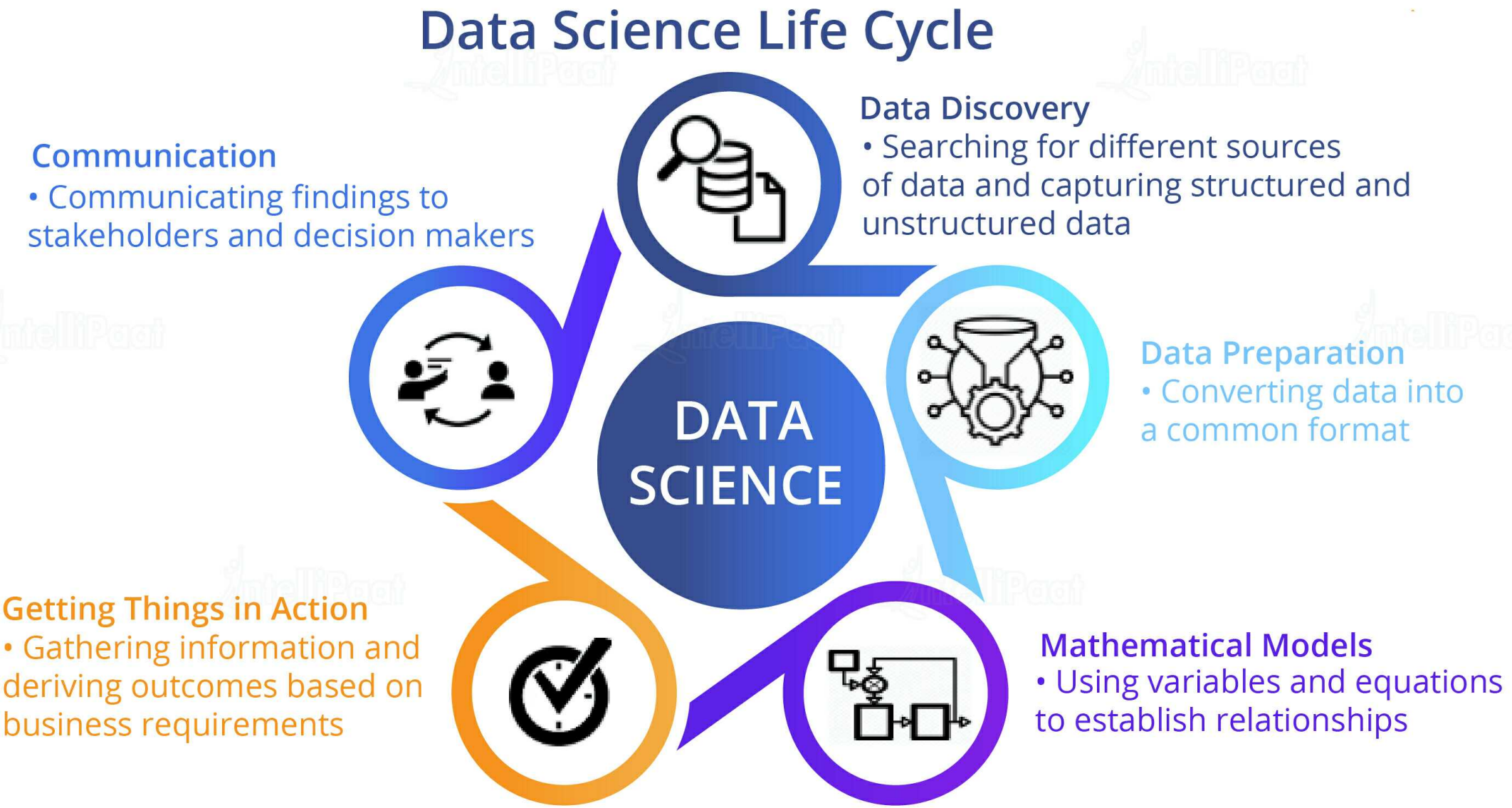
- Moderní přístup pro vyhodnocování metadat všech typů
- Vyžaduje kvalitní metadata – vysokou vizibilitu
- Nutností je dostatečně velký časový vzorek
- Vytváří základní model chování a následně hledá anomálie
- Bigdata přístup
- Model lze postupně zpřesňovat
- Zvýšení přesnosti detekce (nad 80%)



