PE Trojan Detection Based on the Assessment of Static File Features

Wei, Wang
David@antiy.net

Static Assessment vs. Traditional Detection

- Traditional signature detection technology:

  Build a complete sample database and extract malware signatures. Signature based detection is the basis of virus and Trojan detection.

- Static assessment model using neural algorithm:

  Use an intelligent algorithm to analyze and study known samples, extract signatures and assess unknown samples.

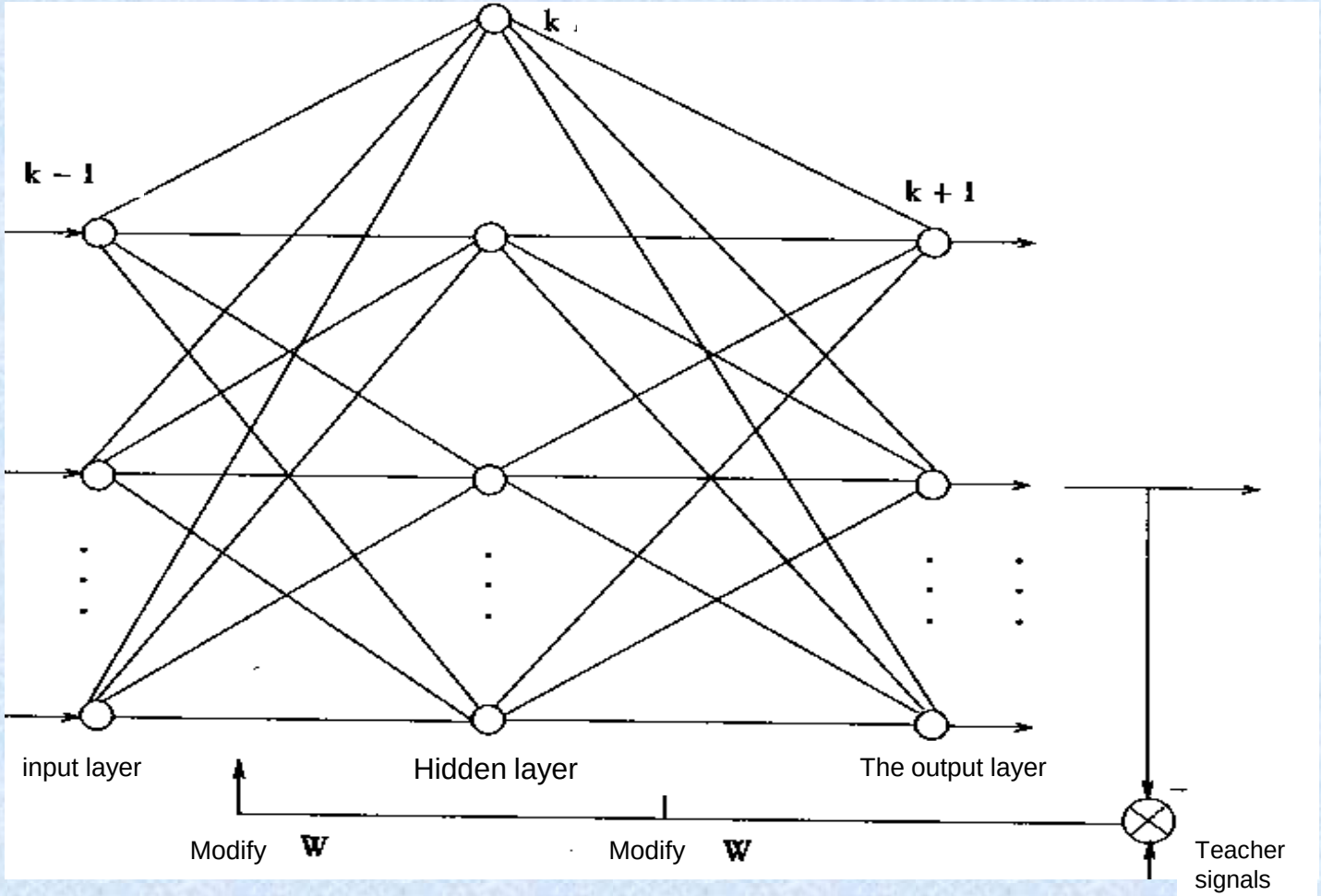
Can static assessment complement signature code detection?

	Signature	Assessment
Virus Database	needed	unneeded
Program Updates	rare	required
Update Downloads	required	rare
Update Size	> 100 KB	<1KB
Unknown Sample	specific signature	target
Accuracy	100%	? ? %
Speed	high	? ?

Building an Intelligent Assessment Model

- Ultimate goal: recognize Trojan files
- Method: develop a categorization machine
- Build a BP (reverse transmission) neural network, then use a batch of sample files to test it; eventually get a reasonable neural network
- Result: Improved rate of Trojan detection

Neural Network Model



Primary Static Assessment Searches

- Script file assessment based on grammar
- Binary file assessment based on file association
- The 2nd one aims at PE files, which is our focus.

