SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Evidenčné číslo: FEI-5382-111266

VYUŽITIE BPF NA ZABEZPEČENIE OS LINUX

BAKALÁRSKA PRÁCA

2023 **Lukáš Grúlik**

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Evidenčné číslo: FEI-5382-111266

VYUŽITIE BPF NA ZABEZPEČENIE OS LINUX

BAKALÁRSKA PRÁCA

Študijný program:	Aplikovaná informatika
Názov študijného odboru:	Informatika
Školiace pracovisko:	Ústav informatiky a matematiky
Vedúci záverečnej práce:	Ing. Roderik Ploszek
Konzultant:	Ing. Roderik Ploszek

Bratislava 2023 **Lukáš Grúlik**

SÚHRN

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Študijný program:	Aplikovaná informatika
Autor:	Lukáš Grúlik
Bakalárska práca:	Využitie BPF na zabezpečenie OS Linux
Vedúci záverečnej práce:	Ing. Roderik Ploszek
Konzultant:	Ing. Roderik Ploszek
Miesto a rok predloženia práce:	Bratislava 2023

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean et est a dui semper facilisis. Pellentesque placerat elit a nunc. Nullam tortor odio, rutrum quis, egestas ut, posuere sed, felis. Vestibulum placerat feugiat nisl. Suspendisse lacinia, odio non feugiat vestibulum, sem erat blandit metus, ac nonummy magna odio pharetra felis. Vivamus vehicula velit non metus faucibus auctor. Nam sed augue. Donec orci. Cras eget diam et dolor dapibus sollicitudin. In lacinia, tellus vitae laoreet ultrices, lectus ligula dictum dui, eget condimentum velit dui vitae ante. Nulla nonummy augue nec pede. Pellentesque ut nulla. Donec at libero. Pellentesque at nisl ac nisi fermentum viverra. Praesent odio. Phasellus tincidunt diam ut ipsum. Donec eget est. A skúška mäkčeňov a dĺžnov.

Kľúčové slová: Linux, eBPF, bezpečnosť

ABSTRACT

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

Study Programme:	Applied Informatics
Author:	Lukáš Grúlik
Bachelor's thesis:	Use of BPF for Linux OS security
Supervisor:	Ing. Roderik Ploszek
Consultant:	Ing. Roderik Ploszek
Place and year of submission:	Bratislava 2023

On the other hand, we denounce with righteous indignation and dislike men who are so beguiled and demoralized by the charms of pleasure of the moment, so blinded by desire, that they cannot foresee the pain and trouble that are bound to ensue; and equal blame belongs to those who fail in their duty through weakness of will, which is the same as saying through shrinking from toil and pain. These cases are perfectly simple and easy to distinguish. In a free hour, when our power of choice is untrammelled and when nothing prevents our being able to do what we like best, every pleasure is to be welcomed and every pain avoided. But in certain circumstances and owing to the claims of duty or the obligations of business it will frequently occur that pleasures have to be repudiated and annoyances accepted. The wise man therefore always holds in these matters to this principle of selection: he rejects pleasures to secure other greater pleasures, or else he endures pains to avoid worse pains.

Keywords: Linux, eBPF, security

Poďakovanie

I would like to express a gratitude to my thesis supervisor.

Obsah

Úvod

Tu bude krasny uvod s diakritikou atd.
A mozno aj viac riadkovy uvod.

1 Teoretické základy eBPF

V súčasnej dobe sa operačné systémy stávajú čoraz komplexnejšími, s rastúcim počtom aplikácií a služieb, ktoré sa na týchto systémoch spúšťajú, rastie aj potreba zabezpečenia pred rôznymi bezpečnostnými rizikami a hrozbami. Jedným z nástrojov, ktorý sa používa na zlepšenie bezpečnosti v operačných systémoch Linux, je eBPF (extended Berkeley Packet Filter). Je to flexibilný a mocný nástroj, ktorý sa čoraz častejšie používa na rôzne účely v operačných systémoch, ako je napríklad monitorovanie a filtrovanie sieťovej aktivity, optimalizáciu výkonu operačného systému, alebo detegovanie útokov a škodlivého kódu. Taktiež je čoraz častejšie používaný pre implementáciu rôznych nástrojov slúžiacich na analýzu dát a monitorovanie systémov, ako napríklad nástroj *perf*, ktorý umožňuje profilovanie výkonu aplikácií v operačnom systéme. Pre maximálny výkon a bezpečnosť systému je dôležité, aby bol eBPF správne nakonfigurovaný a používaný pretože môže spôsobiť veľké problémy pri zlom zaobchádzaní. [1]

