Runtime Packers: The Hidden Problem?

Tom Brosch, Maik Morgenstern
AV-Test GmbH
(www.av-test.org)

Table of Content

- Motivation – what's the problem?
- Issues involved – our test results
  - Detection rates
  - False positives
  - Crashes and other problems
- Conclusions

What's the Problem? (I)

- How many of the malware files out there are runtime-compressed?
  - WildList 03/2006 (www.wildlist.org): Over 92%
  - Only 54 out of 739 files are not packed, according to a quick analysis with PEiD and a manual review
- About 30 different packers and crypters are used
- Some of the top ones are:
  - UPX: 167 files
  - Morphine: 72 files
  - MEW: 59 files
  - FSG: 50 files
  - PESpin: 32 files

What's the Problem? (II)

- A review of some common malware families:
  - Bagle: 62 out of 63 files are runtime packed - 5 different packers, in 7 different versions have been used
  - Mytob: 241 out of 246 - 20 packers, in 32 different versions
  - SDBot: 58 out of 58 - 12 packers, in 16 different versions
- Observation:
  - Nearly every malware is runtime-compressed
  - Many different packers are used throughout one malware family to avoid easy heuristic and 0day detection
- Conclusion:
  - Anti-virus software needs to deal with a lot of packers and have to be prepared for new ones every day

What's the Problem? (III)

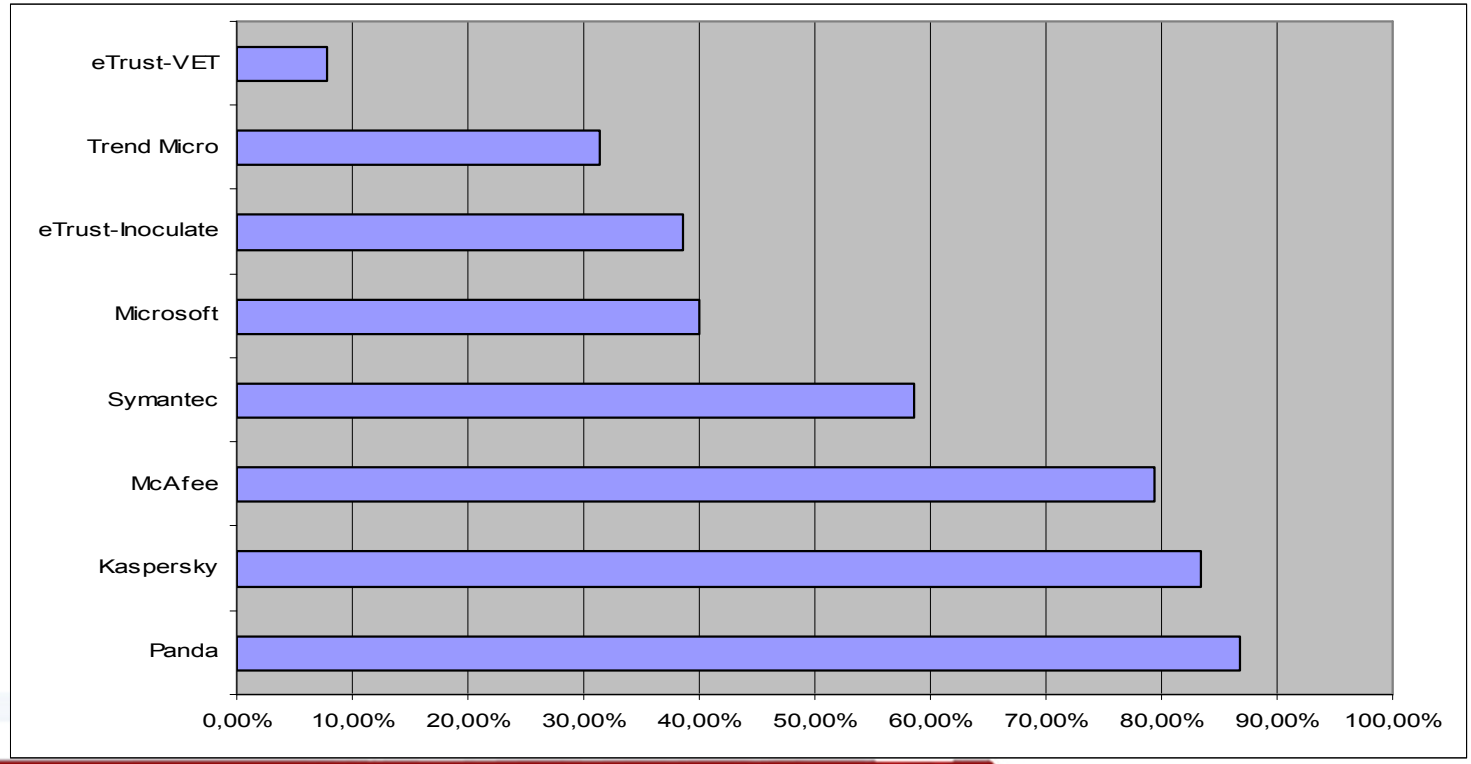
- Nearly all AV products employ unpacking engines. So everything is fine? No, it's not! Why? The engines have many flaws, aren't generic and have a hard time keeping up.
- There is a lot of activity in research in this area:
  - *Defeating polymorphism: beyond emulation*, Adrian E. Stepan *(Microsoft),* Virus Bulletin Conference 2005
  - *Generic unpacking – how to handle modified or unknown PE compression engines?,* Tobias Graf *(Ewido Networks),* Virus Bulletin Conference 2005
  - *Unpacking - a hybrid approach,* Vanja Svajcer, Samir Mody *(Sophos),* Eicar Conference 2006
- Anyway, detection rates are not so good…
  - Microsoft OneCare: 41%
  - Ewido Anti-Malware: 73%
  - Sophos Anti-Virus: 30%

