Využití JupyterLab pro interaktivní analýzu nejen kyberbezpečnostních dat

Seminář o bezpečnosti sítí a služeb 2024

Milan Čermák

CSIRT-MU, Masarykova univerzita, Brno

MUNI
CSIRT-MU

Whoami

LinkedIn – https://www.linkedin.com/in/cermmik/

Google Scholar – https://muni.cz/go/cermak-scholar

- Analytik a výzkumník v oblasti analýzy kyberbezpečnostních dat
- Vedoucí analytického oddělení kyberbezpečnostního týmu MU – CSIRT-MU
- Řešitel národních a mezinárodních výzkumných projektů spolupracující s analytiky a vyšetřovateli trestné činnosti, NÚKIB a mnoha dalšími výzkumnými týmy a společnostmi
- Ph.D. na Fakultě informatiky Masarykovy univerzity (obhájeno v roce 2020)
- Vyučující předmětu Network Forensics a kurzu Network Detection and Response (již brzy)
- Odborné zájmy – forenzní analýza digitálních a síťových dat, tvorba analytických systémů využívající moderní technologie (viz Granef) a web hacking

Pokud vás cokoliv z toho zajímá, tak si s vámi rád popovídám u kávy 😉

Každý větší incident by měl být spojený s analýzou dostupných dat, tak abychom zjistili všechny dopady

- Typicky provádíme stejné kroky, jen s menší obměnou na základě podstaty incidentu
- Velmi často nemáme dostatek času vše detailně prozkoumat a ověřit všechny hypotézy

