cesnet 25

Nové technologie pro oblast bezpečnosti

Tomáš Čejka a Jan Kořenek

Nové technologie

Analýza šifrovaného síťového provozu

- Využití technik strojového učení

Zpracování paketů pro rychlosti 400 Gb/s

- Navazuje na zkušenosti z oblasti HW akcelerace

Analýza šifrovaného provozu

“Rozšířená IP Flow Data”

Co to je?

- “Základní” IP Flow záznam:
  - informace “L1-L4”
- “Rozšířený” IP Flow záznam:
  - informace z vyšších vrstev, až L7
  - hostname,SNI,doména,user-agent,to/from,SIP URI,...
  - NEšifrovaný provoz

S COMBO kartou (HW akcelerace) dokážeme monitorovat rychlé linky!

“Rozšířená IP Flow Data”

K čemu to je?

- Forenzní analýza, větší detail než u “základních” flow dat, ale řádově méně dat než při použití celých paketů
- Pokročilejší detekce bezpečnostních hrozeb
  - Komunikace s C&C servery
  - Hádání hesel hrubou silou, iterace uživatelských účtů... a úspěšné prolomení/kompromitace
  - HeartBleed, ShellShock, ...

Šifrovaný provoz!

- Šifrování je důležité z mnoha důvodů:
  - důvěrnost, integrita, nepopiratelnost
- ale…
  - šifrovat umí i útočníci
  - nebo malware (CC servery, šíření kódu, …)
  - do provozu není vidět, rozšířená IP flow data jsou k ničemu
- nebo ne?

Šifrovaný (/nešifrovaný?) provoz

- TLS <= 1.2 - TLS Fingerprinting
  - “handshake”: podporované algoritmy, certifikát, SNI
  - JA3 fingerprinting
  - Cisco Mercury
- TLS 1.3 šifruje více
- Encrypted DNS

Šifrovaný provoz

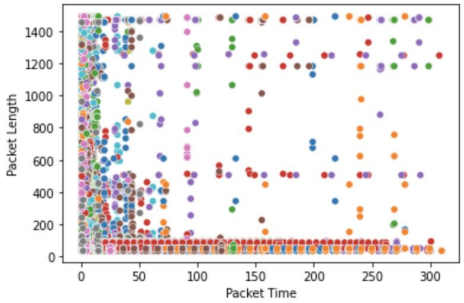
- TLS <= 1.2 - TLS Fingerprinting
  - “handshake”: podporované algoritmy, certifikát, SNI
  - JA3 fingerprinting
  - Cisco Mercury
- TLS 1.3 šifruje více
- Encrypted DNS

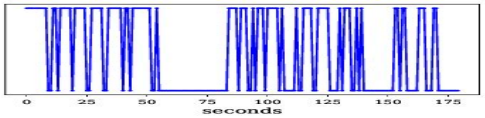
```
20]: # Plot lengths vs. packet time

for i in range(0, len(stats_music)):
    for j in range(0, len(stats_music[i])):
        sns.scatterplot(x=stats_music[i]['PACKET_TIME'][j], y=stats_music[i]['LENGTHS'][j])

plt.ylabel('Packet Length')
plt.xlabel('Packet Time')

plt.show()
```

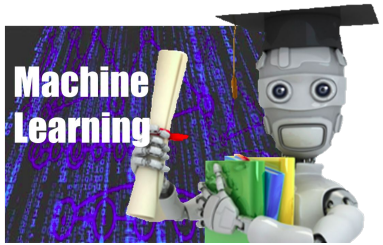




“Postranní kanály”

Strojové učení

- Technologie, která se umí “naučit” funkci pomocí “příkladů”...
- Zdá se, že funguje pro analýzu síťového provozu!
- V posledních letech je to pro nás prioritní oblast:
  - Reputace síťových entit
  - Identifikace typu provozu (podle chování komunikace)
  - Přesnější detekce útoků hrubou silou (v šifrovaném provozu)



Stav poznání

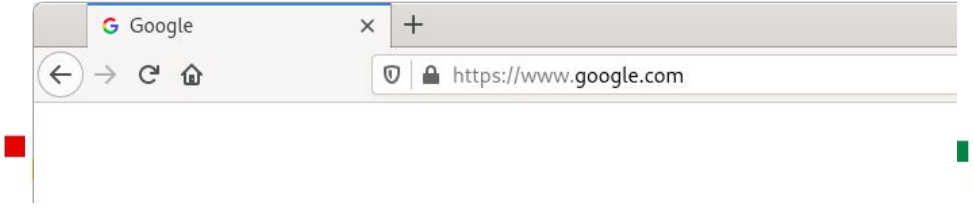
- Existující práce
  - Existují články o analýze provozu pomocí různých metod
  - Rozpoznávání videa podle datového toku
  - Rozpoznávání obsahu VPN
  - Anderson & McGrew (Cisco Systems)
- Náš výzkum
  - cca 6 publikací za poslední rok + 2 podané, 5 plánovaných

Úspěšnost přes 90%!

=> DoH, HTTPS, SSH

Příklady?

- SSH
- TeamViewer
- Tor
- TLS
- WireGuard
- OpenVPN
- QUIC
- HTTPS
- DNS over HTTPS



Budoucnost?

