SmartComLab

MŰEGYETEM 1782

Az MI-alapú DDoS-elhárítás kihívásai és korlátai

Dr. Orosz Péter
BME TMIT SmartCom Lab

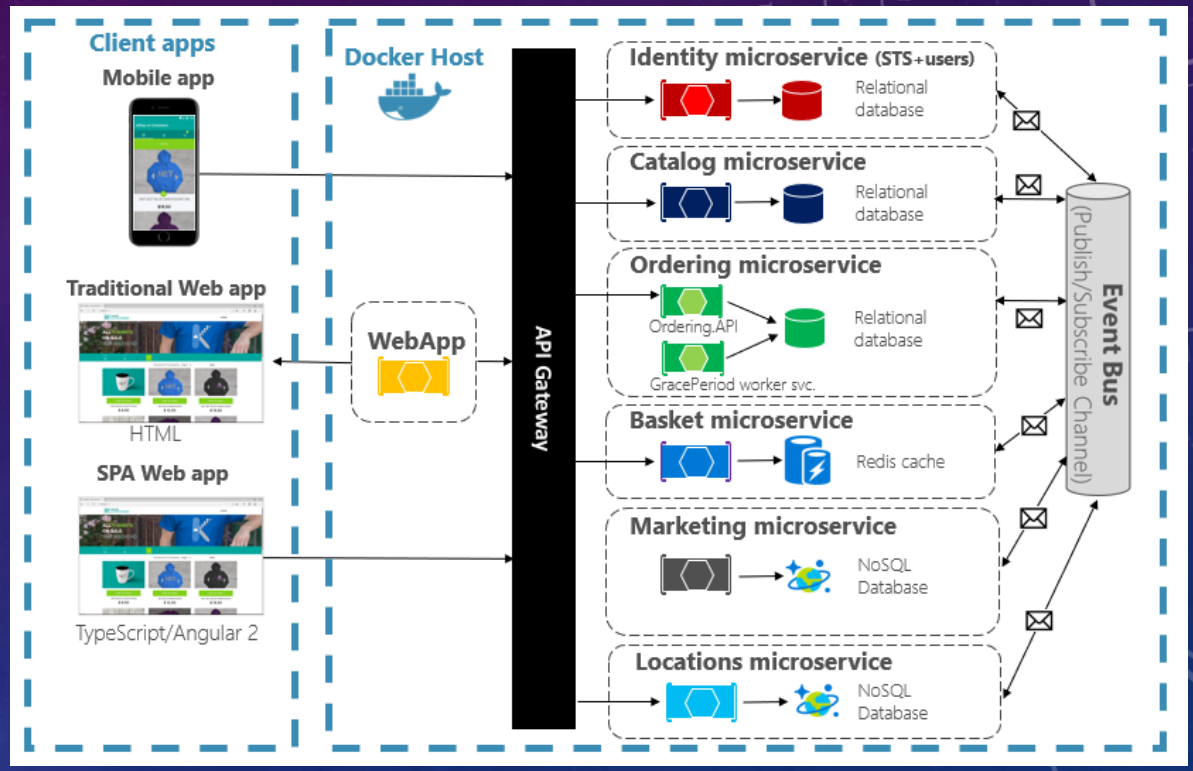
HUNOG 2024

Áttekintés

- Korszerű adatközponti hálózatok forgalmi jellegzetességei
- DDoS trendek, aktualitások és kihívások
- Detekciós és elhárítási módszerek
- A mesterséges intelligencia szerepe a védelmi rendszerekben
- Az MI-támogatott védelmi megoldások lehetőségei és korlátai
- A közeljövő kihívásai