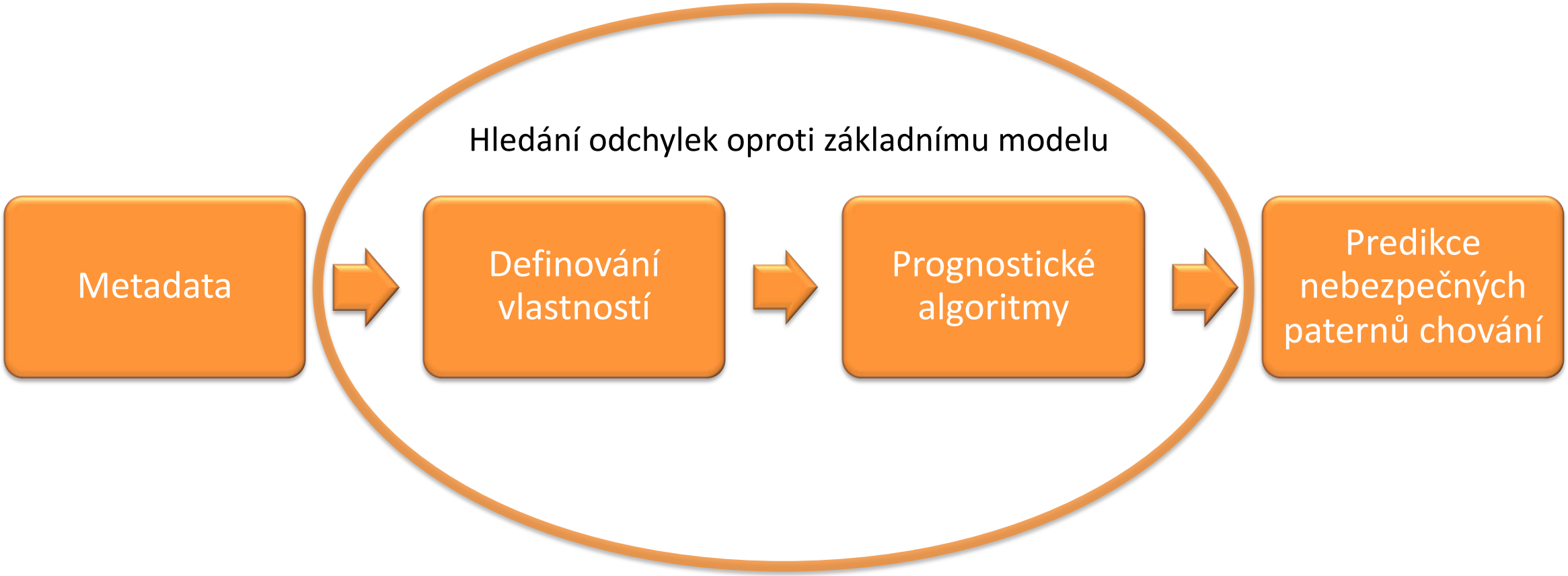
Proč data science?

- Příliš mnoho bezpečnostních alertů a limitované lidské zdroje
- Optimalizace nasazení senzorů
- Zpřesnění detekce
- Predikce možných rizik, ještě dříve než nastanou
- Vytvořený model lze pak snadno modifikovat

DATA SCIENCE



DATA SCIENCE



Nastavení BASELINE modelu

JE TADY NĚJAKÝ PROBLÉM ?

Máme přístup k fyzické síti

- 85% veškerého provozu je **šifrovaný**

Co s tím:

- v případě, že nás zajímá interní obousměrný provoz

->> **nemám problém**

- v případě, že chci sledovat obousměrný internetový provoz z lokální sítě

->> můžu mít problém

Použiji techniku hackerů (man in the middle), kde za pomoci specializovaných zařízení se vůči klientovi tvářím jako legitimní server *(problém je pinning certifikátů)*

Nemáme přístup k síti

- rozhraní k monitorovaným službám musí zpřístupnit provozovatel služby, což může být komplikované získat (zejména u zahraničních subjektů mimo EU).

