Ács Gergely
BME-HIT

Budapest, 2024.11.28

Kutatás.
Innováció.
Hatás.

HUN-REN
Magyar Kutatási Hálózat

https://hun-ren.hu

HUN-REN
Magyar Kutatási Hálózat

Robusztus Malware Detekció és Diffúzós Modellek Adatvédelme (MILAB)

https://hun-ren.hu

Tartalom

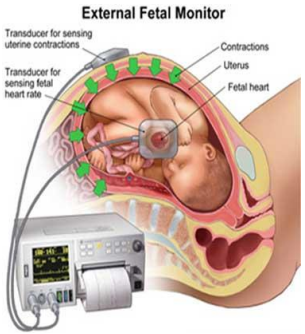
HUN-REN
Magyar Kutatási Hálózat

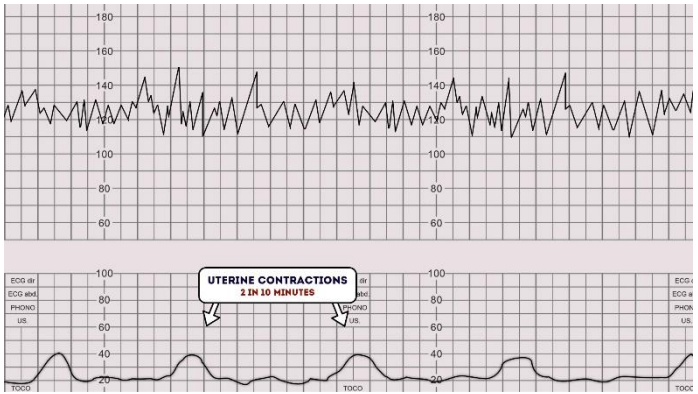
Célok

HUN-REN
Magyar Kutatási Hálózat

1. Robusztus Malware Detekció
   A. IoT malware-ek detekciója gépi modellekkel
   B. Különböző ML alapú megoldások összehasonlítása
   C. Ezen modellek robusztussága evasion és poisoning támadások ellen

2. Diffúziós modellek adatvédelme
   A. Cardiotocography (CTG) adatok generálása diffúziós modellekkel
   B. Ezen modellek adatvédelme





Igényelt erőforrás

HUN-REN
Magyar Kutatási Hálózat

1. Első két évben (2022.04-2024.05):
   - 2db g2.large
     - 8 GB RAM
     - 4 vCPU
     - 16 GB VRAM
     - 1 TB SSD

2. Utolsó fél évben (2024.05-2024.11):
   - 1 db g2.xlarge
     - 32 GB VRAM
     - 8 vCPU
     - 1 TB SSD

- Egy darab gépet egyszerűbb volt menedzselni, és két VM esetén az egyre jutó erőforrás túl kevés volt
- GPU használatot elsősorban a diffúziós modellek igényeltek

Környezet kialakítása

HUN-REN
Magyar Kutatási Hálózat

- Könnyű, de néhány dologra érdemes figyelni

1. /home → 1 TB SSD
   - Miért?
     - Könnyű az adatok migrációja VM-ek között (csak a user és group id-ra kell figyelni)
     - Python virtuális környezet sok helyet igényelhet felhasználónként
       - torch elég nagy, root partíció hamar betelik
       - TMPDIR=~/ beállítása a csomagok (torch) telepítésénél

2. Nem szabad kernelt frissíteni! (vagy rendszert)
   - A GPU-hoz speciális driver kell amit felülírhatnak a telepített csomagok
   - Az `apt install /usr/local/share/elkhcloud/nvidia-linux-grid-525_525.125.06_amd64.deb` gyakran megoldja (nálunk nem)

Környezet kialakítása

HUN-REN
Magyar Kutatási Hálózat

3. Érdemes egy darab VM-et használni (egyszerűbb)
   - Egy darab publikus IP-t kaptunk, a második gép csak az elsőről érhető el
   - VPN-nel is próbálkoztunk, de egy gép mindig kihasználatlan maradt

- Használat: burst-ös (az utóbbi időben inkább napi használat)
  - Első egy évben ritkábban, miután egy gép lett, utána intenzívebb
    - Két VM egyenként nem volt elég a feladatra

