MALWARE

COLETA DISTRIBUÍDA
E
PRÉ-CLASSIFICAÇÃO

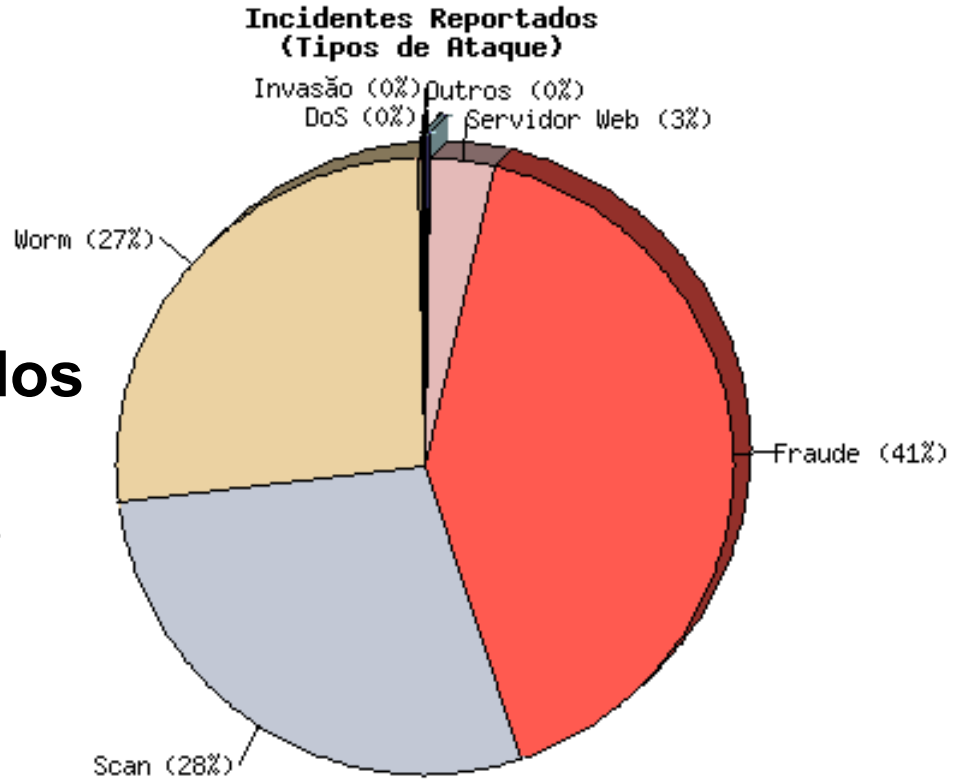
André Grégio, Unicamp
Antonio Montes, CenPRA

Agenda

- Disseminação de *malware*
- Problemas na identificação de *malware*
- Coleta
  - Coleta Distribuída
- Separação de *malware*
- Classificação

Motivação [I]

Incidentes Reportados ao CERT.br Jan. a Mar. / 2008



http://www.cert.br/stats/incidentes/2008-jan-mar/tipos-ataque.html

Motivação [II]

Motivação [III]

Relatório de Pendências Financeiras Spam | X

Brasil Telecom S.A. to me show details May 18 (1 day ago) Reply

Notificação

Comunicamos que seu (CPF/CNPJ) consta em nossos cadastros por motivo de pendências financeiras, com a instituição abaixo relacionada.

(BrasilTelecom S.A.)

Total de Pendências: R$ 1.754,74.

Para sua segurança e praticidade e necessário ABRIR o arquivo do relatório de pendências.

Relatório de Pendências Financeiras **Abrir Relatório**

Se você efetuou a regularização, favor desconsiderar.

Sergio Spinelli Silva Jr
Presidente

Motivação [IV]

AVG	7.5.0.516	2008.05.20	Generic.QFK
BitDefender	7.2	2008.05.20	-
CAT-QuickHeal	9.50	2008.05.19	(Suspicious) - DNAScan
ClamAV	0.92.1	2008.05.20	PUA.Packed.Expressor
DrWeb	4.44.0.09170	2008.05.20	Trojan.DownLoader.25791
eSafe	7.0.15.0	2008.05.19	suspicious Trojan/Worm
eTrust-Vet	31.4.5806	2008.05.20	-
Ewido	4.0	2008.05.20	Downloader.Banload.bes
F-Prot	4.4.2.54	2008.05.16	W32/NewUnknownMalware-P89!Maximus
F-Secure	6.70.13260.0	2008.05.20	Hupigon.gen83
Fortinet	3.14.0.0	2008.05.20	-
GData	2.0.7306.1023	2008.05.20	Win32:VB-AQI
Ikarus	T3.1.1.26.0	2008.05.20	Backdoor.Win32.Poison.ahf
Kaspersky	7.0.0.125	2008.05.20	-
McAfee	5298	2008.05.19	New Malware.dq