1.1 História eBPF

História eBPF sa začína s technológiou nazývanou *Berkeley Packet Filter* v skratke BPF, ktorá vznikla v národnom laboratóriu Lawrenca Berkeleyho, kde ju 19. Decembra 1992 opísali Steven McCanne spolu s Vanom Jacobsonom. BPF bol navrhnutý ako jednoduchý jazyk pre filtrovanie sieťových paketov, ktorý bol implementovaný ako rozšírenie jadra operačného systému. Jeho hlavným cieľom bolo umožniť používateľom filtrovať sieťové pakety bez nutnosti používať externé nástroje ako napríklad `tcpdump`. BPF bol úspešne implementovaný v operačných systémoch ako *BSD* a Linux a stal sa jedným z kľúčových nástrojov pre sieťovú diagnostiku a analýzu. Avšak, s rastúcimi požiadavkami na funkcionalitu a výkon, bolo potrebné rozšíriť BPF o ďalšie možnosti. Začiatkom 21. storočia sa začal vývoj zameraný na vylepšenie technológie BPF. Nová verzia nazvaná eBPF (*extended Berkeley Packet Filter*) vznikla ako rozšírenie k BPF, ktoré bolo navrhnuté tak, aby poskytlo viac funkcií a umožňovalo vykonávať komplexnejšie filtračné skripty na sieťových paketoch. Tento nástroj bol prebratý Linuxovou komunitou, kde sa stal veľmi populárnym pre rôzne účely a v roku 2014 bol implementovaný do jadra Linuxu. Medzi hlavné rozdiely medzi BPF a eBPF patrí podpora pre x86 a arm architektúry, možnosť spustiť aplikácie v jadre operačného systému a možnosť vykonávať viac operácií ako len filtrovanie sieťových paketov. V súčasnosti eBPF umožňuje používateľom načítať a spustiť vlastné programy vo virtuálnom priestore (*sandboxe*) v rámci jadra operačného systému. To znamená, že môže rozšíriť alebo dokonca upraviť spôsob, akým sa jadro správa bez zmeny zdrojového kódu

jadra. Programy spustené týmto spôsobom sú schopné monitorovať systém, zhromažďovať metriky, a dokonca aj vykonávať rôzne úlohy ako napríklad sledovanie a upravovanie sieťovej aktivity. Vďaka týmto rozšíreniam sa eBPF stal veľmi flexibilným nástrojom pre riešenie rôznych problémov v oblasti sieťovej diagnostiky, monitorovania systému a izolácie kontajnerov. [1, 2, 3]

1.2 Princípy eBPF

1.2.1 Ako funguje eBPF

eBPF programy sú riadené udalosťami a spúšťajú sa, keď jadro alebo aplikácia prejde určitým bodom hookom. Preddefinované hooky zahŕňajú systémové volania, vstup/výstup funkcií, sledovacie body jadra, sieťové udalosti a niekoľko ďalších. Ak pre konkrétnu požiadavku neexistuje preddefinovaný hook, je možné vytvoriť `kernel probe (kprobe)` alebo `user probe` (`uprobe`) na pripojenie eBPF programov takmer kdekoľvek v užívateľských aplikáciách alebo jadre.