- Ha konkrét feladaton dolgozunk, akkor intenzívebb használat
  - A VM-en futtatunk és fejlesztünk is (Visual Code Remote SSH)
  - Futtatás: ssh + tmux
  - A GPU kihasználtságunk ilyenkor 50-70% körüli, huzamosabb ideig, attól függ hányan használják
    - 2-3 user használja összesen
    - egy user: 25%

- A publikus IP nagyon kényelmessé teszi az elérést

Adatok és eredmények

HUN-REN
Magyar Kutatási Hálózat

- Malware elemzés
  - Minták: A VM-en nem futtatható ARM és MIPS binárisok
  - Néhány ezer minta, összesen 2-300 MB
  - Feature vektorok tárolása numpy tömbökben
- Diffúziós modellek
  - Publikus CTG adatok
  - Kb. 500 beteg, előfeldolgozás után max 100 000 tanítóminta pytables-ben tárolva
- Minden adat az SSD-n tárolódik a /home könyvtárakban
- Diffúziós modellek: generált adat tárolása csv illetve képek formájában

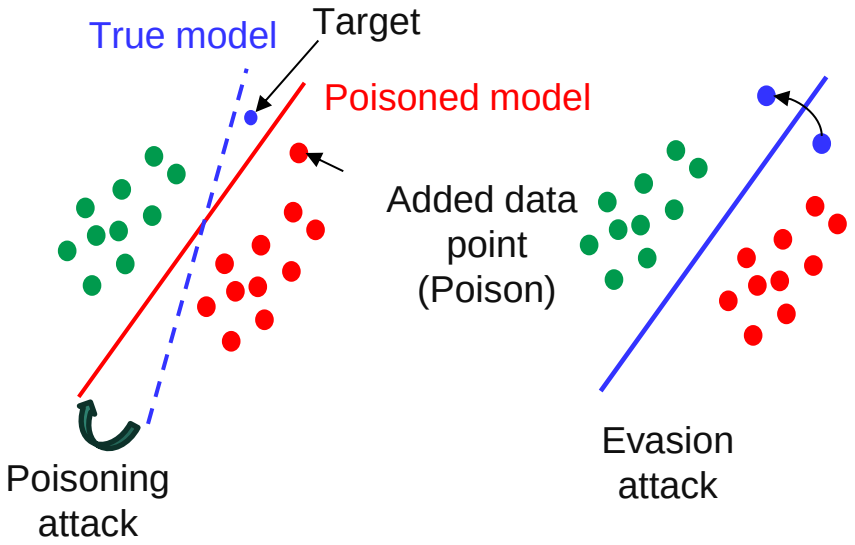
Elért eredmények

HUN-REN
Magyar Kutatási Hálózat

- Különböző ML alapú IoT malware detektorok összehasonlítása
  - Bináris → TLSH → Gépi modell (Random forest, Logistic regression, stb.)
  - >99.9%-os detekciós pontosság
  - Egyszerűbb modellek pontosabbak és kisebb költségűek (IoT-n fontos!)

- ML alapú malware detektorok evasion támadása
  - Különböző evasion technikák tesztelése
    - Benign programok hozzáfűzése
    - Gradiens alapú optimalizáció (módosítások számának minimalizásával)
  - Védekezés: Robusztus tanítás