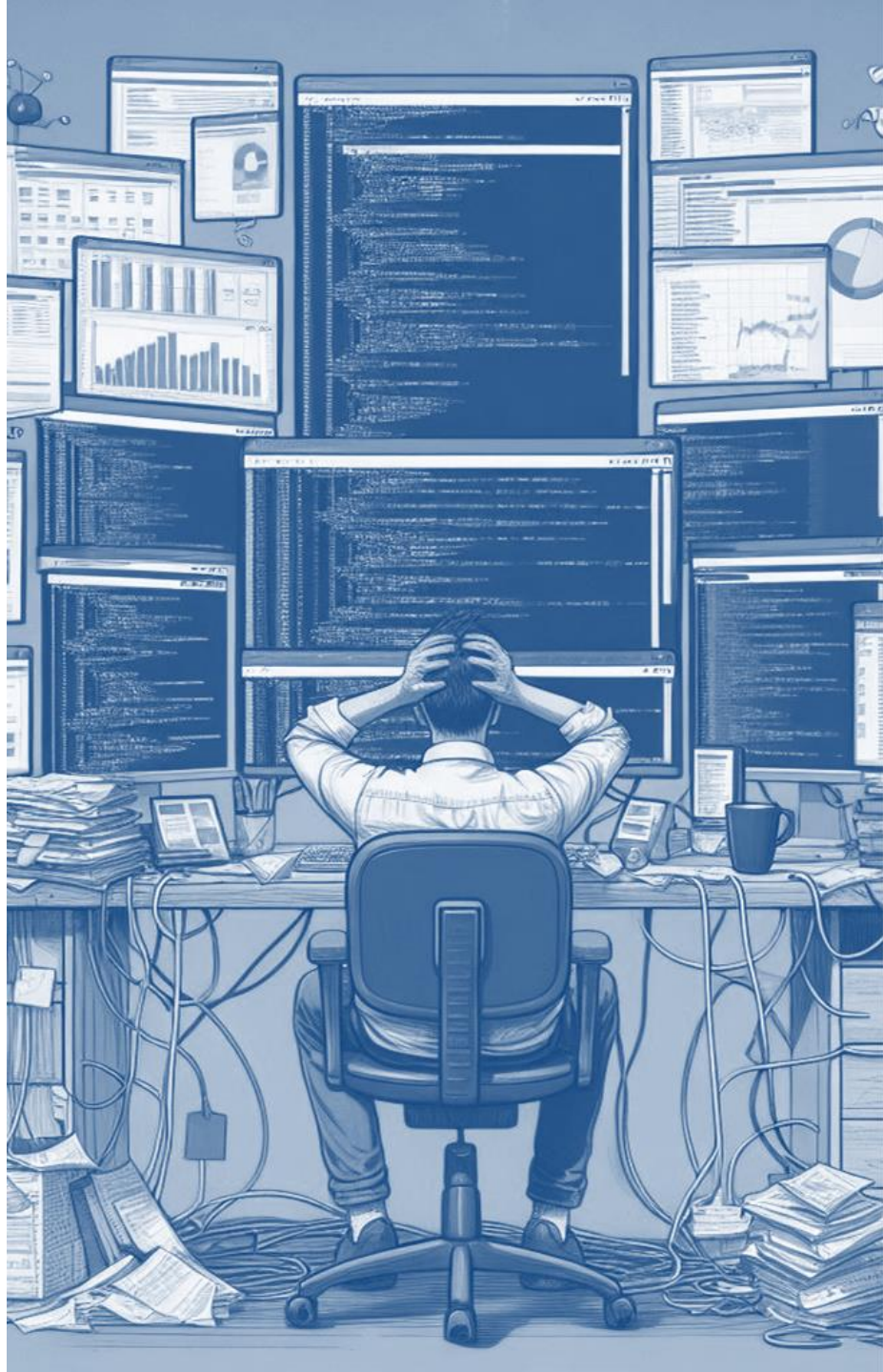
Typický průběh analýzy dat

Scénář: Server v naší síti skenuje své okolí a hádá hesla na RDP

1) Detekování incidentu (SIEM)
2) Získání dalších informací (SIEM)
3) Informace o serveru (asset management)
4) Potvrzení incidentu (síťové toky)
5) Analýza souvisejícího provozu (síťové toky)
6) Hledání původu útoku (síťové toky)
7) Analýza chování útočníka (systémové logy)
8) Dohledání informací o útočníkovi (CTI)
9) Analýza nahraného SW (externí nástroj)
10) Analýza procesů strojů s RDP (symon)
11) Korelace a tvorba časové osy (poznámky)
12) Dokumentace a uložení (poznámky)
13) Sdílení poznatků (CTI)

Těžký život analytika

- Často se opakující rutinní analytické postupy
- Dlouhé čekání na výsledky (obzvlášť při dotazech na delší časové úseky)
- Používání různých analytických nástrojů s rozdílnou strukturou dat a odlišným uživatelským rozhraním
- Složitá korelace dat z různých zdrojů (jiné formáty dat, nesynchronizované časové známky, …)
- Velké množství analytických úkolů s požadavkem na co nejrychlejší řešení
- Obtížné sdílení know-how s ostatními analytiky
- Potřeba zpracování výsledků a poznatků do jednoduše pochopitelné formy



V datové analytice se typicky používá nástroj Jupyter Notebook, tak proč jej nevyužít i pro naše potřeby?

- Snadné programování v jazyce Python s využití knihoven pro zpracování a vizualizaci dat
- Univerzální řešení umožňující jednoduché sdílení prováděných analýz

Jupyter Notebook a JupyterLab

Jednoduché prostředí umožňující psát analytický kód v jazyce Python, dokumentovat v Markdown a zobrazit výsledky v textové i grafické formě. Vše v rámci jednoho souboru, který je možné jednoduše sdílet.

Odkaz na projekt: https://jupyter.org/

- JupyterLab je webové grafické prostředí umožňující spouštět a zobrazit Jupyter Notebooky (víceuživatelské prostředí je realizováno pomocí JupyterHub)
- Pro jednoduchou instalaci na serveru je možné využít projekt The Littlest JupyterHub, pro nasazení v Kubernetes projekt Zero to JupyterHub
- Typicky využívané knihovny: Pandas, Numpy, Plotly, Bokeh
- Pěkný úvod do Jupyter Notebooků nabízí série článků od Pavla Tišnovského na Root.cz: https://www.root.cz/clanky/jupyter-notebook-nastroj-pro-programatory-vyzkumniky-i-lektory/

Výhody Jupyter Notebooků

- Python ve spojení s vhodnými knihovnami umožňuje rychlý vývoj metod pro zpracování heterogenních dat a jejich korelaci
- Možnost jednoduše změnit proměnnou nebo blok kódu a spustit jej znovu při zachování předcházejících výsledků výpočtu
- Lze vytvořit analytický playbook se všemi kroky pro danou analýzu s možností dokumentování jednotlivých kroků
- Většina databází, služeb a nástrojů má dostupný Python modul nebo umožňuje dotazování skrze API
- Při správném naprogramování je možné provést celou analýzu na jedno kliknutí a nechat běžet i několik dní
- Jupyter Notebook je možné jednoduše sdílet včetně uložených výsledků analýz