1.2.2 Ako sa píšu programy eBPF

V mnohých scenároch sa eBPF nepoužíva priamo, ale nepriamo prostredníctvom projektov ako je napríklad Cilium, bcc alebo bpftrace, ktoré poskytujú abstrakciu nad eBPF a nevyžadujú priame písanie programov, ale namiesto toho ponúkajú možnosť špecifikovať definície založené na zámeroch, ktoré sa potom implementujú pomocou eBPF. Ak neexistuje abstrakcia vyššej úrovne, je potrebné programy písať priamo. Jadro Linuxu očakáva, že programy eBPF budú načítané vo forme bajtkódu (`bytecode`). Aj keď je samozrejme možné napísať bytecode priamo, bežnejšou vývojovou praxou je využitie kompilátora, ako je LLVM, na kompiláciu pseudo-C kódu do eBPF bajtkódu.

1.3 Architektúra eBPF

1.3.1 Kompilácia JIT

Krok kompilácie `Just-in-Time` (`JIT`) prekladá všeobecný bajtový kód programu do inštrukčnej sady špecifickej pre stroj s cieľom optimalizovať rýchlosť vykonávania programu. Vďaka tomu sa programy eBPF spúšťajú rovnako efektívne ako natívne skompilovaný kód jadra alebo ako kód načítaný ako modul jadra.

1.3.2 eBPF Maps

Dôležitým aspektom programov eBPF je schopnosť zdieľať zhromaždené informácie a ukladať stav. Na tento účel môžu programy eBPF využívať koncept máp pre ukladanie a načítavanie údajov v širokom súbore dátových štruktúr. K mapám možno pristupovať z eBPF programov, ako aj z aplikácií v používateľskom priestore prostredníctvom

systémového volania.

1.3.3 eBPF Helpers

eBPF programy nemôžu volať ľubovoľné funkcie jadra. Ak by sa to povolilo, programy eBPF by sa viazali na konkrétne verzie jadra a skomplikovala by sa kompatibilita programov. Namiesto toho môžu programy eBPF uskutočňovať volania funkcií do pomocných funkcií, čo je dobre známe a stabilné API, ktoré jadro ponúka. Súbor dostupných pomocných volaní sa neustále vyvíja. Príklady dostupných pomocných volaní:

- Prístup k mape eBPF
- Generovanie náhodných čísel
- Získať aktuálny čas a dátum
- Získať kontext procesu/skupiny
- Manipulácia so sieťovými paketmi a logika presmerovania

1.4 eBPF nástroje

Programovanie eBPF je neuveriteľne výkonné, ale aj zložité. Z toho dôvodu vzniklo niekoľko projektov a dodávateľov, ktorí stavajú na platforme eBPF s cieľom vytvoriť novú generáciu nástrojov, ktoré budú pokrývať pozorovateľnosť, bezpečnosť, sieťovanie a ďalej.

1.4.1 Knižnica libbpf C/C++

je generická knižnica eBPF založená na jazyku C/C++, ktorá pomáha oddeliť načítavanie objektových súborov eBPF generovaných kompilátorom clang/LLVM do jadra a vo všeobecnosti abstrahuje interakciu so systémovým volaním BPF poskytovaním ľahko použiteľných API knižníc pre aplikácie.

1.4.2 BCC

Umožňuje používateľom písať programy v jazyku python s vloženými programami eBPF. Tento framework je primárne zameraný na prípady použitia, ktoré zahŕňajú profilovanie-/sledovanie aplikácií a systémov, kde sa program eBPF používa na zber štatistík alebo generovanie udalostí V používateľskom priestore zbiera údaje a zobrazuje ich v ľudsky čitateľnej forme.

1.4.3 bpftrace

bpftrace je vysokoúrovňový trasovací jazyk pre Linux eBPF, ktorý je k dispozícii v jadrách Linuxu od verzie 4.x. bpftrace používa LLVM ako backend na kompiláciu skriptov do bajtkódu eBPF a využíva BCC na interakciu s linuxovým subsystémom eBPF, ako aj

existujúce možnosti trasovania Linuxu Jazyk bpftrace je inšpirovaný jazykmi awk, C a predchádzajúcimi trasovačmi, ako sú DTrace a SystemTap.

