

Válasz

Dr. Krasznay Csaba Docens Úr

„Zsarolóvírus-támadások időkritikus felismerése gépi tanulás alkalmazási lehetőségeinek vizsgálatával centralizált fájl tárolási környezetekben”

című PhD értekezéshez írt

opponensi véleményére

Mindenekelőtt szeretném kifejezni köszönetemet Dr. Krasznay Csaba Docens Úrnak amiért dolgozatomat ilyen alaposággal elolvasta és értékelte. Részletes és körültekintő opponensi véleménye számomra megtiszteltetés, különösen azért, mert a bíráló nem csupán a dolgozat egyes részleteire, hanem annak koncepcionális és módszertani kérdéseire is kiterjed. A megfogalmazott kritikai észrevételeket köszönettel fogadom, és az érintett pontok tekintetében a bíráló megállapításaival egyetértek. Köszönöm, hogy opponensem több olyan kérdésre is ráirányította a figyelmet, amelyek a dolgozat egyes állításainak pontosítását, illetve bizonyos összefüggések árnyaltabb bemutatását tették szükségessé. Ezúton is szeretném külön megköszönni a dolgozat szerkezetére és felépítésére vonatkozó értékes javaslatokat, valamint mindazon meglátásokat, amelyek segítségével felismerhettem és tovább javíthattam a dolgozat azon részeit, ahol erre lehetőség nyílt. Úgy gondolom, hogy az észrevételek nagyban hozzájárultak a dolgozat és így a téziseim végső formájának és tartalmának kialakításához, a javaslatok pedig nem csupán jelen dolgozat vonatkozásában voltak hasznosak számomra, hanem további munkáim során is segítséget jelentenek majd. A dolgozathoz kapcsolódó kérdésekre az alábbiakban igyekszem teljesen és legjobb tudásom szerint reagálni.

1. „A szerző a dolgozatban részletesen tárgyalja a concept drift problémáját és a modell rendszeres újratanításának lehetőségét. Egy ténylegesen üzemelő rendszer esetében milyen mechanizmust tartana alkalmasnak annak felismerésére, hogy a normál felhasználói viselkedés vagy a ransomware-ek működése olyan mértékben megváltozott, amely már a modell újratanítását teszi szükségessé?”

Alapvetően különválasztanám a védekezési stratégiát annak függvényében, hogy az adat-exfiltrációs szakaszban szeretnénk a támadást észlelni, vagy az adatok titkosításának szakaszában. Az ARDS védelmi modell koncepciója szerint lehetőség nyílik a fájl kiszolgálóról történő adatolvasás és a tűzfalon keresztül a hálózathoz kilépő adatforgalom (upstream) időbeli mintázatai közötti tendenciális összefüggések felismerésére akár statisztikai, akár mesterséges intelligencia alapú módszerekkel. A titkosítási szakaszban történő felismerésnél az XGBoost gépi tanulási modell hatékony segítséget nyújthat. Véleményem szerint

egy produktív környezet esetén a következő események alapján indikálhatnánk az újratanítást:

- az antivírus szoftverekhez hasonlóan az újonnan megjelenő fenyegetettségek esetén automatikusan;
- a felhasználói visszajelzések alapján, elsősorban a fals pozitív visszajelzésekre alapozva.

A fenti indikátorok mellett a modell teljesítményét folyamatosan, véletlenszerűen kiválasztott egyre bővülő zsarolóvírusminta halmazon lehetne tesztelni a „concept drift” nyomon követése érdekében. Az alkalmazott gépi tanulási modellek korábbi verziójának újratanítását követően a modellt ismét perzisztálhatjuk (esetünkben „keras”, vagy „h5” fájlformátumban), ami lehetővé teszi a későbbi visszatöltést, a további tanítást, illetve detekció céljából történő felhasználást. Fontos kiemelni azonban, hogy ahogy a mesterséges intelligenciát egyre inkább használják a kibertámadások területén úgy egyre kifinomultabb fertőző, adatexfiltrációs és titkosítási stratégiák jelennek meg. Ezek az eljárások például az időbeli késleltetéseket, vagy a rejtőzködési stratégiákat a támadás során akár környezetre szabottan, dinamikusan változtathatják, ami a védelmi oldalon még komplexebb, szintén mesterséges intelligenciával támogatott képességek fejlesztését teszi szükségessé.