Újfajta forgalmi mintázatok – újfajta kihívások

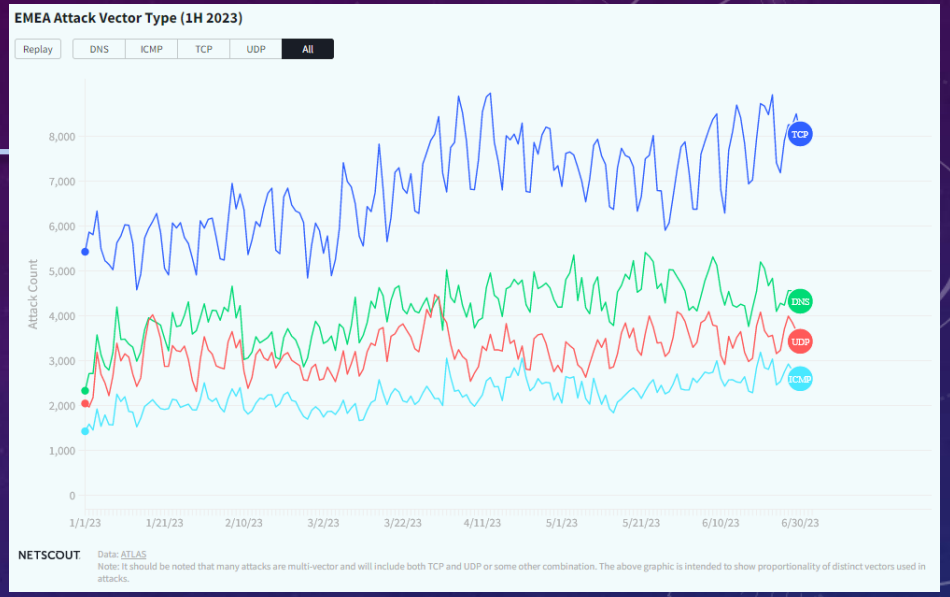
- Cloud-native alkalmazások terjedése: virtualizáció/kontérnerizáció
- Microservice-alapú architektúrák
- Jelentőssé vált a kelet-nyugat forgalom: alacsony K-Ny késleltetés, fix sebesség a klaszterek között
- Az operátorok számára komoly kihívás az alert noise és alert fatigue jelenségek kezelése
- A root-cause elemzés egyre összetettebb, egyre időigényesebb feladat
- Incidensek időegységre vetített növekvő száma
- Teljesítmény- és biztonsági incidensek hatékony elkülönítése

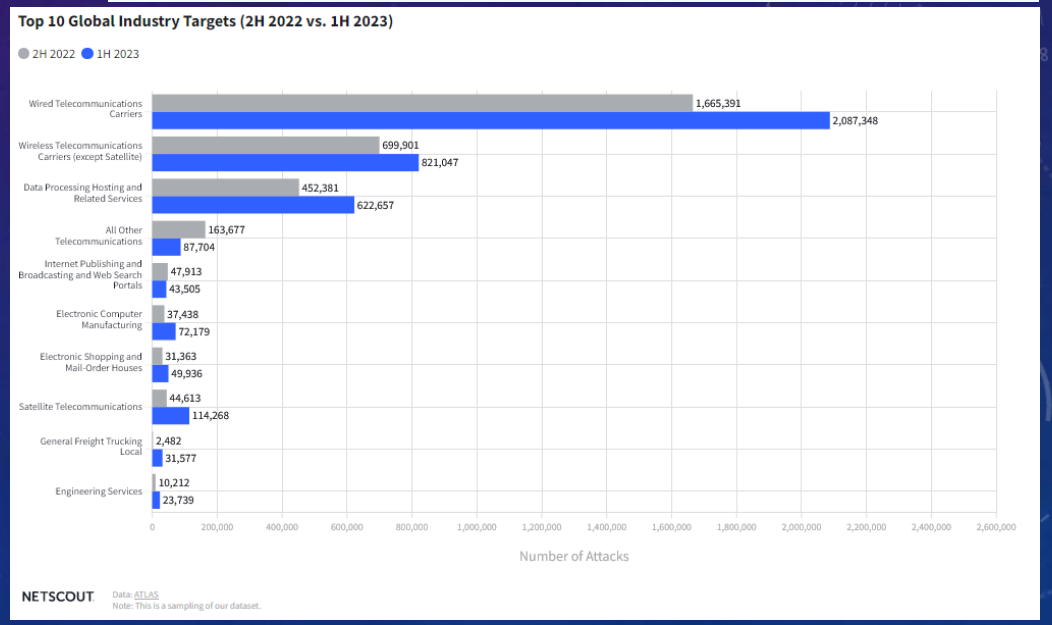


Forrás: microsoft.com

DDoS trendek: 2020-2024

- **Pandémia és háborús konfliktus:** a támadások számának jelentős növekedése (a szomszédos országokban is),
- Radikális változás a módszerekben és a végrehajtás időtartamában
- Volumetrikus, hit-and-run és multivektor támadások
- L3-L4 dominancia, TLS titkosított támadás, elosztott cél, 5G támadó oldalon
- Olcsón bérelhető botnetek
- Economic DDoS
- IoT- és mobil- és szerver-alapú botnetek
- Gyakoribbak a radar alatti esetek
- Felhasználói viselkedés utánzása böngészőben (HTTP támadások)