File **Relatorio.ExE** received on **05.20.2008 15:23:45 (CET)**
Current status: **finished**
Result: **21/32 (65.63%)**

Commom Malware Enumeration [I]

- Iniciativa gerenciada/mantida pelo MITRE:
  - Prover identificadores comuns para *malware* na tentativa de indexá-los;
  - Reduzir a confusão ao se referenciar ameaças em incidentes envolvendo *malware*;
  - Melhorar a comunicação e compartilhamento de informações entre *antivirus vendors* e a comunidade de segurança.

Commom Malware Enumeration [II]

- Identificadores CME => “ameaças de *malware*”.
  - Uma ameaça pode ser identificada por uma assinatura, pode (ou não) explorar uma vulnerabilidade, depender da ação de usuário…
  - Uma ameaça engloba os arquivos envolvidos em uma “epidemia” de *malware*:
    - Ex.: os componentes do Nimda -- ISS *buffer overflow byte stream*, o arquivo passado via TFTP, etc. -- são referenciados por um único identificador CME.

Commom Malware Enumeration [III]

- Fragmento da lista CME:

CME-ID	Aliases	Description
CME-711	*Aladdin*: Win32.Small.dam *Authentium*: W32/Downloader.AYDY *AVIRA*: TR/Dldr.Small.DBX *CA*: Win32/Pecoan *ClamAV*: Trojan.Downloader-647 *ESET*: Win32/Fuclip.A *Fortinet*: W32/Small.DAM!tr *F-Secure*: Small.DAM *Grisoft*: Downloader.Tibs *Kaspersky*: Trojan-Downloader.Win32.Small.dam *McAfee*: Downloader-BAI!M711 *Microsoft*: Win32/Nuwar.N@MM!CME-711 *Norman*: W32/Tibs.gen12 *Panda*: Trj/Alanchum.NX!CME-711 *Sophos*: Troj/DwnLdr-FYD *Symantec*: Trojan.Peacomm *Trend Micro*: TROJ_SMALL.EDW	CME-711 is a Trojan Downloader that is spread as an attachment to emails with news headlines as the subject lines which downloads additional security threats,

http://cme.mitre.org/data/list.html

Coleta de *malware* [I]

- *Nepenthes*: *honeypot* de baixa interação que emula vulnerabilidades conhecidas e efetua *download* de *malware* tentando explorá-las.
  - Módulos para *parsing* de *shellcodes* -> extrai informações sobre o *malware* em propagação;
  - Módulos de *fetching* -> usa informações de *shellcode* para fazer o *download* via HTTP, FTP, TFTP;
  - Emulação uma *shell* de *Windows*;
  - Pode enviar *malware* para uma *sandbox* remota.
- Unicidade baseada em MD5…

Coleta de *malware* [II]

- *Donutd*:
  - *Daemon* escrito em *python* pela equipe de segurança do CenPRA para *download* de *spam* e integração com a arquitetura de coleta de *malware* (*Pighunter*).
    - Módulos:
      - main.py
      - getMailIMAP.py
      - getMailPop.py
      - unpack.py
        - » Armazena as mensagens e os anexos

Coleta Distribuída [I]

- *Nepenthes* + *Honeyd* em modo *proxy*

```
[…]
add windows tcp port 110 proxy <IP_Nepenthes>:110
add windows tcp port 113 proxy <IP_Nepenthes>:113
add windows tcp port 135 proxy <IP_Nepenthes>:135
add windows tcp port 137 proxy <IP_Nepenthes>:137
add windows tcp port 138 proxy <IP_Nepenthes>:138
add windows tcp port 139 proxy <IP_Nepenthes>:139
add windows tcp port 143 proxy <IP_Nepenthes>:143
add windows tcp port 220 proxy <IP_Nepenthes>:220
add windows tcp port 443 proxy <IP_Nepenthes>:443
add windows tcp port 445 proxy <IP_Nepenthes>:445
add windows tcp port 465 proxy <IP_Nepenthes>:465
add windows tcp port 993 proxy <IP_Nepenthes>:993
[…]
```

Coleta Distribuída [II]

- Aproveitamento de recursos computacionais:
  - *Honeypots* distribuídos podem encaminhar conexões para um *Nepenthes.*
- Análise de tendências:
  - o *malware* que atacou a rede A nos serviços XY é o mesmo que atacou a rede B ou é uma variante?
- Se os *logs* não forem filtrados, armazenam a conexão via *proxy.*

Coleta Distribuída [III]

- Classe C com ~70 endereços em *proxy*;
- Período: 1 mês.