1.4.4 Knižnica eBPF Go

poskytuje všeobecnú knižnicu eBPF, ktorá oddeľuje proces získania bajtkódu eBPF, načítanie a správu programov eBPF. Programy eBPF sa zvyčajne vytvárajú napísaním jazyka vyššej úrovne a potom sa pomocou kompilátora clang/LLVM skompilujú do bajtkódu eBPF.

1.5 Použitie eBPF

1.5.1 Bezpečnosť a eBPF

Počas vývoja eBPF bola bezpečnosť najdôležitejším aspektom pri zvažovaní začlenenia eBPF do jadra Linuxu. eBPF bezpečnosť je zabezpečená prostredníctvom niekoľkých vrstiev:

Požadované oprávnenia

Pokiaľ nie je povolený neprivilegovaný eBPF, všetky procesy, ktoré majú v úmysle načítať programy eBPF do jadra Linuxu, musia byť spustené v privilegovanom režime (root) alebo musia vyžadovať schopnosť `CAP_BPF`. To znamená, že nedôveryhodné programy nemôžu načítať programy eBPF. Ak je zapnutý neprivilegovaný režim eBPF, neprivilegované procesy môžu načítať určité programy eBPF s výhradou obmedzenej sady funkcií a s obmedzeným prístupom k jadru.

Overovač (Verifier)

Ak je procesu povolené načítať program eBPF, všetky programy stále prechádzajú cez overovač eBPF. Overovač eBPF zabezpečuje bezpečnosť samotného programu. To znamená, že napr:

- Programy eBPF môžu obsahovať tzv. ohraničené slučky, ale program je prijatý len vtedy, ak overovateľ môže zabezpečiť, že slučka obsahuje výstupnú podmienku, ktorá sa zaručene stane pravdivou.
- Programy nesmú používať žiadne neinicializované premenné ani pristupovať do pamäte mimo hraníc.
- Programy sa musia zmestiť do požiadaviek na veľkosť systému. Nie je možné načítať ľubovoľne veľké programy eBPF.
- Program musí mať konečnú zložitosť. Overovač vyhodnotí všetky možné cesty vykonávania a musí byť schopný dokončiť analýzu v medziach nakonfigurovanej hornej

hranice zložitosti.

Hardening (Tvrdenie)

Po úspešnom dokončení overovania program eBPF prejde procesom „tvrdenia“ podľa toho, či je program načítaný z privilegovaného alebo neprivilegovaného procesu. Tento krok zahŕňa: Ochranu vykonávania programu: Pamäť jadra, v ktorej sa nachádza program eBPF, je chránená a je určená len na čítanie. Pokiaľ sa program pokúsi niečo modifikovať, jadro sa zrúti aby neumožnilo pokračovať vo vykonávaní poškodeného/manipulovaného programu. Zmiernenie proti Spectre: Pri špekulácii môžu procesory nesprávne predpovedať vetvy a zanechať pozorovateľné vedľajšie efekty, ktoré by sa mohli extrahovať prostredníctvom bočného kanála. Konštantné zaslepenie: Všetky konštanty v kóde sú zaslepené, aby sa zabránilo útokom JIT spraying.

1.5.2 Seccomp

Mechanizmus seccomp() umožňuje procesu načítať BPF program na obmedzenie jeho budúceho používania systémových volaní. Jedná sa o jednoduchý, ale flexibilný mechanizmus sandboxingu, ktorý sa široko používa. Tieto filtračné programy však bežia na "klasickom"virtuálnom stroji BPF, a nie na rozšírenom stroji eBPF, ktorý sa používa na iných miestach jadra. Účelom programu BPF pod funkciou seccomp() je rozhodovať o tom, či má byť dané systémové volanie povolené. Prechodom na eBPF by sa seccomp() programom sprístupnilo množstvo nových funkcií vrátane máp, pomocných funkcií, ukladania na jednotlivé úlohy, expresívnejšej inštrukčnej sady a ďalších. Programy pre eBPF možno písať v jazyku C, čo nie je možné pre programy klasického BPF. Tento problém, viedol k vytvoreniu špeciálnych jazykov, ako je easyseccomp. Kvôli bezpečnostným problémom nie je zatiaľ možné integrovať eBPF do systému seccomp()