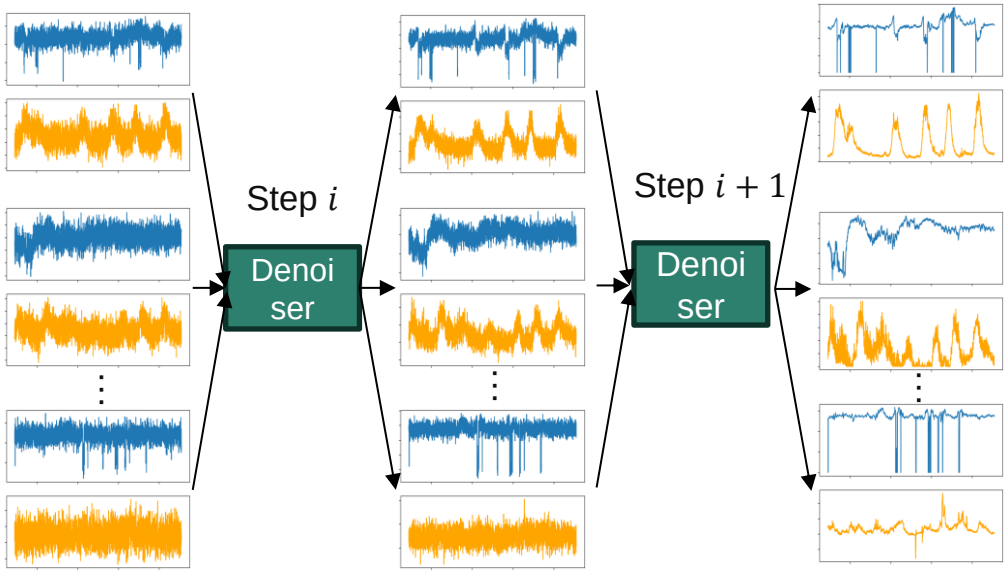
- ML alapú malware detektorok evasion támadása
  - Gradiens alapú poisoning (Witches' Brew)
  - Folyamatban…



Elért eredmények

HUN-REN
Magyar Kutatási Hálózat

- Diffúziós modellek
  - Latent Diffusion model for CTG generation
    - Diffúziós modellek + VQ-VAE
    - Undersampling → Diffúzió → Super-resolution
  - Diffúziós modellek adatvédelme
    - Sikeres membership támadás tervezése és kiértékelése
    - Differentially Private SGD for Diffusion



Publikációk

HUN-REN
Magyar Kutatási Hálózat

- N. Neubrandt, L. Buttyan, R. Nagy, Comparative analysis of similarity-based IoT malware detectors, under submission
- J. Sandor and R. Nagy and L. Buttyán, Increasing the Robustness of a Machine Learning-based IoT Malware Detection Method with Adversarial Training, WiseML'23
- E. Glocker, Privacy Analysis of Diffusion-based Cardiotocography Data Generation, TDK, 2024

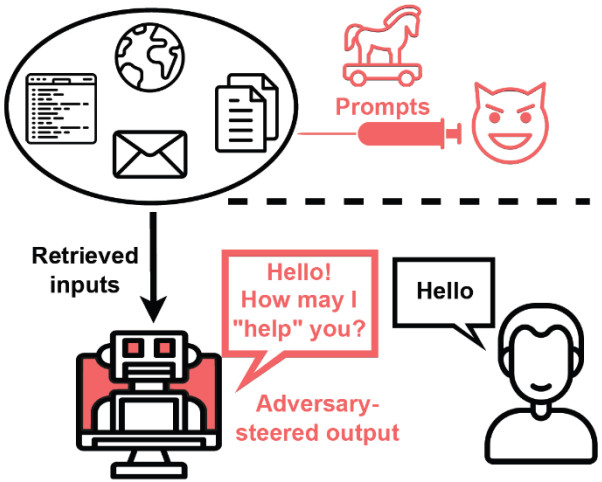
Jövőbeni tervek

HUN-REN
Magyar Kutatási Hálózat

- A kijelölt feladatokra elég a kapacitás, nem igényelnek komplex modelleket
  - Idősorok diffúziós generálása
  - Malware analízis

- Ezeket a feladatokat folytatjuk 2025 végéig

- Jövőbeni (mostani) projekt: LLM-ek biztonsága
  - Jóval több GPU memória kellene



Tapasztalatok

HUN-REN
Magyar Kutatási Hálózat

- Elégedettek vagyunk
  - Könnyen elérhető, gyors, nem volt fennakadás
- Felhasználói támogatás
  - Gyors, segítőkészek
  - Dokumentációk kellően részletesek
- Más rendszerekkel összehasonlítva
  - Nagyobb rendelkezésre állás mint más rendszereknél amit használunk
    - Soha nem volt fennakadás vagy kiesés
  - A számítási kapacitása csak kisebb feladatokra elég
- Min lehetne változtatni?
  - Rendszerfrissítés veszélyeire felhívni a figyelmet

HUN-REN

Magyar Kutatási Hálózat

https://hun-ren.hu