Nepenthes-Single	**Nepenthes+honeyd**
~23 *downloads*/dia	~71 *downloads*/dia
Novembro/2007	**Dezembro/2007**

Coleta Distribuída [IV]

- Espaço amostral de 100 *malware* com MD5 único:
  - 12 não foram identificados pelo antivírus
    - 5 corrompidos
    - 7 inéditos
  - E os outros 88? MD5 único…

MD5 único…

- 5865e732663d75b501ffd7d98bc49005
  - Trojan horse BackDoor.RBot.BI
- 65258e79ee78ac97dc33dd3a7cd482ff
  - Trojan horse BackDoor.RBot.BI
- 9a119463e60ced4488e4309f37dd2d02
  - Trojan horse BackDoor.RBot.BI
- C819d7746a18cbcaca2ff066a31086e9
  - Trojan horse BackDoor.RBot.BI

Triagem de *malware* [I]

- Definição de algumas classes iniciais:
  - *Backdoor*;
  - *Trojan;*
  - *Worm;*
  - *Downloader;*
  - *Bot;*
  - *Outros;*
  - *Não identificados.*

Triagem de *malware* [II]

Antivirus			**Result**
Avast	-	-	Win32:Kolab-S
AVG	-	-	Dropper.Delf.ACL
ClamAV	-	-	**Worm**.Kolab-110
F-Secure	-	-	Net-**Worm**.Win32.Kolab.ep
Fortinet	-	-	W32/Kolab.EP!**worm**.im
Kaspersky	-	-	Net-**Worm**.Win32.Kolab.ep
McAfee	-	-	Generic.dx
Panda	-	-	-
Symantec	-	-	**Trojan Horse**

Triagem de *malware* [III]

- Diretórios nomeados por classes servem de repositório para *malware* separados por "funcionalidade";
  - Há um diretório onde são armazenados os *malware* não identificados.
    - Corrompidos?
    - *Packed*?
  - Próximo passo: identificar padrões…

Idéia Geral

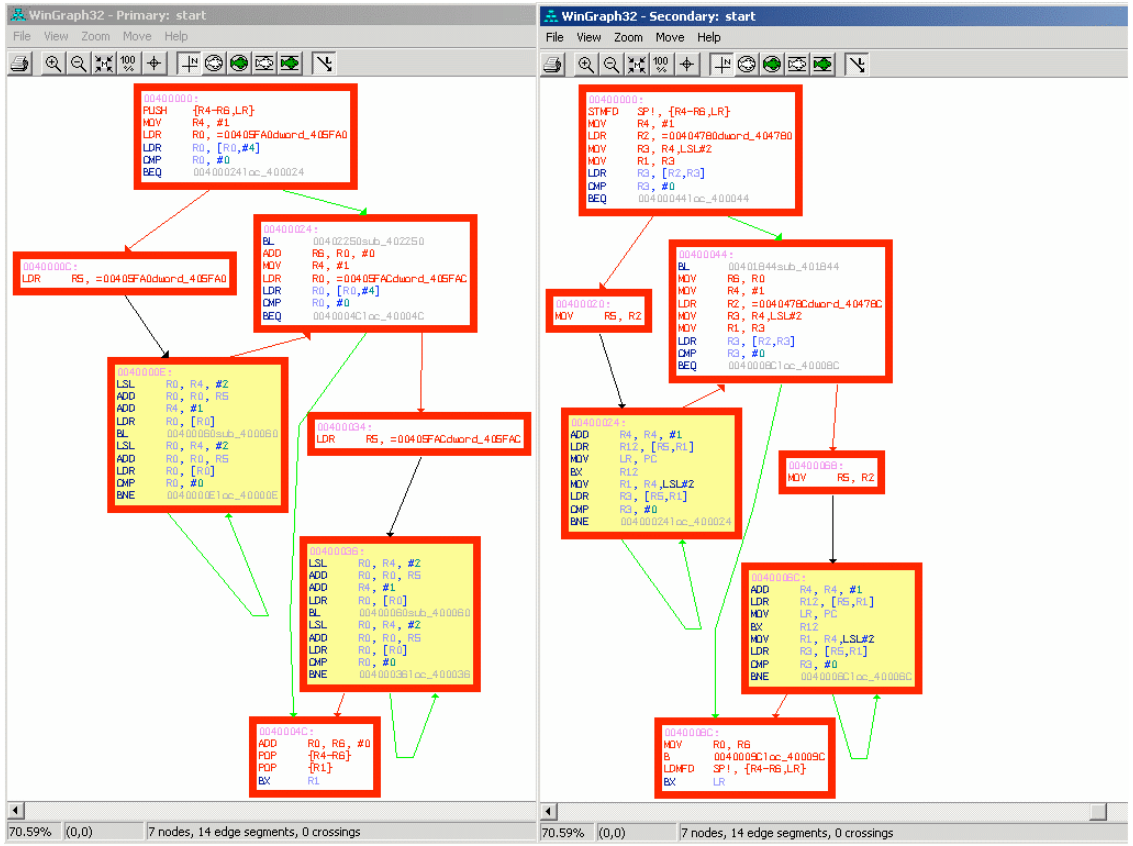
- O que faz com que um *malware* seja classificado como sendo de um tipo ou de outro?
  - Reformulando: Sua categoria/comportamento indica alguma similaridade no binário?
- Se eu tiver um repositório com elementos separados por categoria, posso comparar os não classificados e buscar uma aproximação