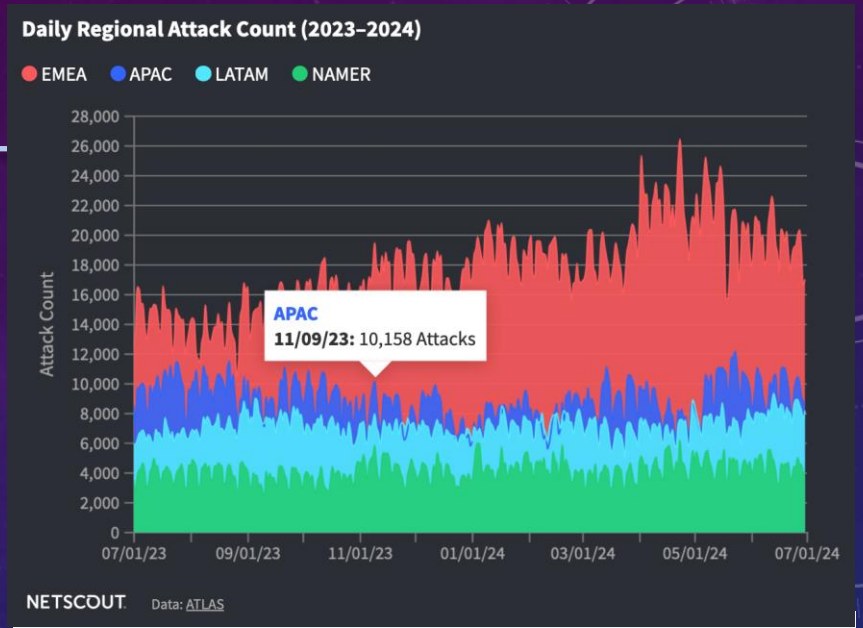


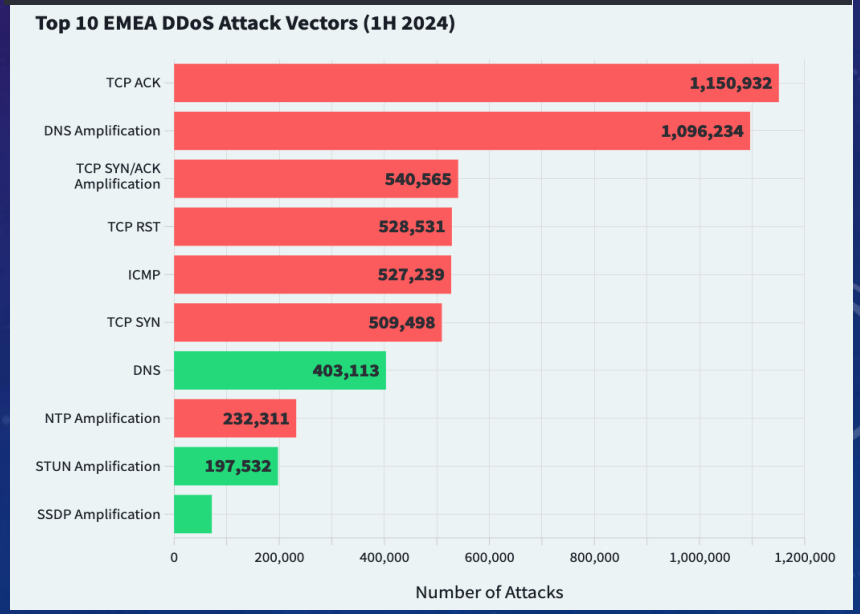


SmartComLab

DDoS támadások: EMEA 2024H1

- Legnagyobb támadás (H1): 352 Mpps, multivektor, Németország, távközlési szolgáltató
- Támadások átlagos hosszának változása: ~50%
- A támadások 75%-a kevesebb mint 15 percig tartott
- Multivektor támadások: ismét emelkedik a >10 vektorszámú esetek száma
- EMEA régióban a DDoS esetek számának növekedése: 25% 2023H2-höz képest
- Domináns vektorok: TCP ACK/RST/SYN, DNS amplification, ICMP, UDP
- Célországok: Németország, Franciaország, Lengyelország, Szaúd-Arábia
- Kiemelt célszegmens: gaming, távközlési operátorok, felhőszolgáltatók