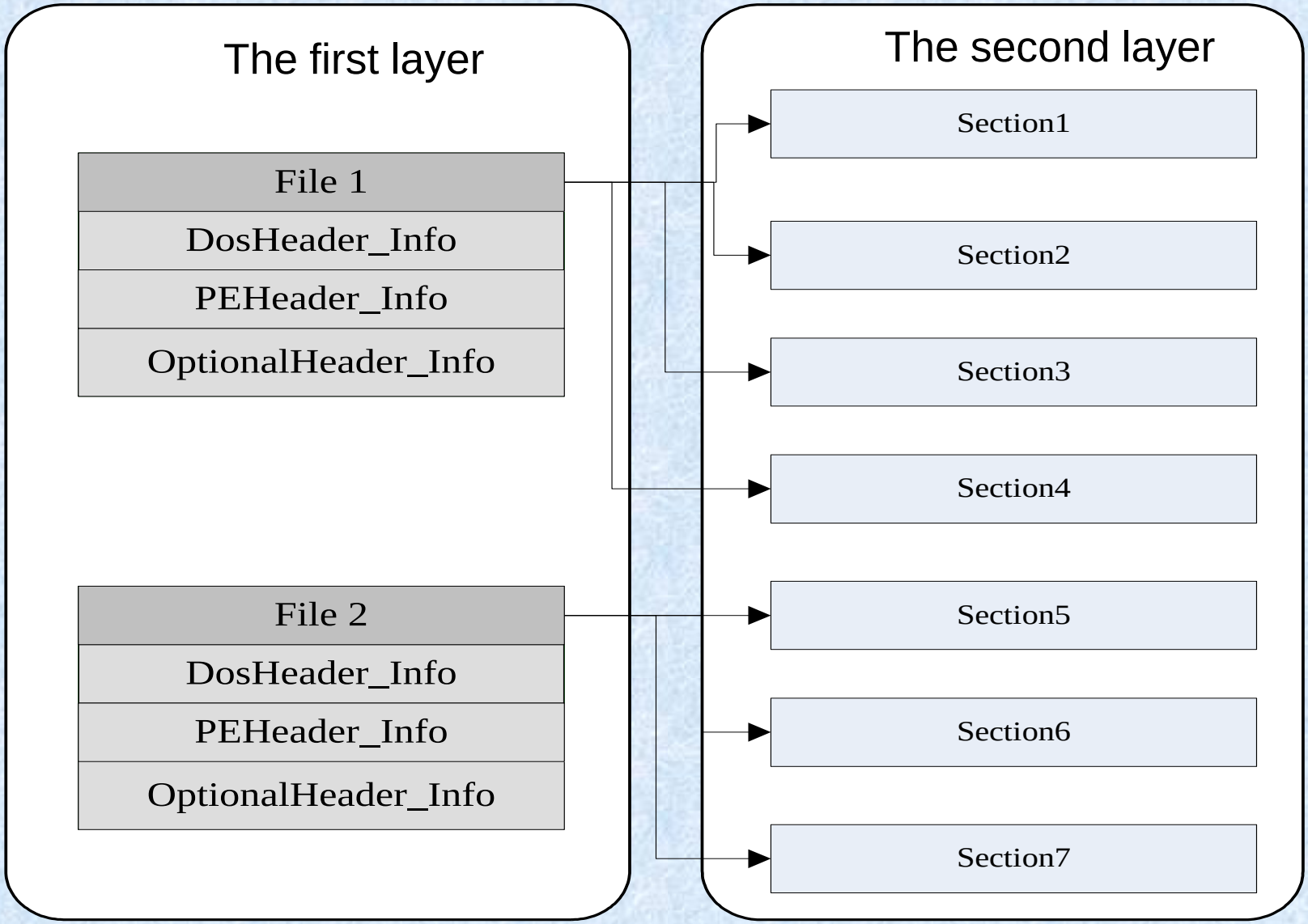
The Structure of PE Files

PE file information:

- PE DosHeader
- PE OptionalHeader
- PE SectionHeader
- PE ImportSectionHeader
- … …

Conflicts Between PE Structure and BP Network

BP Network Requirement	PE File
One dimensional I/O	Several one-to-many relationships One FileHeader corresponds to several SectionHeaders. The input header corresponds to several input functions.
Data Type	Attributes based on strings (names of input functions)
Normalized Data [0,1]	Various data types


The first layer
File 1
DosHeader_Info
PEHeader_Info
OptionalHeader_Info
File 2
DosHeader_Info
PEHeader_Info
OptionalHeader_Info
The second layer
Section1
Section2
Section3
Section4
Section5
Section6
Section7

Solutions for a Multi-Dimensional Structure

Hierarchical tests based on changes of granularity:

1. Test attributes of the same level (one-to-one structure) individually, ordered from lowest to highest granularity
2. The low granularity attributes participate in the higher levels of tests
3. Until the highest granularity is achieved. That is, the file correspondence is one to one

String Attribute Normalization

- In PE files, there are plenty of import function names, while the BP network input data range is [0,1]
- Introduce a method of statistical probability on the imported function names, and then normalize them. The probability value will range between 0 and 1 to meet the needs of the neural network.

Experiment Result

- Samples in the test:
  - Trojan Files: 158
  - Normal Files: 228
  - Unstudied Trojan Samples: 152
  - Unstudied Normal Samples: 1000
- Accuracy rate of unknown file detection is 90% higher.☺
- The false positive rate with normal files is 7%.☹

Result Analysis

- The alarm rate for unstudied Trojans is good
- If the false positive rate can’t be brought below 0.05%, it is not applicable.

Demo Time

- Now it is time for a demonstration, you can go the bathroom or get some coffee.

Where Are We?

- Currently, we are in the experimental phase.
- Our network has been used in filtering unknown files (we receive malware reports from 100,000 users, network sniffers and honey pots per month).
- The early results indicate that we are headed in the right direction and our work is of value.
- It's a lot of work to extract PE file information. Currently, we only analyze the DosHeader, FileHeader, OptionalHeader, SectionHeader, and ImportSection. Some important sections such as ResourceSection and ExportSection haven't been analyzed.
- It hasn't yet been integrated with our automatic signature extraction mechanism.

Some Thoughts

- Static assessment based on the neural network should be able to defy all unknown viruses with a 50% alarm rate.
- We can see that avoidance of such mechanisms is not difficult. But then, no Trojan detection technique can defeat targeted avoidance.
- We hope in April or May of 2005, this technology will be much more mature.

Thank You