Classificação de *malware* [I]

- Técnica baseada na complexidade de Kolmogorov:
  - Compressão dos binários executáveis (bzip2, …)
    - A complexidade de uma *string* é o comprimento da menor *string* que a descreve.
  - *Cluster* com diferentes versões de um *malware* formando uma família
    - Geração de árvore para análise preliminar;
    - Pode ajudar mesmo no casos de *malware* com UPX.

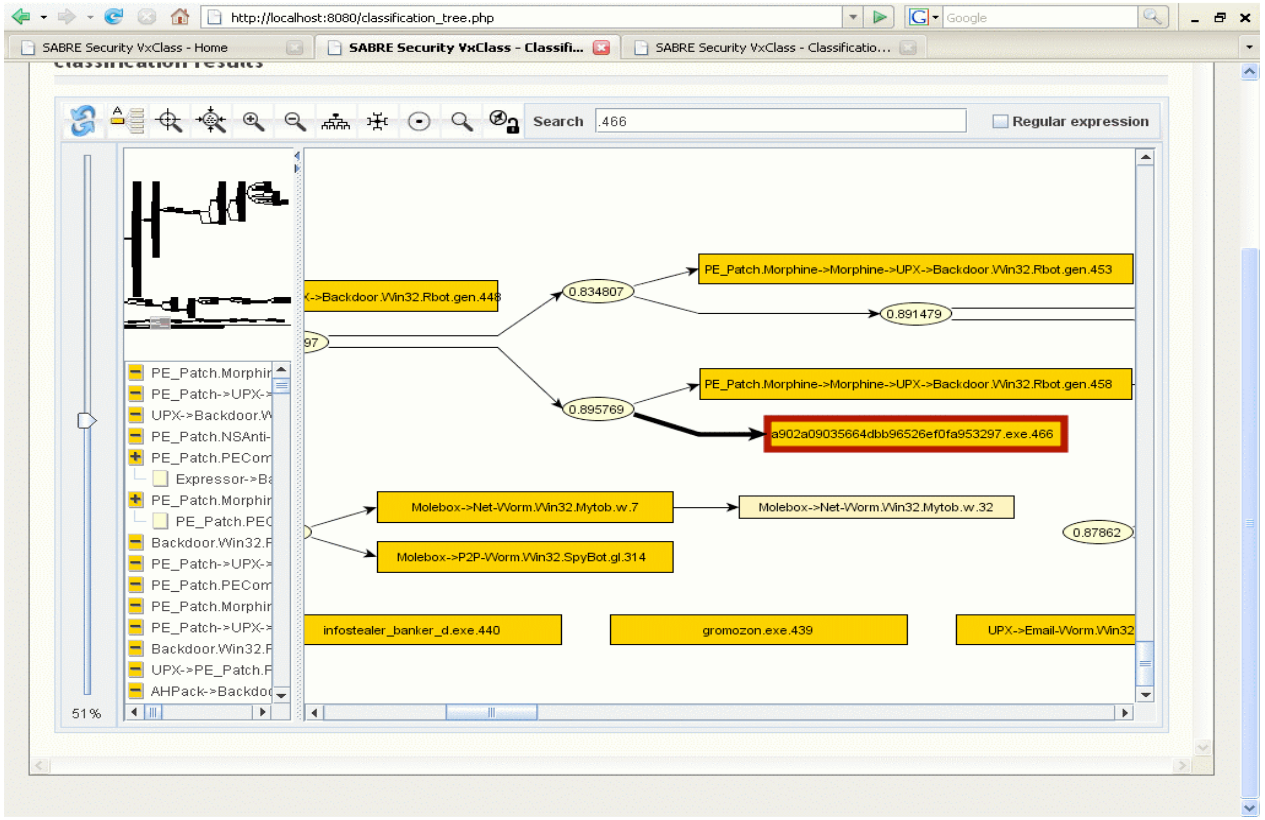
(Wehner, S. Analyzing Worms and Network Traffic using Compression, http://arxiv.org/abs/cs/0504045)