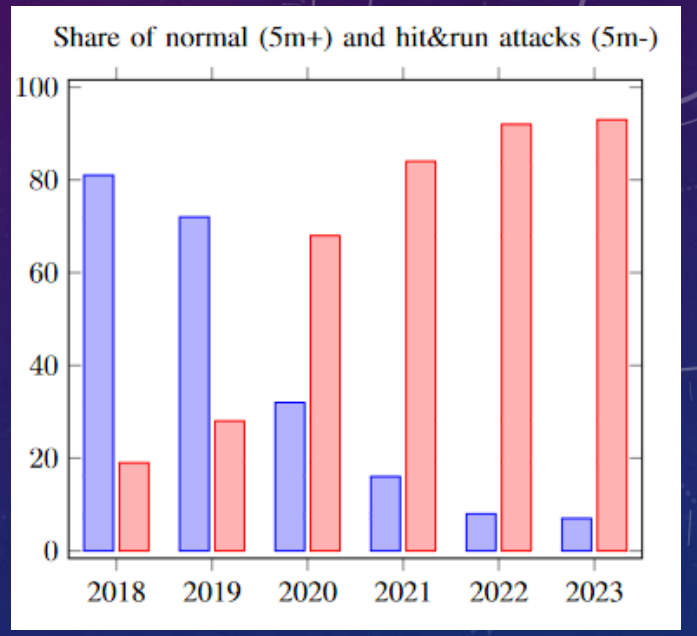




Forrás: NetScout TIR 2024H1

DDoS kihívások

- Komplexebb és egyben rövidebb támadások felismerése és elhárítása
- Csökkenő erőforrásigény támadó oldalon: a rövidebb támadáshoz rövidebb időre kell bérelni a botnetet
- Növekvő erőforrásigény az áldozat oldalán
- Linux-alapú támadás-források: Windows WLS elterjedése, keresztfordítási lehetőség a két operációs rendszer között
- Számos IoT eszköz Linux alapokra épül, de jellemzően alacsony a védelmi szintjük
- Mobil- és böngésző-alapú botnetek terjedése



DDoS kihívások (folyt.)

- Támadó oldalon is megjelent a generatív AI
- Az applikáció-specifikus támadások e2e titkosítottak (nem lehet kiszűrni az aggregációs pontokon)
- Létező megoldás: a védelmi rendszer reverse proxiként beékelődik a kliens és a szerver közzé:
  - DNS bejegyzés módosítása
  - Titkos kulcsok megszerzése
  - Visszafejtett forgalom szűrése
  - Tiszta forgalom titkosítása
  - Továbbítás az eredeti szerver felé

DDoS detekciós módszerek

Forgalmi adatok kinyerése

- In-band detekció: a hálózati forgalom csomagalapú elemzése
- Out-of-band detekció: routerek (pl. Netflow, Sflow, IPFix) flow rekordjainak elemzése

A detekció hatékonyságának objektív vizsgálata

- Detekciós idő
- Fals pozitív ráta
- Fals negatív ráta

Hit and run hatások

- Felhőszolgáltatások jelentős QoS degradációja
  - switch-ek csomagpufferei 1-5 ms alatt telíthetőek
  - a TCP torlódásvezérlés megzavarása: az eredeti átviteli ráta visszaállítása másodperces nagyságrend alatt megy végbe
  - érintett mind az É-D, mind pedig a K-Ny forgalom
- Radar alatti jelenlét: a legtöbb DDoS detekciós rendszer a másodperces nagyságrendben képes azonosítani
- Elhárítási láncban emberi döntés: jelentős késleltetés, figyelem fáradása a nagy számú biztonsági esemény miatt (akár 1000-2000 esemény naponta)

Hit and run új kutatási eredmények

Jelenleg az egyik legkutatottabb támadástípus

Új eredmények:

- Alacsony fals rátájú, alacsony késleltetésű módszerek
- TCP teljesítmény stabilizálása támadás alatt
- Frekvenciatartomány elemzése az időtartomány helyett
- Gépi tanulás a frekvenciatartományi mintázatok azonosítására
- Entrópia detekciós algoritmusok

