Fakulta elektrotechniky a informatiky
Vysoká škola báňská - Technická univerzita Ostrava


# Detekce malware na základě informací z PE hlaviček

**ING. JIŘÍ ZNOJ**

# Úvod

- téměř 50% světové populace má přístup k internetu
- oproti roku 2000 se jedná o 936% nárůst uživatel internetu
- malware = malicious software (škodlivý software)
- V prvním čtvrtletí 2017 - více jak 30 milionů nových vzorků malware detekováno firmou MacAfee
- 357 milionů variant malware bylo dosud objeveno firmou Symantec

# Úvod

- **PE (Portable Executable)** – formát spustitelných souborů a knihoven používaný ve Windows
- **PE Header** – kolekce dat definující obsah PE souboru
- podle Yujie Fan a kol. chování PE souborů může být dobře reprezentováno pomocí API, což je jedna z nejlepších metod heuristické analýzy

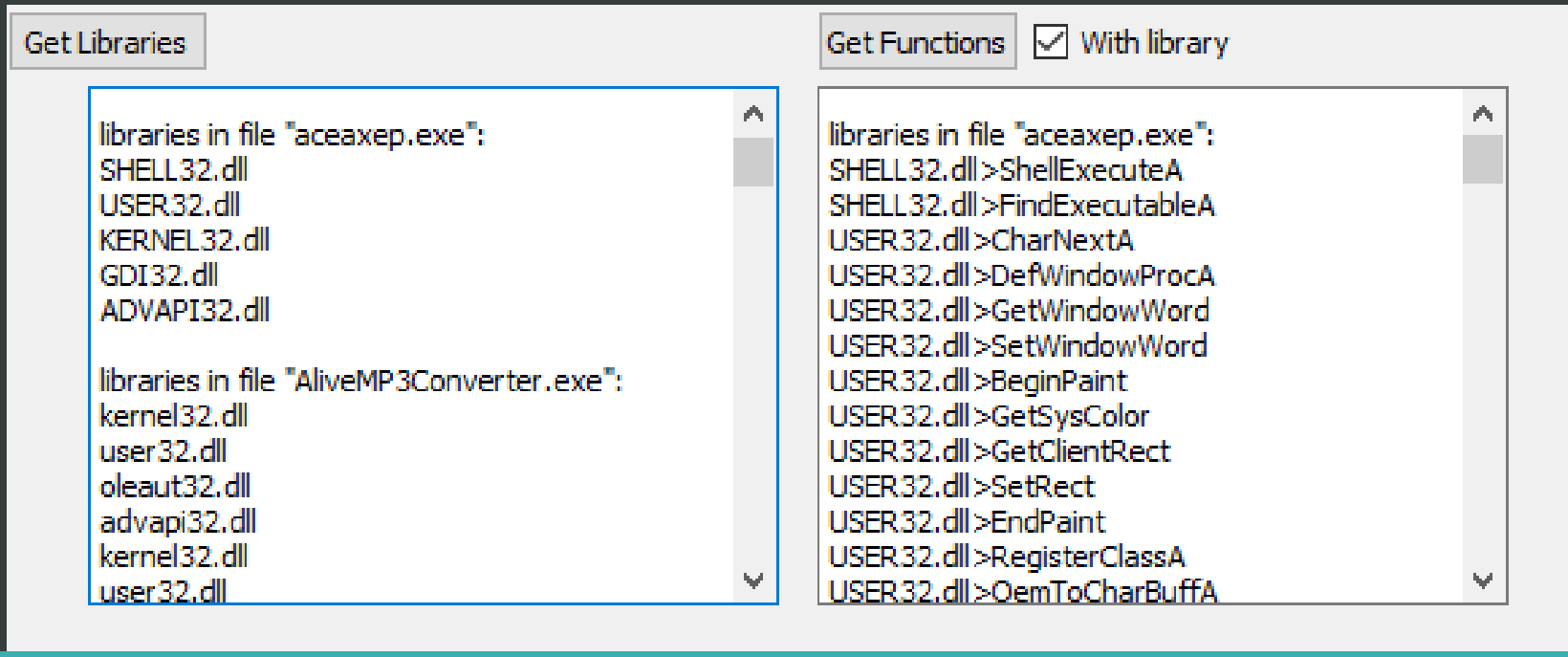
# Vybrané související práce

- D. Schultz a kol. (2001) – myšlenka využít PE soubory pro detekci malware
- Kolter a kol. (2006) – data z PE souborů reprezentována jako n-gramy a dosažení lepších výsledků
- Shafiq a kol. (2009) – framework PE-Miner, předzpracování dat, optimalizace algoritmů
- Kolosnjaji a kol. (2016) – použití neuronových sítí pro počítání volání API z PE hlaviček v kombinaci s výstupem ze sandboxu.

# Antivirus k detekci „Zero day“ útoků

# Knihovna pro čtení informací z PE souborů

- psáno v C++
- podpora 64-bitových souborů
- výstup v požadovaném tvaru bez redundantních dat



# Sběr testovacích dat

- Malware
  - Microsoft Malware Classification Challenge (BIG 2015) – více jak 17 GB testovacích dat, více jak 17 GB trénovacích dat
  - VX Heaven – více jak 270 000 vzorků (62 GB)

- Goodware
  - Skript pro úložiště studna, sourceforge a filehippo

```
http://www.willingsoftware.com/download/willingwebcam553.exe
step35.exe - download finnished.
http://www.labf.com/download/aceaxep.exe
http://www.avs4you.com/downloads/AVSDVDPlayer.exe
http://ibrowser.fcodersoft.com/ftp/ibrowser.zip
willingwebcam553.exe - download finnished.
http://www.brooksnet.com/software/intelliscribe/intelliscribe.exe
ipscan-win32-3.3.1.exe - download finnished.
```

# Vybrané metody použitelné k detekci malware

- Statistical modeling (Naïve Bayes)
- Decision trees construction (C4.5)
- Constructing rules (RIPPER)
- Linear models (Support vector machines)
- Instance-based learning (k-nearest-neighbors)
- Clustering (K-means)
- Transformations (boosting, bagging, pruning)

# Závěr, plány do budoucna

- Vyhledány a prozkoumány výzkumy zabývající se detekcí malware z PE hlaviček
- Zmapování použitelných a použitých technik k detekci malware z PE hlaviček
- Vytvoření knihovny pro čtení dat z PE hlaviček
- Vytvořen skript na získávání goodware programů, vytváření dostatečného množství testovacích dat

# Závěr, plány do budoucna

- implementace existujících algoritmů
- implementace nových, nebo modifikovaných existujících algoritmů a jejich porovnání s existujícími
- cílem je funkční řešení schopné detekovat některé „zero-day“ útoky