Number of observations that no IP not communicated at all

In [5]:
```
noncommunicative_obs = df[df.isnull().all(axis=1)].index
noncommunicative_obs = noncommunicative_obs.shape[0]
print("Noncommunicative Observations: ", noncommunicative_obs)
print("Percentage of totoal Observationss: ", noncommunicative_obs/df
```

```
Noncommunicative Observations:  1329
Percentage of totoal Observationss:  15.17123287671233
```

Number of IPs that have not communicated at all

In [6]:
```
noncommunicative_ips = df.columns[df.isnull().all()]
noncommunicative_ips = noncommunicative_ips.shape[0]
print("Noncommunicative IPs: ", noncommunicative_ips)
print("Percentage of totoal IPs: ", noncommunicative_ips/df.shape[1]*
```

```
Noncommunicative IPs:  703
Percentage of totoal IPs:  1.07269287109375
```

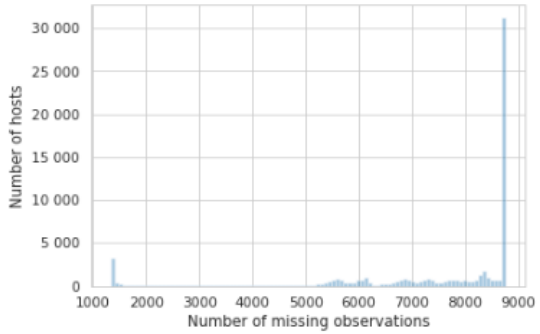
Distribution of the number of missing observation over the whole network

In [7]:
```
sns.set_style("whitegrid")
sns.set_context("paper", font_scale=1.2 )

# Plot the data
sns_plot = sns.distplot(df.isnull().sum(),bins=100,hist_kws={'cumulat

# Adjust plot properties
sns_plot.set(ylabel = "Number of hosts", xlabel="Number of missing ob
ylabels = ['{:,.0f}'.format(y).replace(',', ' ') for y in sns_plot.ge
sns_plot.set_yticklabels(ylabels)

#sns_plot.figure.savefig('./fig/NA_distr_big.svg')
plt.show()
```

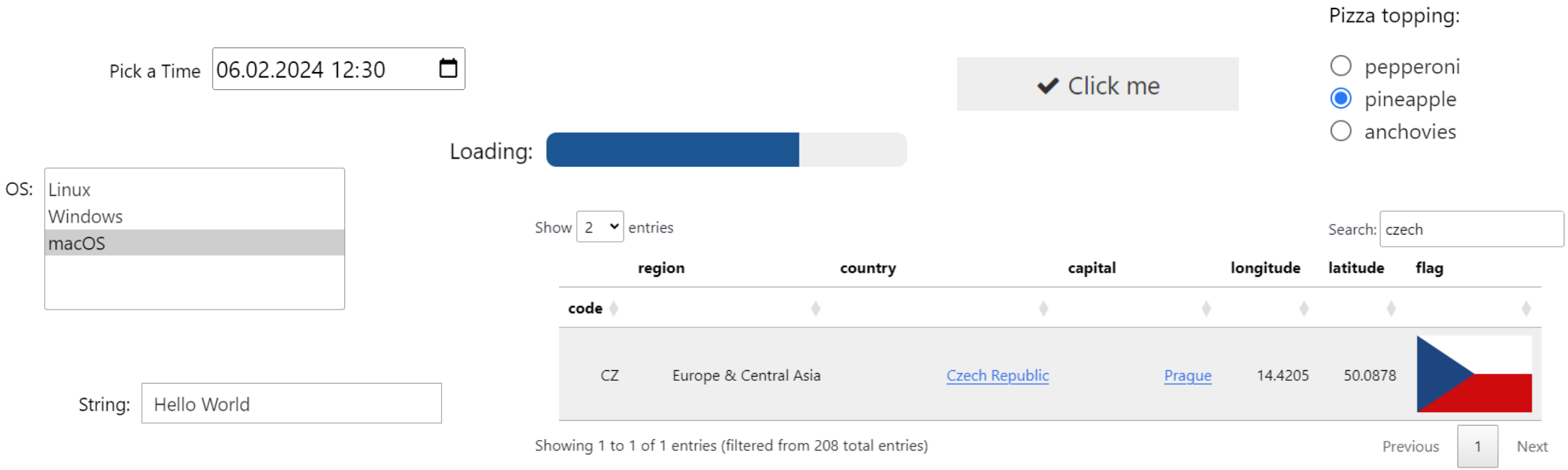


Teoreticky máme nástroj řešící většinu problémů současných analýz… jen uživatelská přívětivost pokulhává