- Analýza šifrovaného provozu bez narušení soukromí
- Klasifikace typů provozu (některé protokoly jsou víceúčelové)
- Detekce škodlivého provozu nad šifrovaným provozem

ETA Publikace za poslední rok

- Hynek, K.; Beneš, T.; Čejka, T.; Kubátová, H. **Refined detection of SSH brute-force attackers using machine learning** In: ICT Systems Security and Privacy Protection. Cham: Springer, 2020. p. 49-63. IFIP Advances in Information and Communication Technology. vol. 580. ISSN 1868-4238. ISBN 978-3-030-58200-5.
- Vekshin, D.; Hynek, K.; Čejka, T. **DoH Insight: Detecting DNS over HTTPS by Machine Learning** In: ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and Security. New York: ACM, 2020. p. 1-8. ISBN 978-1-4503-8833-7.
- Hynek, K.; Čejka, T. **Privacy Illusion: Beware of Unpadded DoH** In: 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Montreal: IEEE, 2020. p. 621-628. ISSN 2644-3163. ISBN 978-1-7281-8416-6.
- Luxemburk, J.; Hynek, K.; Čejka, T. **Detection of HTTPS Brute-Force Attacks with Packet-Level Feature** Set In: 11th Annual Computing and Communication Workshop and Conference (CCWC2021). Piscataway (New Jersey): IEEE, 2021. p. 0115-0123. ISBN 978-0-7381-4394-1.
- Hynek, K.; Vekshin, D.; Luxemburk, J.; Čejka, T.; Wasicek, A.; **Emerging Threat Models for DoH Abuse on the Internet** Submitted: Computer & Security Journal
- Husák, M., Bartoš, V., Sokol, P., Gajdoš, A.: Predictive methods in cyber defense: Current experience and research challenges. Future Generation Computer Systems, vol. 115, Elsevier, February 2021. ISSN 0167-739X.

Více na https://www.liberouter.org/publications/ , v případě problémů s přístupem k článkům se ozvěte na cejkat@cesnet.cz.

=> DoH, HTTPS, SSH

cesnet

400 Gb/s

Zpracování provozu na rychlosti 400 Gb/s

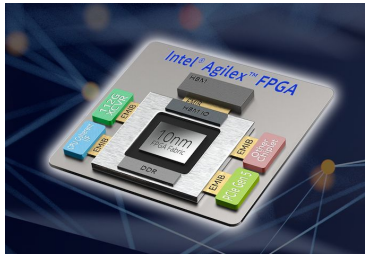
Vycházíme ze zkušeností se 100 GE kartami a HW akcelerací

Několik nových problémů k řešení

- Nutnost zpracovat 4 pakety najednou (paralelně) - nový systém zpracování
- Pro packet capture řešení je nutná podpora PCIe gen 4 nebo gen 5

Spolupráce se společnostmi Intel a Reflex CES

- Návrh nové akcelerační karty s čipy Intel Agilex
- Předpokládáme první vzorky 400GE karet do konce roku 2021



Využití systému DPDK

Systém DPDK poskytuje vynikající výkonnost i při využití pouze komoditních síťových karet (Intel, Mellanox, …)

Škálovatelné řešení čištění DDoS útoků v distribuovaném prostředí

- DPDK zajišťuje rychlý příjem paketů a rychlé paralelní zpracování
- Složitá SW analýza dat vyžaduje běh v distribuovaném prostředí
- Díky DPDK je možné použít filtrační mechanismy i na komoditních kartách

Vhodný standard pro akcelerační technologie a vysokorychlostní zpracování dat

Příprava sond pro 400GE

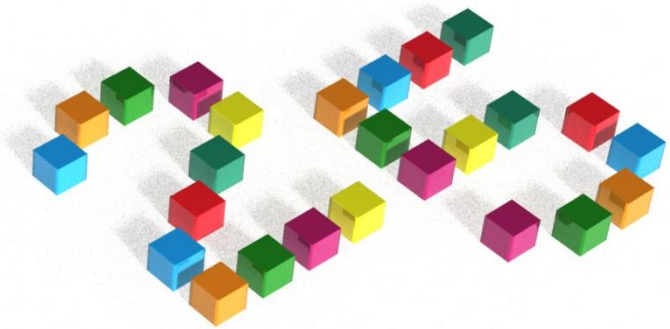
Připravujeme prototyp sondy pro sběr informací o síťových tocích pracující na plné rychlosti 400GE

Podpora IDS systémů přímo na sondě (Suricata)

- Flow tabulka pro shunting
- Pre-filtrace v hardware na základě pravidel

Analýza šifrovaného provozu pomocí ML klasifikátorů

- Sběr velkého množství parametrů (stovky)
- Podpora velké množiny klasifikátorů provozu (stovky až tisíce)



Děkuji za pozornost!

Klasifikátory využívající strojové učení

Vynikající výsledky dosahují klasifikátory na bázi rozhodovacích stromů (Random Forest)

- Vysoká přesnost klasifikace i ve srovnání s hlubokými NN
- Snadná interpretovatelnost chování vyvinutých klasifikátorů

Náročná úloha jak pro sběr vlastností, tak i pro vyhodnocení velkého počtu rozhodovacích stromů

- Řádově desítky vlastností a stovky až tisíce rozhodovacích stromů
- Online zpracování vyžaduje značný výpočetní výkon
- Vhodná HW akcelerace pro FPGA SmartNIC