Vědci z CESNETu publikovali unikátní datovou sadu v časopise Nature Scientific Data

5.12.2024 - | CESNET

Výzkumníci z Oddělení nástrojů pro administraci a bezpečnost sdružení CESNET vytvořili a publikovali datovou sadu, která je cenným nástrojem pro pochopení dynamických změn v síťovém provozu.

Tato unikátní datová sada představuje zásadní krok v řešení kybernetických hrozeb a svou jedinečností si vysloužila publikaci v prestižním časopise Nature Scientific Data.

Důležitost modelů strojového učení pro detekci bezpečnostních hrozeb na počítačových sítích je dlouhodobě známá jak vědecké, tak odborné komunitě. Výzkumníci sdružení CESNET zkoumají využití metod strojového učení nad síťovým provozem v rámci projektu "Analýza šifrovaného provozu pomocí síťových toků", který byl vybrán mezi podpořenými projekty v rámci výzvy IMPAKT 1 Ministerstva vnitra ČR. Ačkoliv bylo již během projektu vytvořeno několik vysoce inovativních a přesných detektorů strojového učení, jejich masovému nasazení stále brání několik obtížně řešitelných problémů. Jedním z nejčastěji zmiňovaných je problém tzv. datového posunu—jev, kdy model strojového učení byl vyvinut na datech, která zastarala a již neodrážejí aktuální stav.

Datové sady v běžném životě a jejich fungování
Možná jste se někdy setkali se situací, kdy jste se pokoušeli přihlásit do svého telefonu nebo počítače pomocí rozpoznání obličeje (například Apple Face ID nebo Windows Hello), ale zařízení vás jednoduše nepoznávalo. To se stalo, protože systém byl natrénován na vaší historické podobě, která se mohla změnit—například kvůli probdělé noci došlo mírnému otoku obličeje, nebo jste změnili účes, který nyní zasahuje jinak do obličeje. V takovém případě se projevil datový posun; trénovací data (vaše podoba) byla zastaralá a ověření nefungovalo správně.
Biometrické ověření obličeje však efektivně čelí problému datového posunu prostřednictvím pravidelného přetrénovávání. Pokaždé, když zařízení úspěšně ověří váš obličej, aktualizuje vaši podobu, aby vás příště opět rozpoznalo. Tento systém obvykle funguje, protože naše podoba se mění relativně pomalu. Nicméně, pokud dojde k náhlé změně (například pokud se muži oholí), ověření často selhává a je nutné aktivovat záložní metodu—zadání hesla.

Důležitost datových sad pro bezpečnost síťového provozu
Podobný problém se objevuje také v oblasti kybernetické bezpečnosti. Na rozdíl od většiny běžných situací je však datový posun v kybernetické bezpečnosti obvykle náhlý a nepředvídatelný. Kyberzločinci mohou nalézt nové metody útoků, nebo nasazení nových služeb na síti může dramaticky ovlivnit charakter provozu. I drobné aktualizace certifikátů mohou zásadně změnit charakter síťových dat a tím narušit funkčnost strojového učení.

V oblasti kyberbezpečnosti obvykle nemáme záložní metody detekce, které by fungovaly na 100 %, a proto je zásadní tento fenomén zkoumat. Vzhledem k faktické absenci dostupných datových sad vhodných pro tento výzkum měli vědci dosud omezené možnosti—naštěstí právě vznikla nová datová sada, která tento výzkum umožňuje.

Rok síťového provozu v přelomové datové sadě
Tým vědců ze sdružení CESNET a Fakulty informačních technologií ČVUT v Praze ve složení Karel

Hynek, Jan Luxemburk, Jaroslav Pešek, Tomáš Čejka a Pavel Šiška vytvořil a publikoval unikátní datovou sadu v prestižním časopise Nature Scientific Data, která zahrnuje celý rok anonymizovaného síťového provozu z páteřních linek národní akademické sítě. Vědecká komunita dosud disponovala datovými sadami zachycující pár dnů, či týdnu, kvůli náročnosti dlouhodobého sběru a objemu celkových dat. Vytvoření datové sady obsahující celý rok provozu nemá obdoby, proto je zásadním krokem v řešení výzev, jakým je například datový posun a jeho negativní vliv na bezpečnost síťového provozu.

Nově vytvořená datová sada nejen umožňuje zkoumat postupně se snižující přesnost existujících algoritmů, ale také podporuje vývoj nových metod, které budou schopny adaptivně reagovat na neustále se měnící podmínky v síťovém provozu. Výzkumníkům a odborníkům zabývajícím se síťovou bezpečností poskytuje cenné nástroje pro analýzu chování strojového učení v dynamickém a rychle se měnícím prostředí kybernetických hrozeb. Vzhledem k rychlému vývoji technologií a metod útoků je klíčové, aby vědecká a odborná komunita pokračovala ve výzkumu a implementaci efektivních řešení, která zajistí ochranu před kybernetickými hrozbami a zlepšila celkovou bezpečnost digitálního prostředí.

Sdružení CESNET se v tomto kontextu profiluje jako lídr v oblasti síťové bezpečnosti – nejenže provádí špičkový výzkum, ale také aktivně vytváří podmínky pro jeho realizaci a podporuje další rozvoj v této oblasti. Datová sada publikovaná v prestižním časopise Nature Scientific Data je pak jedním z příkladů kvalitních výsledků, který umožňuje odborné komunitě efektivně reagovat na aktuální i budoucí výzvy v oblasti kybernetické bezpečnosti.

Celé znění článku v anglickém jazyce si můžete přečíst zde.

https://www.cesnet.cz/novinky/vedci-z-cesnetu-publikovali-unikatni-datovou-sadu-v-casopise-nature-scientific-data-253