CESNET a FIT ČVUT zveřejnili dosud největší datovou sadu určenou pro detekci hrozeb a predikci síťového provozu

31.7.2025 - Ivana Macnarová | Fakulta informačních technologií ČVUT v Praze

S detekcí anomálií se setkáváme v každodenním životě často a aniž bychom o tom věděli – ať už jde o podezřelou platbu z jiné země nebo neobvyklou částku zaznamenanou bankovním systémem, odchylky ve zdravotních údajích zachycené chytrými hodinkami nebo náhlou změnu chování při online nákupech, která může indikovat zneužití účtu. Ve všech těchto případech se jedná o detekci anomálií. Tedy situací, které se odchylují od běžného chování a mohou být indikátorem rizika. Podobné principy se uplatňují i v oblasti kybernetické bezpečnosti, kde anomálie v síťovém provozu často signalizují hrozby, chyby nebo kritické změny v chování zařízení.

V oblasti správy a zabezpečení sítí hraje detekce anomálií klíčovou roli. Moderní útoky na infrastrukturu, jako jsou distribuované útoky na dostupnost služeb (DDoS), šíření malwaru nebo zneužití kompromitovaných zařízení, se často skrývají v běžném provozu a unikají tradičním pravidlům detekce.

Tým výzkumníků ze sdružení CESNET a FIT ČVUT – Josef Koumar, Karel Hynek, Tomáš Čejka a Pavel Šiška – publikoval v prestižním časopise Nature Scientific Data dosud nejrozsáhlejší veřejně dostupnou datovou sadu svého druhu. Obsahuje více než 800 tisíc časových řad vytvořených agregací reálného, anonymizovaného síťového provozu ze zařízení, sítí a institucí z páteřních linek národní akademické sítě CESNET.

Na rozdíl od běžně používaných uměle vytvořených laboratorních datových sad, které měla vědecká komunita dosud k dispozici, zachycuje tento dataset rozsáhlý a různorodý provoz reálných počítačových sítí. Jde o bezprecedentní počin, který výrazně posouvá možnosti výzkumu v oblasti kybernetické bezpečnosti a správy sítí. Umožňuje vývoj vysoce přesné umělé inteligence pro detekci anomálií a hlavně její komplexní a robustní testování v reálných podmínkách s různorodým provozem. Výrazně tak zvyšuje věrohodnost výsledků detekce, například útoků typu DDoS nebo podezřelého chování infikovaných zařízení.

Význam přínosu posiluje také publikování open-source knihovny CESNET TS-Zoo, která usnadňuje práci s datovou sadou a zároveň umožňuje snadné sdílení metodologie prostřednictvím benchmarků. Kombinace realistického datasetu a open-source nástroje přispívá k vyšší transparentnosti metod a reprodukovatelnosti experimentů – tedy ke kvalitnějším a ověřitelným výsledkům v celém výzkumném ekosystému.

http://fit.cvut.cz/cs/zivot-na-fit/aktualne/zpravy/23252-cesnet-a-fit-cvut-zverejnili-dosud-nejvetsi-datovou-sadu-urcenou-pro-detekci-hrozeb-a-predikci-sitoveho-provozu