- V rámci analýzy je potřeba vždy nějakým způsobem měnit kód (byť jen hodnotu proměnné)
- Neznalý člověk se v celém prostředí nevyzná a může být zmatený z některých funkcí

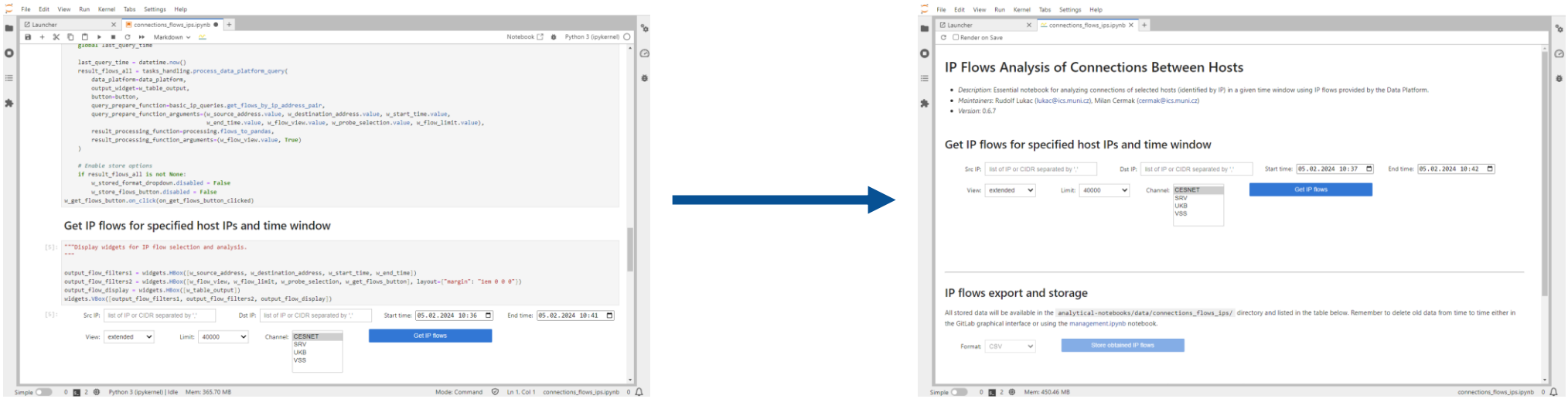
Jupyter Widgets a interaktivní prostředí

- Široká nabídka interaktivních widgetů, které nám nabízí možnost interakce v Jupyter Notebooku bez potřeby zasahovat do kódu
- Spojení output a button widgetů nám umožňuje jednoduše měnit již zobrazený obsah
- Pro zobrazení tabulek doporučuji využít navíc knihovnu itables



Voilà aneb zbavme se zobrazení kódu

- Rozšíření Voilà umožňuje spustit Jupyter Notebook a zobrazit pouze jeho výstupy (dokumentace v Markdown, widgety a jakékoliv další výstupy provedených funkcí)
- V kombinaci s Jupyter Widgets nám umožňuje vytvářet interaktivní analytické dashboardy a playbooky, které jsou jednoduše použitelné i pro méně znalé uživatele
- **Tip:** Prohlížeč souborů v JupyterLab je možné nastavit tak aby defaultně zobrazoval Jupyter Notebook pomocí Voilà preview namísto zobrazení Notebooku s kódem



Interaktivní Jupyter Notebooky nám skvěle pomáhají při analýze incidentů, jen jsme museli vyřešit pár komplikací

- Vše je dostupné v rámci jednoho GitLab repozitáře, který si uživatel v JupyterLab stáhne
- Pro námi používané služby máme vytvořeny konektory, přičemž uživatel musí pouze vyplnit autentizační údaje v konfiguračního souboru

