KCKB UPJS

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH

Financované Európskou úniou NextGenerationEU

PLÁN OBNOVY

MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

3. Základné aspekty asembléru a nízkoúrovňová analýza

Rámec nízkoúrovňovej analýzy

- **Význam asembléru:** Nie archaický jazyk, ale priama reprezentácia strojového kódu a jediný "zdroj pravdy" pri absencii zdrojového kódu.
- **Doména malvéru:** Škodlivý kód, exploity a perzistencia operujú na úrovni inštrukcií, kde sa stierajú hranice medzi dátami a kódom.
- **Ciele:**
  - Pochopiť transformáciu logiky C do asembléru (GCC na Linuxe).
  - Identifikovať kryptografické rutiny a skryté funkcie.
  - Analyzovať techniky obfuskácie, anti-debuggingu a anti-VM.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Architektúra x86 vs. x64

- **Spätná kompatibilita:** x64 je rozšírením x86.
  - *Legacy Mode / Compatibility Mode:* Umožňuje beh 32-bit aplikácií.
  - *Heaven's Gate (primárne známe z prostredia Windows/WoW64):* Technika malvéru prepínajúca medzi 32-bit a 64-bit segmentmi kódu v jednom procese (sťažuje debugging).
- **Adresný priestor:**
  - x86: Limit 4 GB (2^32).
  - x64: Teoreticky 16 EB (2^64), prakticky 48-bitová virtuálna adresa (klasický štandard, 256 TB) alebo 57-bitová (moderné CPU, 128 PB).
- **Kánonický tvar adresy (48-bitová adresa):** Bity 48 až 63 musia kopírovať bit 47, inak nastáva *#GP - General Protection Fault* (dôležité pri analýze pádov exploitov).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Hierarchia registrov (GPR)

- **Rozšírenie:** Pôvodné registre (EAX, EBX...) rozšírené na 64-bit (RAX, RBX...) + 8 nových registrov (R8 – R15).
- **Hierarchický prístup:**
  - 64-bit: `RAX`
  - 32-bit: `EAX` (spodných 32 bitov)
  - 16-bit: `AX`
  - 8-bit: `AH` (horná polovica AX), `AL` (dolná polovica AX)
- **Nové možnosti:** Prístup k spodným 8 bitom aj pri registroch indexov (SIL, DIL) a nových registroch (R8B, R8W, R8D).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Zero-Extension a jeho bezpečnostné implikácie

- **Pravidlo Zero-Extension:**
  - Zápis do **32-bitovej** časti registra v 64-bitovom režime (napr. `MOV EAX, 1`) automaticky **vynuluje** horných 32 bitov celého registra (RAX).
- **Výnimka:**
  - Zápis do **8-bit** alebo **16-bit** časti (napr. `MOV AL, 1`) **neovplyvňuje** horné bity.
- **Význam pre analytika:** Nutnosť správne sledovať tok dát (taint analysis), aby nedošlo k mylnej interpretácii hodnôt v registroch pri optimalizovanom kóde.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Segmentácia v Linuxe: FS a GS

- **Flat Memory Model:** Segmentácia sa bežne nepoužíva, okrem FS a GS.
- **Register FS (User-space):**
  - Implementácia **Thread Local Storage (TLS)**.
  - **Stack Canary:** Ochrana proti pretečeniu zásobníka. Hodnota uložená na `FS:[0x28]`.
  - *Detekcia:* Inštrukcia `MOV RAX, QWORD PTR FS:[0x28]` v prológu funkcie.
- **Register GS (Kernel-space):**
  - Prístup k dátam špecifickým pre CPU jadro (Per-CPU data) pre efektívny SMP (Symmetric Multiprocessing).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Syntax Asembléru: Intel vs. AT&T

- **AT&T (GNU Štandard):**
  - Zdroj -> Cieľ (`movq $5, %rax`).
  - Prefixy `%` (registre) a `$` (hodnoty).
  - Pamäť: `offset(base, index, scale)`.
- **Intel (RE Štandard - IDA, Ghidra):**
  - Cieľ <- Zdroj (`mov rax, 5`).
  - Bez prefixov, algebraický zápis.
  - Pamäť: `[base + index*scale + displacement]`.
  - Explicitná veľkosť operandov: `QWORD PTR`, `DWORD PTR`.