MI alkalmazási lehetőségek

- 0-day detekció
- Másodlagos csatornák: ticketing és log elemzés (eddig nem lehetett emberi erő nélkül feldolgozni)
- Human-in-the-loop helyett human-on-the-loop működés
- Újfajta analitikai képességek, validálás, finomhangolás
- Szabályok és konfigurációk generálása
- Elemi biztonsági eseményekből incidensek automatikus létrehozása
- Tudásbázis automatikus felépítése

Megjegyzés:

Nincs értelme kicsatolni MI modellre egyszerű támadási mintákat, pl. TCP Syn flood

Kiemelt funkciók

- Enrichment (események, riasztások)
- Anomália-detekció
- Korrelációs vizsgálatok
- Esemény-szintű zajcsökkentés
- Automatizált diagnózis és incidenskezelés
- Tudásbázis építés
- Teljesítmény- és biztonságspecifikus események nagy megbízhatóságú elkülönítése

MI alkalmazási korlátok

- Csomagszintű válogatás MI-vel nem lehetséges a számítási igény miatt (minden csomagról el kell dönteni, hogy a támadáshoz tartozik-e)
- Tanítási adathalmaz teljes cimkézése
- Az automatizált döntés megmagyarázhatósága
- Megmagyarázhatóság vs. komplexitás -> fordított arányosság
- Valósidejűség vs. erőforrásigény
- Fals ráta mértéke

Megmagyarázhatóság

Többféleképpen érthető

Az AI pipeline képes arra, hogy rámutasson, hogy mi és hol történt. Vagyis mely ponton következett be a hibás döntés, de arra már nem képes, hogy az okot feltárja.

Viszont...

A fő kérdés továbbra is az, hogy hogyan lehet legközelebb elkerülni a fals detekciót?

A legfontosabb jelenlegi korlát

Az AI túlhasználata – extrém sok energia

Aki túl sok mindenre használ AI-t, az ugyanúgy versenyhátrányba kerülhet, mint aki egyáltalán nem használja.

Megoldás:

Fejlett előfeldolgozás klasszikus algoritmusokkal

- Kevesebb energia
- Nagyobb pontosság a testre szabott tulajdonságok miatt
- Nyers csomagok helyett n-tuple átadása az AI modellnek, ami tovább gazdagítható security exchange információkkal (IP listázás, geo infó, trust score, stb.)

Az AI modell túl- és alultanítása

A hálózati események osztályozásakor jellemző a kiegyensúlyozatlan adathalmazok előállítása, ami magas fals rátához vezethet.

Azoknak az adatközpontoknak, amelyek jellemzően KKV-k számára hosztolnak, nehezebb a dolguk, mivel a felhasználói forgalmi mixük jóval összetettebb, mint pl. egy Google vagy Facebook szolgáltatási adatközponté.