Test Results

- Used test sets and test setup
  - Malware test set: 10 common malware files, packed with about 40 different runtime packers in over 500 versions and options: out of approx. 5000 generated files, **2941** manual-reviewed files were finally used for the test (the remaining samples didn't work properly)
  - False positive test set: 10 clean files (parts of Windows and standard applications), packed with the same options as above
  - For details of used packers and versions see additional material
  - Test setup of anti-virus products:
    - 27 command-line versions in an automated environment
    - 7 GUI versions tested manually on Windows XP SP2 (English)
    - Latest updates and signatures from around June 20th were used (for exact version details see additional material)
    - Test systems were 15 identical Pentium IV 2.8 GHz with 512 MB and 250 GB hard disk drives

Detection Rates (I)

- Malware test set results range from 10% to over 80%
- But: be prepared for false positives on several products!



Detection Rates (II)

- Packers used in WildList malware
  - Interestingly the products perform pretty good (nearly always over their own average) on packers used in WildList'ed malware (= common malware files)
  - However, they perform usually worse on packers not used on WildList'ed malware
  - Many packers aren't detected at all by some products

Detection Rates (III)

- Top 5 packers (WildList malware):

	Kaspersky	McAfee	Symantec	Microsoft
Average on the malware test set	83%	79%	58%	39%
ASpack	95%	97%	95%	81%
FSG	100%	100%	56%	100%
Morphine	100%	70%	100%	0%
UPX	96%	97%	92%	100%
MEW	100%	86%	53%	80%

Detection Rates (IV)

- "Poor" performance on other packers:

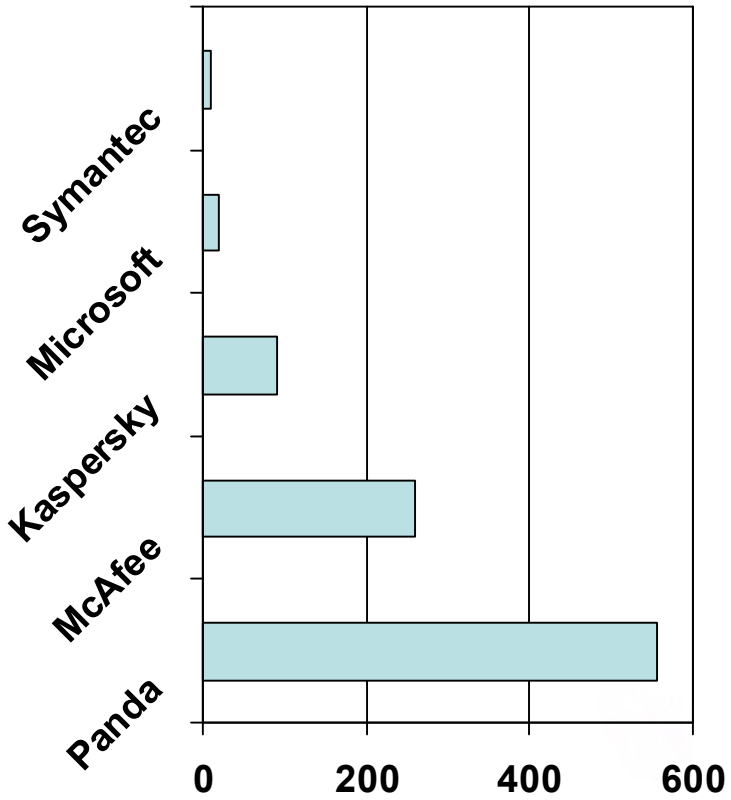
Kaspersky	McAfee	Symantec	Microsoft
Armadillo: 6%	Armadillo: 14%	Acprotect: 33%	ASProtect: 26%
ASProtect: 80%	Obsidium: 18%	Exe32pack: 18%	PEBundle: 14%
PEBundle: 81%	Yoda's Protector: 55%	Neolite: 22%	PECompact: 24%

Detection Rates

- Packers not detected at all
  - Microsoft: Armadillo, Krypton, Obsidium
  - Symantec: Armadillo, ASProtect, Shrinker
  - McAfee: Epack, PELock
  - Kaspersky: SVK-Protector
  - Trend Micro: Acprotect, Armadillo, Cexe

False Positives (I)

- "Suspicious" problem:
  - Panda had good detection rates (86%), but with a lot "suspicious" warnings (because the runtime packer got flagged), which in turn results to many false positives (556) now, since the packer gets flagged again. The same issues occurred on eSafe (2091 false positives) and Fortinet (1854 false positives), for example.