2. A vizsgálat centralizált, virtualizált fájlkiszolgáló-infrastruktúrára vonatkozik. A szerző megítélése szerint mely részei vihetők át változtatás nélkül, illetve melyeket kellene alapvetően újragondolni felhőalapú vagy hibrid adattárolási környezetben?

A felépített virtuális infrastruktúra ARDS fájlkiszolgálójának telepítése és a szolgáltatás-környezet felépítése során fontos szempont volt, hogy a későbbiekben a kiszolgáló és a hozzá kapcsolódó védelmi szolgáltatások könnyen migrálhatóak és lehetőség szerint környezetfüggetlenek legyenek. Ennek eredményeként a kutatás során használt Debian 12.4.0 Bookworm, a Samba 4.19.3 fájlkiszolgáló, a Docker-alapú naplózási szolgáltatások (Fluent Bit, InfluxDB), valamint az XGBoost modell Python-környezete egyaránt könnyen mozgatható felhőbe. A Debian-alapú rendszerekről az S3-kompatibilis objektumtárolók szabványos klienseszközökön és API-kon keresztül is elérhetőek. A kiszolgálói környezet felépítéséből adódóan a rendszer mind vertikálisan, mind horizontálisan könnyen skálázható. Hibrid adattárolási környezetben a hálózati sávszélesség és a késleltetés kritikus tényezők, mivel eltérő lokáció esetén a különböző szolgáltatások (adattárolás, naplózás, detekció) integrációja a hálózati kapcsolat minőségétől és rendelkezésre állásától nagyban függ. Ilyenkor a hálózati kapcsolat kiesése megfelelő redundancia hiányában egyetlen meghibásodási pontként (single point of failure) veszélyeztetheti a rendszer működését. Az alapvető újragondolás egyik szolgáltatáskomponens esetében sem szükséges, azonban a különböző tárolási technológiák eltérő interfészek és protokollok támogatását igényelhetik (pl. S3 API, NVMe over Fabrics, RBD, iSCSI). Ez a gyakorlatban azt jelenti, hogy az

operációs rendszernek képesnek kell lennie a centralizált adatvagyon elhelyezését biztosító tárolórendszer kezelésére.

3. A relatív teljesítménymutató alkalmazása szükségszerűen referencia-rendszer választását igényli. Hogyan befolyásolja a referencia megválasztása az eredmények értelmezését, és milyen szempontok alapján lehet egy általánosabb rendszer-összehasonlító vizsgálatban megfelelő referencia-konfigurációt kiválasztani?

A referencia-rendszer megválasztása alapvetően meghatározza az eredmények értelmezését és összehasonlíthatóságát, mivel a relatív teljesítménymutató az eredményeket ehhez a rendszerhez viszonyítja. Éppen ezért célszerű olyan referencia-rendszert választani, amely teljesítmény szempontjából reprezentatív és releváns. Nem szabad például szándékosan gyengébb teljesítményű rendszert alkalmazni a vizsgált környezet kedvezőbb megítélése érdekében. Általános rendszer-összehasonlító vizsgálatoknál elsődleges referencia lehet a jelenlegi produktív környezet, mivel ez biztosítja a vizsgálat szempontjából leginkább releváns viszonyítási alapot. Amennyiben ilyen rendszer nem áll rendelkezésre, célszerű lehet széles körben elterjedt, iparágilag elfogadott, az adott szolgáltatást tipikusan megvalósító megoldásokat választani. A fájlkiszolgálók esetében ilyen referencia lehet a SMB/CIFS protokollok különböző verzióit támogató megoldások közül például a Synology DSM, a Microsoft Windows Server vagy a TrueNAS. Átfogóbb vizsgálat igénye esetén több referencia-rendszer alkalmazása is indokolt lehet, ami lehetővé teszi, hogy a vizsgált rendszer teljesítményét több, eltérő jellemzőkkel rendelkező környezethez viszonyítsuk és ezáltal árnyaltabb képet kapjunk a relatív teljesítményről.

Végül még egyszer köszönöm Docens Úr támogató véleményezését és téziseim elfogadását. Kérem tisztelt Docens Urat, hogy a bírálóiban felvetett kérdésekre, javaslatokra adott válaszait, magyarázataimat szíveskedjen elfogadni.

Veszprém, 2026. szeptember 28.



Arányi Gábor

doktorjelölt,

okl. mérnökinformatikus