* Jedným z prvých použití virtuálneho stroja BPF mimo siete bola implementácia politík kontroly prístupu pre systémové volanie seccomp(). [2]

1.5.3 LSM (Linux Security Modules)

Framework bezpečnostného modulu Linuxu (LSM) poskytuje mechanizmus na pripojenie rôznych bezpečnostných kontrol pomocou nových rozšírení jadra. Primárnymi používateľmi rozhrania LSM sú rozšírenia MAC (Mandatory Access Control), ktoré poskytujú komplexnú bezpečnostnú politiku. Okrem väčších rozšírení MAC možno pomocou rozhrania LSM vytvárať aj ďalšie rozšírenia, ktoré poskytujú špecifické zmeny fungovania systému, ak tieto úpravy nie sú k dispozícii v základnej funkcii samotného systému Linux. [3] Z pohľadu bezpečnostného správania sa lepšie mapuje na LSM ako na filtre seccomp, ktoré sú založené na zachytávaní syscallov. Rôzne bezpečnostné správanie sa môže realizovať

prostredníctvom viacerých systémových volaní, takže by bolo ľahké jedno alebo viacero z nich prehliadnuť, zatiaľ čo hooky LSM zachytávajú správanie, ktoré je predmetom záujmu. Zámerom je, aby eBPF helpre boli "presné a granulárne". Na rozdiel od API sledovania BPF nebudú mať všeobecný prístup k vnútorným dátovým štruktúram jadra. KRSI vyžaduje na svoju prácu `/CAP_SYS_ADMIN`. * `/CAP_SYS_ADMIN` je potrebný na vykonávanie celého radu administratívnych operácií, ktoré je ťažké z kontajnerov vypustiť, ak sa v kontajneri vykonávajú privilegované operácie.

1.5.4 KRSI

Prototyp KRSI je implementovaný ako bezpečnostný modul Linuxu (LSM), ktorý umožňuje pripojenie programov eBPF k bezpečnostným hákom jadra.

Hlavným cieľom KRSI je sledovať celkové správanie systému za účelom odhalenia útokov. KRSI exportuje novú hierarchiu súborového systému pod `/sys/kernel/security/bpf` s jedným súborom pre každý hák. K danému háku môže byť pripojených viac ako jeden program. Pri každom volaní bezpečnostného háku sa postupne zavolajú všetky pripojené programy BPF, a ak niektorý program BPF vráti chybový stav, požadovaná akcia sa zamietne.

KRSI vs Landlock

KRSI je nástroj pre správcov systému, ktorí sa zaujímajú o monitorovanie správania systému ako celku. Landlock je určený na to, aby umožnil neprivilegovaným používateľom sandboxovať programy, ktoré spúšťajú. Na pripojenie programu BPF k háku prostredníctvom Landlock nie sú potrebné žiadne oprávnenia.

2 Aplikácie eBPF

2.0.1 Detekcia útokov

3 Porovnanie eBPF s inými nástrojmi

4 Implementácia vlastného riešenia pomocou eBPF

5 Testovanie a vyhodnotenie výsledkov implementácie

Záver

Conclusion is going to be where?
Here.

Zoznam použitej literatúry

1. BORKMANN, Daniel a STAROVOITOV, Alexei. *eBPF* [online]. 2022. [cit. 2022-11-24]. Dostupné z : `https://ebpf.io/`.
2. CARTER, Eric. *Introducing Container Observability with eBPF* [online]. 2019. [cit. 2023-01-10]. Dostupné z : `https://sysdig.com/blog/introducing-container-observability-with-ebpf-and-sysdig/`.
3. WIKIPEDIA. *Berkeley Packet Filter* [online]. 2023. [cit. 2023-01-10]. Dostupné z : `https://en.wikipedia.org/wiki/Berkeley_Packet_Filter`.