Zadání

Úroveň č. 1

Naprogramujte konzolovou aplikaci, která bude provádět porovnání obecných binárních souborů na základě jejich podobnosti.

Vstup:

Dva obecné binární soubory, s velikostí max. 1MB

Výstup:

Číslo 0 až 100, přičemž výsledek nula znamená, že v souborech není žádná podobnost; číslo 100 znamená úplnou binární shodnost. Čísla mezi nulou a stovkou pak znamenají menší či větší míru podobnosti. Aplikace nechť zobrazuje změřené číslo na konzoli. Změřené číslo bude také vraceno ve výstupní hodnotě aplikace – tak aby bylo možno hodnotu ověřovat ve skriptech (batch files) v hodnotě ERRORLEVEL.

Úroveň č. 2

Totéž jako v úrovni 1, ovšem na vstupu budou dva spustitelné programy pro Windows, tj. programy ve formátu PE (portable executable) , s velikostí max. 1MB
Poznámka: Podobnost u takového typu souborů (PE) bývá obvykle způsobena

- Stejným původem, tj. ze stejných zdrojových kódů, ovšem částečně změněných. Tou změnou může být aktualizace či oprava chyby, jiná jazyková verze anebo záměrně způsobený polymorfismus. Polymorfismu se na úrovni zdrojových kódů dosahuje výměnou datových typů za jinou bitovou šířku, různými direktivami pro překladač (např. úroveň optimalizace pro rychlost) apod.
- Dodatečnou úpravou již přeložených programů různými obfuskátory a kryptory.
- Sdílením veřejně dostupných či placených knihoven funkcí, jako OpenSSL apod. Tyto knihovny jsou přítomny v binárce programu a tvoří určitou míru podobnosti pro všechny programy, které danou knihovnu využívají.

Úroveň č. 3

Navrhněte aplikaci, které s využitím aplikace z úrovně 2 identifikuje v zadaném balíku programů skupiny dle podobnosti.

Vstup:

Balík spustitelných programů pro Windows, v množství cca 100 až 1000 kusů. V balíku budou soubory z několika rodin malware (čili soubory s velkou podobností) a také soubory, které jsou individuálního původu a nemají podobnost s ničím ostatním.

Výstup:

Seznam všech souborů na vstupu, ovšem rozdělený do několika skupin. V každé skupině budou ty programy, které aplikace určí jako příbuzné (tj. pocházející z jedné rodiny malware). Aplikace také do zvláštní skupiny vyčlení ty soubory, u kterých není zjevná podobnost s čímkoliv ostatním.

Hodnocení

Úroveň č. 1

Ve finále se bude porovnávat 20 párů souborů. Při vyhodnocení se sečtou odchylky výsledků týmu od aritmetického průměru hodnot vypočtených týmy.
Body se udělí proporcionálně dle výsledků ostatních týmů. Maximum bodů je 10.

Úroveň č. 2

Ve finále se bude porovnávat 20 párů programů. Při vyhodnocení se sečtou odchylky výsledků týmu od aritmetického průměru hodnot vypočtených týmy.
Body se udělí proporcionálně dle výsledků ostatních týmů. Maximum bodů je 100.

Úroveň č. 3

Body se udělí na základě správného přiřazování do kategorií. Za každé správné přiřazení je bod. Poté se počet bodů vynásobí vhodným koeficientem, aby měl nejlepší tým maximum bodů. Maximum bodů je 1000.

Pokud program týmu nebude schopný řešit úlohy určité úrovně, tým získá v úrovni 0 bodů. Výsledné bodové hodnocení je součet bodů ze tří úrovní.

Výsledné hodnocení se skládá z 50% z bodů a z 50% z názoru poroty.

Výpočet hodnocení pro úrovně 2 a 3

Pro každý pár souborů se vypočítá aritmetický průměr výsledku všech týmů.
Pro každý tým se sečtou odchylky od průměru odchylek všech týmů.
Sečtou se odchylky všech týmů.
Pro každý tým se spočítá: **a** = celkový součet odchylek / součet odchylek týmu
Pokud je součet odchylek týmu roven 0, pak **a** = maximum bodů v úrovni.
Vypočítá se koeficient **k** = maximum bodů v úrovni / nejvyšší **a**
Všechna a se vynásobí **k** a to je výsledný počet bodů v kategorii

Průběh soutěže

Na vypracování zadání je 8 hodin čistého času. Dostanete také vzorové soubory.
Po 8-mi hodinách dostanete reálné soubory, na jejichž porovnání a výpočet čísel máte 30 minut.
Poté následují krátké 5 minutové prezentace, kde odprezentujete svoje řešení.

Pravidla

- Je povoleno používat všech dostupných zdrojů informací včetně internetu.
- Všechny týmu musí pracovat samostatně bez cizí pomoci a rad mimo tým.
- O výsledku rozhodne porota.
- Pořadatel má právo diskvalifikovat ze soutěže tým, který jedná proti duchu zadání
- Jsme studenti techniky, ne práva, proto platí smysl věty v zadání a ne „slovíčkaření“
- S týmy ohledně zadání komunikuje pouze Topic Responsible
- V případě nejasností na dotazy odpovídá Topic Responsible