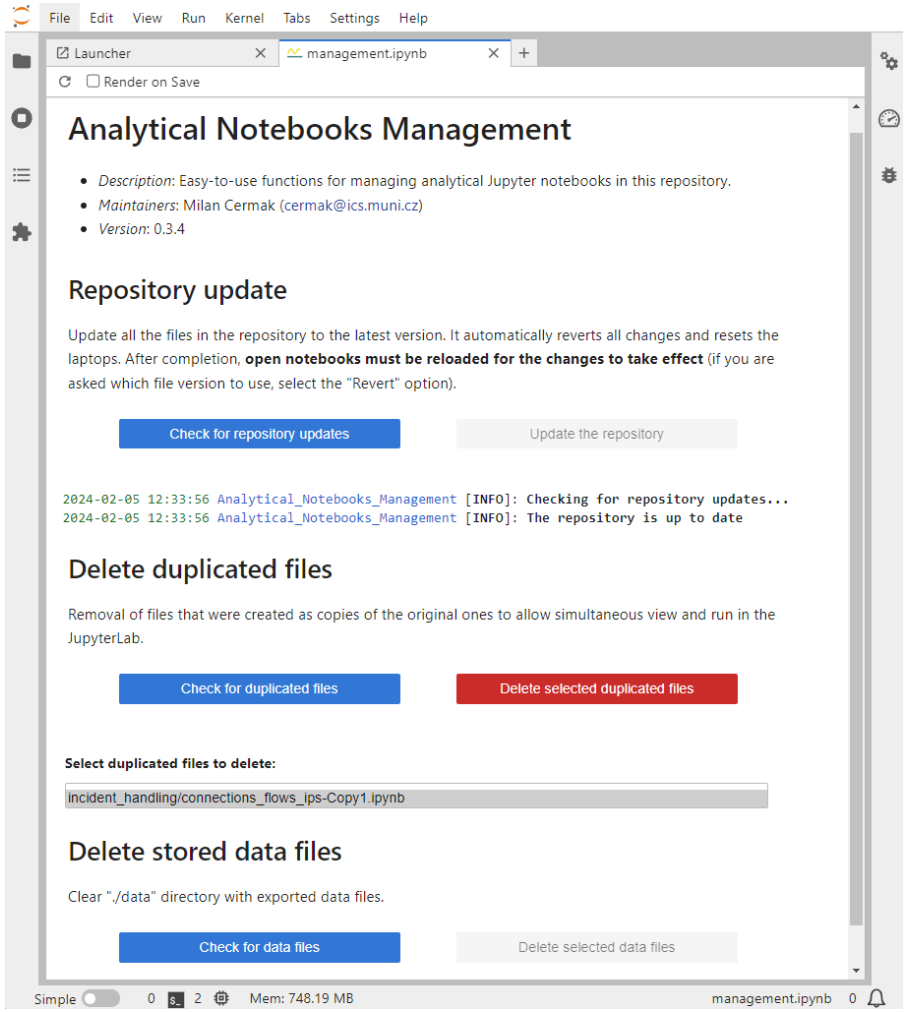
Správa pomocí Management Notebooku

Problémy v použitelnosti

- Aktualizace pomocí `git pull` je zbytečně složitá
- Jupyter Notebook může běžet jen v jedné instanci
  → lze obejít pomocí vytváření kopií (duplikování)
- Potřeba průběžného mazání exportovaných dat