Classificação de *malware* [II]



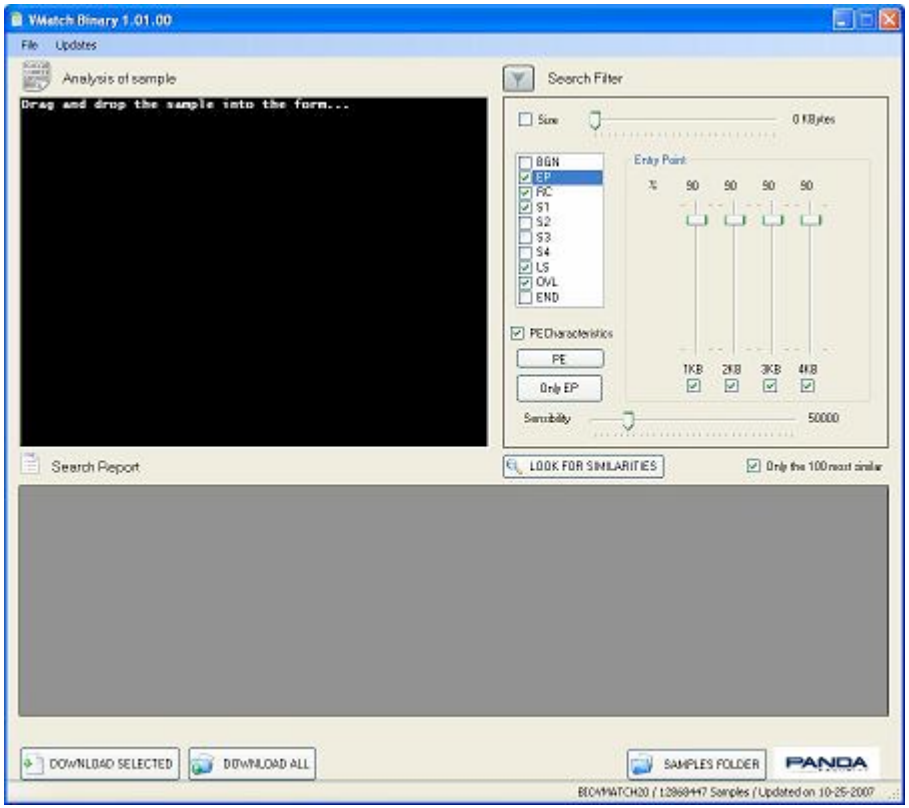
http://www.zynamics.com/content/_images/vxclass_scr4.gif

Classificação de *malware* [III]



http://www.zynamics.com/index.php?page=vxclass

Classificação de *malware* [IV]



http://pandalabs.pandasecurity.com/archive/Automatic-classification-of-malware-_2800_II_2900_.aspx

Classificação de *malware* [V]

- Engenharia reversa:
  - Busca por similaridades entre seções de código;
    - Mineração de dados;
    - Detecção de padrões em blocos comuns.
- Análise dos grafos do fluxo de execução.
- Análise comportamental:
  - Criação de perfis que indiquem programas maliciosos em execução.

Classificação de *malware* [VI]

- Malware 1:
  - `.idata:`
  - `|and|cli|popf|lock|jae|ja|adc|js|lea|jbe|xor|xor|imul|icebp|xor|cmc|addr16|xor|cld|outsl|rorl|sub|sub|jg|xchg|lcall|jb|push|in|incl`
- Malware 2:
  - `.text:`
  - `|mov|push|pushl|mov|xor|mov|dec|push|dec|jne|outsb|arpl|aaa|add|insl|xor|mov|aaa|sbbl|push|sub|lret|pusha|sahf|in|subl|and|inc|jle|pop|xchg|mov`

Conclusões

- Abordagens com grafos orientados ou algoritmos bioinformáticos podem ser o começo.

- Problemas para classificação:
  - *Packing;*
  - Falsos-positivos:
    - Um *botclient* pode não ser malicioso;
    - A calculadora do *Windows* com *packer* também…
  - *NOPs*.

Obrigado!

- Contato:

André Grégio
gregio@las.ic.unicamp.br

Antonio Montes
antonio.montes@cenpra.gov.br