Poznámka: V tejto prednáške a v praxi RE sa preferuje Intel syntax.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Kľúčové inštrukcie a ich zneužitie

- **MOV vs. LEA:**
  - `LEA` (Load Effective Address): Vypočíta adresu, ale **nepristupuje** k pamäti.
  - *Malvér využitie:* Rýchla aritmetika bez zmeny flagov (napr. `LEA EAX,` = `EBX + 8`).
- **XOR Idiom:**
  - `XOR EAX, EAX`: Efektívny spôsob vynulovania registra (zero-extends do RAX).
  - *Malvér využitie:* Základná obfuskácia reťazcov a dát.

Financované Európskou úniou
NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Riadenie toku a rozhodovacia logika

- **Porovnanie:**
  - `CMP`: Aritmetické odčítanie (neukladá výsledok, len mení flagy).
  - `TEST`: Logický AND (často `TEST RAX, RAX` pre kontrolu NULL).
- **Podmienené skoky (Jcc):**
  - Závislosť na RFLAGS (ZF, SF, OF, CF).
  - Rozdiel `JA` (unsigned) vs. `JG` (signed).
- **SYSCALL:**
  - Rozhranie pre systémové volania v 64-bit Linuxe.
  - Analýza tabuľky syscallov odhaľuje interakciu malvéru s OS (súbory, sieť).
  - 64-bitový malvér na Linuxe môže (na obídenie niektorých bezpečnostných monitorov alebo nástrojov ako strace, ak sú zle nakonfigurované) zámerne použiť staré 32-bitové prerušenie int 0x80 namiesto inštrukcie syscall. Kernel túto inštrukciu stále spracuje (cez compatibility layer).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

System V AMD64 ABI

- **Odovzdávanie argumentov (v tomto poradí):**
  - **RDI**
  - **RSI**
  - **RDX**
  - **RCX**
  - **R8**
  - **R9**
- **Zásobník:** Ďalšie argumenty (7+) uložené na stacku.
- **Návratová hodnota:** Register **RAX**.
- **Red Zone:** 128 bajtov pod RSP, ktoré môžu listové funkcie využiť bez posúvania stack pointera (optimalizácia).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Stack Frame a preklad z C

- **Frame Pointer Omission:**
  - Moderné GCC (`-O2`) nepoužíva RBP ako bázu zásobníka.
  - Premenné adresované relatívne k **RSP**, čo komplikuje manuálnu analýzu (offsety sa menia s `PUSH`/`POP`).
- **Padding a zarovnanie:**
  - Štruktúry v C sú zarovnané na násobky veľkosti členov.
  - Medzi `char` (1B) a `int` (4B) vkladá kompilátor 3 bajty paddingu.
  - *Chyba analytika:* Interpretácia paddingu ako skrytej premennej.
  - atribút __attribute__((packed)): vývojári malvéru, alebo C2 sieťových protokolov často tento atribút používajú, aby padding úplne zrušili.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Switch-Case konštrukcie

- **Riedke hodnoty:** Séria `CMP` a `JNE` (If-Else reťazec).
- **Husté hodnoty (Jump Table):**
  - Vytvorenie poľa adries v sekcii `.rodata`.
  - Nepriamy skok (pri explicitne načítanej adrese): `JMP RAX` (kde RAX obsahuje adresu načítanú z tabuľky).
- **Dekompilácia:** Ak nástroj nerozpozná tabuľku, kód jednotlivých vetiev sa javí ako mŕtvy kód (orphaned blocks).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Obfuskácia: Control Flow Flattening