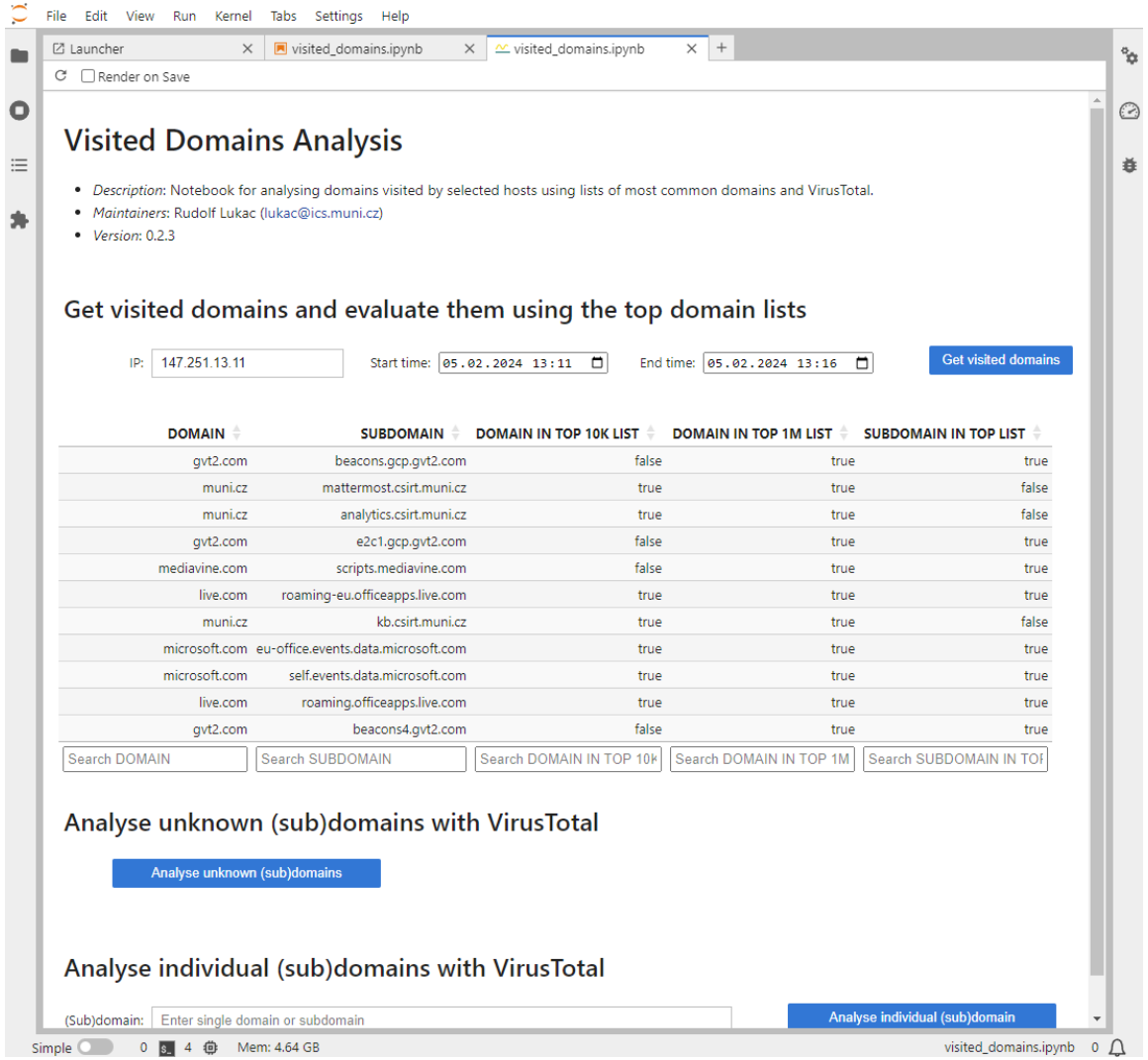
Management Jupyter Notebook

- Uživatel jen jednou stáhne celé prostředí a dál už využívá poskytnuté funkce
- Automatické řešení nekonzistence repozitáře pokud uživatel udělá nějaké změny



Typický analytický playbook

- Jupyter Notebook obsahuje vše co potřebuje analytik při řešení incidentu
- Vše řízeno přes jednoduché vstupy (text, časové rozmezí, selektor, …) a tlačítka spouštějící danou akci
- Každý krok lze vysvětlit a zdokumentovat
- V rámci jednoho úkonu se mohou korelovat data z více zdrojů
- Možnost propojení externích služeb (např. VirusTotal) bez nutnosti přepínání mezi okny
- Velmi rychlý vývoj a prototypování



Shrnutí a pár poznámek na závěr

Shrnutí

- Spojení nástrojů JupyterLab, Jupyter Widgets a Voilà umožňuje vytvářet interaktivní analytické playbooky jednoduše použitelné i pro běžné uživatele
- Obecné řešení pro různé typy analýz nad všemi dostupnými daty (nemusíme se omezovat jenom na kyberbezpečnostní doménu)
- Pomocí knihoven a konektorů je možné napojit libovolnou službu nebo databázi
- Analytici kteří neumí programovat GUI aplikace mohou jednoduše tvořit interaktivní analýzy (stačí když mají dostupné ukázky již hotových Jupyter Notebooků)
- Kód z Jupyter Notebooků můžeme jednoduše přepsat do API a využít v jiných systémech

Pokud vás téma zaujalo a máte nějaké otázky nebo další podněty, tak budu rád, když se mi ozvete na cermak@ics.muni.cz