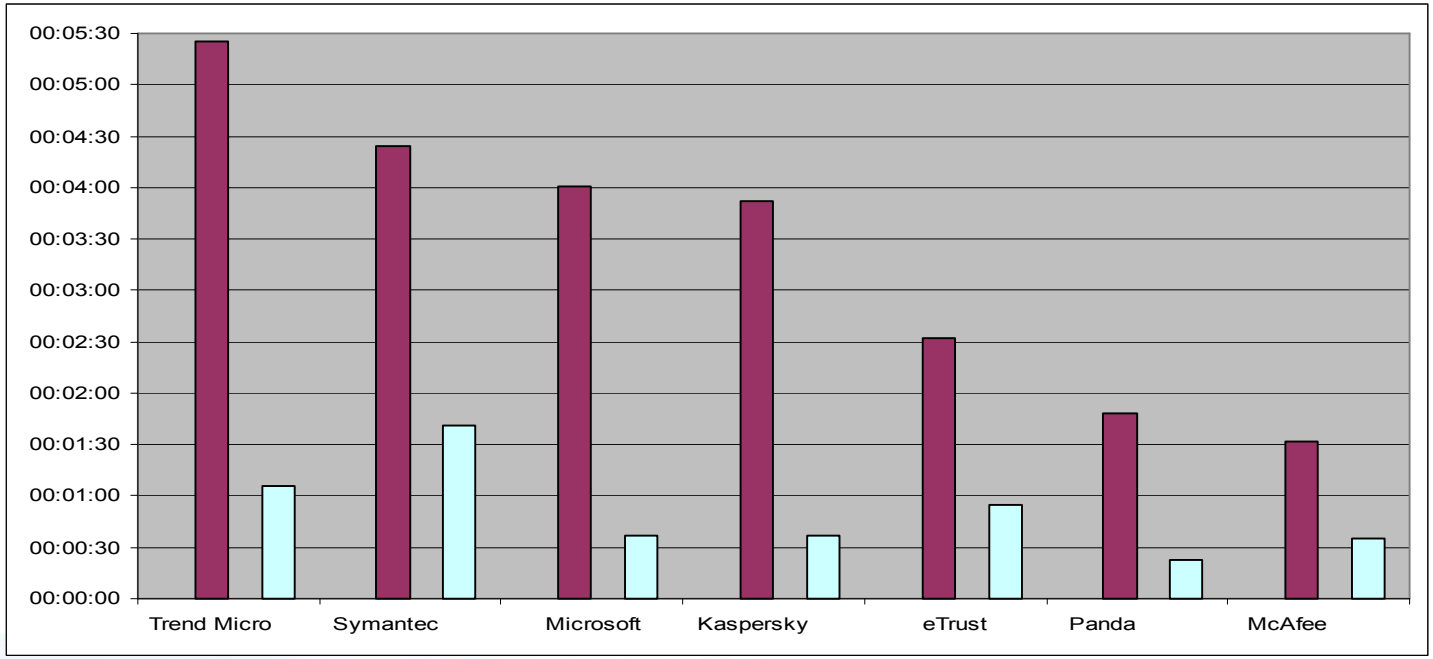


False Positives (II)

- Only Trend Micro and CA eTrust triggered no false positives on our test set used
- Other products ranged between ten and several hundred up to 2091 (as shown before)
- Common false positives:
  - Exebundle gets flagged as malicious by McAfee, Microsoft, Kaspersky and Panda
  - Many scanners wrongly flag packed files as certain malware
    - Kaspersky, for example, flagged several Armadillo-packed files as ‘Backdoor.Win32.Agobot.afn’ or ‘Backdoor.Win32.Rbot.ip’

Crashes and other Problems (I)

- Scanning speed
  - Dr. Web took around 20 seconds on 2941 non-packed files, but over 2 hours (!) on the same files in a packed state
  - Trend Micro needed 1 minute vs. 5 minutes
  - Usual increase in scanning time is between the factor 1.5 and 10



Crashes and other Problems (II)

- Two of the problems we faced in our test:
  - Panda (GUI version):
    - When scanning certain files packed with Cexe packer, the scan simply stops at that file without any error
    - So just place the Cexe ahead of your malicious files and they won‘t be scanned
  - Trend Micro command-line scanner:
    - When scanning certain files packed with Petite, the scan will stall
    - Easy DoS attack possible if you “accidentally” send a file like that to an e-mail gateway running a Trend Micro product
- For other problems in security software see: *Insecurity in Security Software*, Maik Morgenstern, Andreas Marx (AV-Test GmbH), Virus Bulletin Conference 2005

Conclusions (I)

- Three main issues:
  - Detection rates: detection of packers commonly found ‘in the wild’ is OK, however, detection of other (less commonly used) packers still needs to get a lot better!
  - False positives: nearly all products triggered false positives and some just flag many packed PE files as being “suspicious”. Also, several false positive-causing files were detected as a certain malware (bad signatures are in use).
  - Crashes and speed problems: scanning packed files increases scanning times and the system load a lot. Some scanners had serious problems when scanning packed files. We had several security problems with archive files last year, so can we expect runtime packer problems next year?

Conclusions (II)

- Proposals:
  - AV vendors need to support runtime packers not found on the WildList, or else virus writers will just switch to yet undetected packers (in order to avoid 0day detection)
  - Of course it's not possible to catch every packer (version) out there, so heuristic or generic approaches should be combined with the dedicated unpacking engines
  - But some heuristic approaches need to get a lot better than just flagging all packed PE files (this might work on a gateway level, but not on desktops or server systems)
  - In addition to this, signatures need to be more carefully chosen to avoid false positives that way
  - Possible problems in unpacking engines should be reviewed and fixed to avoid the issues we have seen with archive files last year

The End

Thank you for your attention.

Any questions?

Tom Brosch, Maik Morgenstern

www.av-test.org