ŠIFROVANÝ PROVOZ

Jedním z úkolů kybernetické bezpečnosti je zabezpečit integritu spojení - nemožnost vstoupit do probíhajícího provozu

Paradox – který nám komplikuje situaci v případě monitoringu

Má vůbec smysl monitorovat šifrovaný provoz ?

Rozhodně MÁ

Samotné navazování šifrovaného spojení mi prozradí množství užitečných informací:

- parametry šifrování
- parametry použitých certifikátů
- informace o komunikujících klientech
- informace vycházející ze samotných flow dat (IP, porty, poloha klienta a podobně)
- informace z dalších jiných zdrojů

NEŠIFROVANÝ PROVOZ

- **Umíme rozkódovat encapsulovaný provoz** – na aplikační úrovni je provoz analyzován pomocí inspekce protokolů a formátů dat, které jsou známé a nezašifrované
- **Provoz si můžeme přiblížit** - zajímá nás poloha vzniku provozu, případně linková úroveň provozu, pokud máme technické a legislativní možnosti je mít
- **Můžeme se selektivně zaměřit na určitý provoz**, na určité komunikační porty, protokoly, obsah, formát či jiné parametry
- **K selektivně definovanému provozu** si mohu **připojit full packet capture** nástroj

CO TO ZNAMENÁ V REALITĚ ?

Zadání:

Zajímá mě provoz a jeho charakteristika-profil (četnost, hodina, den v týdnu) pro telefonní číslo +420 xxx xxx xxx na email franta.vomacka@seznam.cz

Vzhledem k tomu, že se jedná o dlouhodobé sledování, nemohu zpracovat veškerý provoz v dané lokalitě

Řešení: Vytvoříme si METADATA

ŘEŠENÍ

Situace:
- snažíme se popsat dlouhodobě chování vybraného uživatele a nemůžeme shromažďovat u daného poskytovatele veškerý provoz

Řešení:
- musíme vybrat a vydefinovat **určitou charakteristiku provozu** a budeme zkoumat a popisovat **jenom tuto část provozu** – vytvořili jsme si tak svoje **metadata**

- zajímají nás veškerá telefonní čísla, které posílají email do lokality BTS123 atd., filtry postupně zpřesňujeme

- analýzou zadání jsme zúžili množství dat, které potřebujeme ukládat, abychom následně z těchto dat vybrali jenom ta, která nás skutečně zajímají

JAK POSTUPOVAT

- **Definovat si záměr:**

- co se snažíme naší aktivitou docílit
- může to být samotné hledání určitých informací
- může to být doplňková informace, která nám obohatí jiný typ hledání

- **Rozmyslet si konkrétní parametry hledání, výpočty nad hledanými daty**

- **Délka trvání**

- **Zvážit možnosti nástroje, který na hledání použijeme**

VYHODNOCENÍ PŘÍKLADU

Na základě porovnání množství provozu procházejícím zmíněnou lokalitou jsme zjistili:

- emailová služba je méně používána než tomu bylo v minulosti, což znamená:

  - Jenom 5% provozu byla e-mailová komunikace
  - 95% byl jiný provoz, který nás nezajímal = úspora prostoru pro další hledání

Závěr:

Velice jednoduchý příklad **má za cíl demonstrovat, co metadata jsou** a jak je k nim potřeba přistupovat

DOPORUČENÍ

Co je vaším cílem?

- Je potřeba si správně definovat cíle monitoringu, kybernetické obrany …
- Zvolit správnou strategii, která je rychlá, efektivní a technologicky nejméně náročná, založená na metadatech v kombinaci s data science přístupem.
- Použít vhodné technologické nástroje, které generují hodnotná metadata v kombinaci se selektivním PCAPem.

OTÁZKY?

Děkuji za pozornost!

martin.pulpan@pointers.cz