- **Princíp:** Transformácia stromovej štruktúry programu na plochý graf riadený centrálnym dispečerom.
- **Komponenty:**
  - **Stavová premenná:** Určuje, ktorý blok sa vykoná.
  - **Dispečer:** Switch statement v nekonečnom cykle.
  - **“Sploštené” bloky:** Aktualizujú stav a vracajú sa k dispečerovi.
- **Výsledok:** "Špagetový kód", strata vizuálnej následnosti blokov v grafoch (IDA Pro/Ghidra).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Obfuskácia: Opaque Predicates

- **Definícia:** Podmienené skoky, ktorých výsledok je známy v čase kompilácie (vždy True/False), ale pre statický analyzátor neznámy.
- **Cieľ:**
  - Vytvorenie falošných vetiev v CFG.
  - Vloženie "Junk Code" (napr. bajt `0xE8` uprostred inštrukcie) na desynchronizáciu disassemblera.
- **Príklad:** Matematické invarianty (napr. `7y^2 - 1 != x^2`), ktoré vyžadujú SMT solver na vyriešenie.
- Táto technika spoľahlivo oklame Linear Sweep disassemblery (ako je štandardný objdump), ktoré čítajú bajty rad za radom. Avšak moderné Recursive Descent disassemblery (ako IDA Pro, Ghidra), ktoré sledujú tok riadenia programu (sledujú inštrukcie JMP, nečítajú naslepo), túto techniku často automaticky prekonajú, pokiaľ nie je kombinovaná s inou obfuskáciou.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Linux Infekcia: LD_PRELOAD

- **Mechanizmus:** Premenná `LD_PRELOAD` alebo súbor `/etc/ld.so.preload` vynúti načítanie škodlivej knižnice pred `libc`.
- **Hooking:**
  - Malvér exportuje funkcie ako `readdir`, `open`.
  - Volá originálnu funkciu cez `dlsym(RTLD_NEXT, ...)`.
  - Filtruje výstup (skrýva súbory útočníka).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Linux Infekcia: .init_array

- **Sekcie ELF:** `.init_array` a `.fini_array` obsahujú pointery na funkcie vykonávané **pred** a **po** funkcii `main`.
- **Vektor útoku:**
  - Vloženie shellkódu do "code cave".
  - Presmerovanie pointera v `.init_array` na tento shellkód.
- **Nebezpečenstvo:** Analytici sa sústredia na `main`, pričom škodlivý kód sa vykoná už počas inicializácie C runtime.

Financované Európskou úniou
NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Anti-Debugging Techniky

- **RDTSC (Read Time-Stamp Counter):**
  - Meranie počtu cyklov procesora (Timing Attack).
  - Detekcia spomalenia spôsobeného krokovaním v debuggeri (Single-stepping).
- **CPUID:**
  - Vrátenie informácií o procesore. Zavolaním inštrukcie CPUID s hodnotou EAX=1 sa do registra ECX načítajú informácie o vlastnostiach CPU
  - Detekcia príznaku "Hypervisor Present" v registri ECX (detekcia VMware, VirtualBox).

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

Záver

Reverzné inžinierstvo vyžaduje syntézu znalostí:

- Architektúra x64 a registre.
- Konvencie System V ABI.
- Štruktúra ELF súborov.
- Optimalizačné vzory GCC.

Schopnosť odhaliť obfuskáciu a anti-analýzu je kľúčová pre modernú analýzu hrozieb.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

KCKB UPJS

UNIVERZITA
PAVLA JOZEFA ŠAFÁRIKA
V KOŠICIACH

Financované
Európskou úniou
NextGenerationEU

PLÁN OBNOVY

MINISTERSTVO
INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA
A INFORMATIZÁCIE
SLOVENSKEJ REPUBLIKY

Ďakujem za pozornosť.