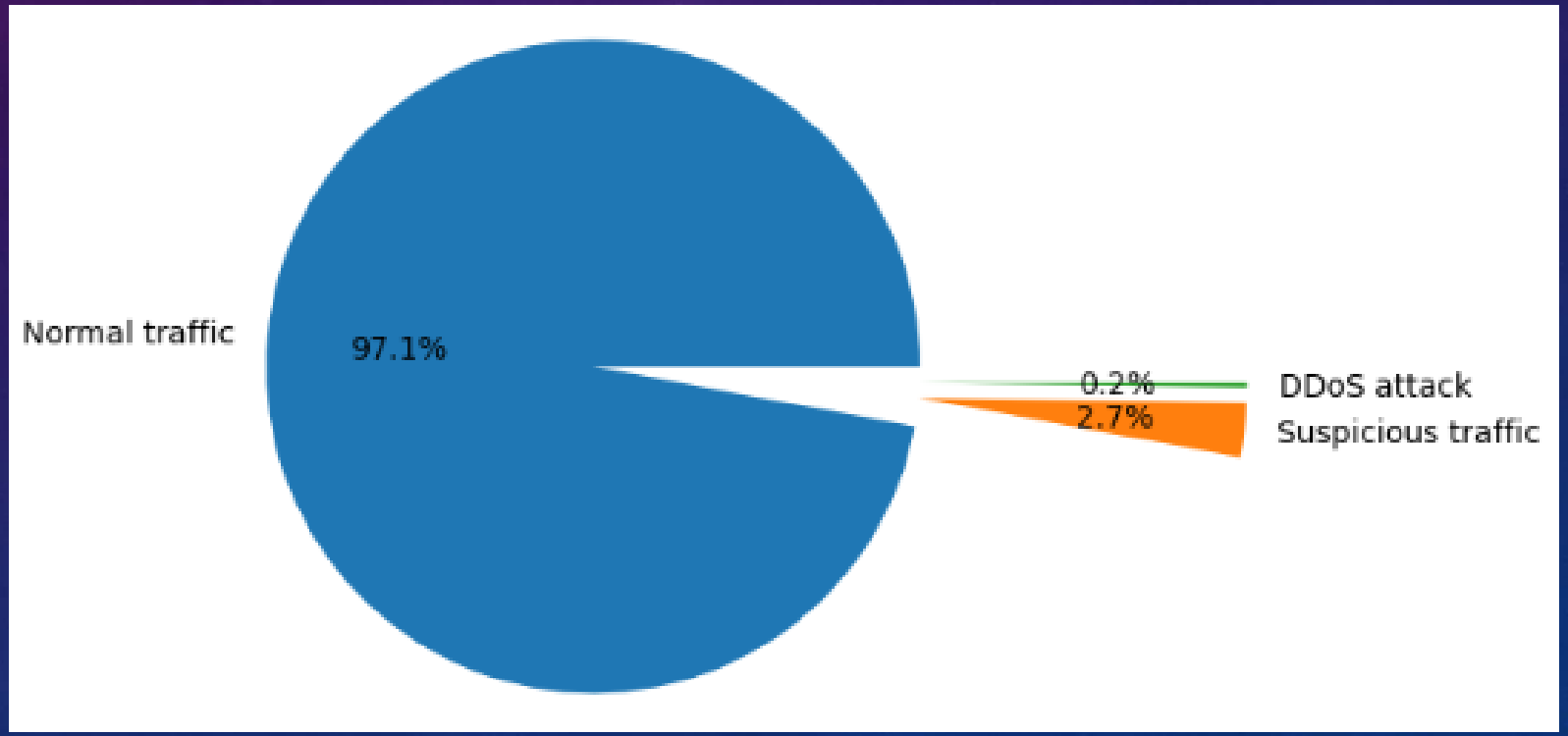
Elosztott védelem: a Cloudflare saját architektúrája

- Paraméterezhető forgalomszűrő: eBPF szűrő (n-tuple, akár még payload is, payload hash).
- AI modul paraméterezi a szűrőt
- Az AI modul mintavételezett forgalmat kap (1:10 vagy 1:100) a detekcióhoz. Az eBPF felküldi az AI-modulnak mind a felhasználók mind a támadás forgalmát. Ez alapján programozza be az AI a szűrőket.
- Az eBPF szűrők a végpontokon működnek (CDN szervereken)

Esettanulmány: a tanító adathalmaz összetétele és a modell hibaaránya

Random-forest modell
Fák száma: 60
Tanítási minták száma: 50.000
Adatok felosztása: 80%-20%

Valós \ Becsült	DDoS támadás	Normál forgalom	Gyanús forgalom
DDoS támadás	102	3	11
Normál forgalom	0	49276	0
Gyanús forgalom	16	0	1360



Összefoglalás

- Egyre összetettebb infrastruktúrák és szolgáltatások
  - Cloud-native és microservice-alapú alkalmazások
  - Növekvő terhek az üzemeltetőkön, adminisztrátorokon
- Human-in-the-loop helyett human-on-the-loop működés
- Újszerű analitikai képességek: enrichment, validáció, kereszt-korreláció, anomália-detekció
- Események és incidensek láthatóságának növelése
- Viselkedésalapú vizsgálatok
- Tudásbázis építése gépi és mélytanulás segítségével
- Automatizált elhárítási folyamatok
- Korlátok: AI túlhasználat – energiafelhasználás – megmagyarázhatóság – adathalmaz

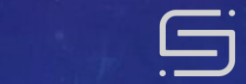
A részletekről érdeklődőknek

Péter Orosz, Balázs Nagy, Pál Varga:

„*Detection strategies for post-pandemic DDoS profiles*”, Infocommunications Journal Q4 2023

Köszönöm a figyelmet!

smartcomlab.tmit.bme.hu



SmartComLab