

#### **Texture Vector**

#### Using Texture Vector Analysis to Identify File Similarity

### **Texture Vector Similarity**

• Can texture vectors imply similarity?



### **Texture Vector Anatomy**

- A Texture Vector is a vector of transforms on a section of data
- Transforms are:
  - Standard Deviation
  - Mean
  - Mode
  - Mode Count
  - Entropy
- Collectively, they produce the file's "fingerprint" as a spectrogram:

0 1	50000	100000	150000	200000	250000	300000	350000	400000
<b>- 1</b> 1								- H.
ΠU								
					, in the second s	_	a contraction of the second	

## **Similarity Measures**

• Texture-Vector similarity: Euclidean distance:

 $s = w_1(dv_1)^2 + w_2(dv_2)^2 + w_3(dv_3)^2 + w_4(dv_4)^2 + w_5(dv_5)^2$ 

219512

- File similarity measure = f(histogram(f(offsets between similar texture-vectors)))
  - histogram is from similar texture-vector offsets
     similarity measure is from histogram variance

#### Data

- 1,134 files
- 23 groups of similar files
- 642,411 similarity comparisons



Similarity to node 410: Stay in group, Standard Deviation similarity > 1.000

Node 410: /smallwork/bdallen/executable\_files\_500/nvrshu\_dll/MY01-023\_WINDOWS\_system32\_nvrshu.dll.tmp Group: 'nvrshu\_dll', Size: 262144, Modtime: 2011-04-07 17:16:28, MD5: 920AE502475DD6E93FD47B58956B190C



# Similarity by File Group

- Files within file groups are more similar than
  - files across file groups
    - usually

Family		-	1 1	) 2	-	4	5	6	-	7	8	0	10	11	12
20003775 dll	1	1	5 12	2 54	2	1 1	38	3.0		36	28	22	3.3	51	2.3
hthsery dll		1	2 35	1 12	0	7 0	0.7	0.7		0.5	0.7	0.6	0.3	0.0	0.6
ccalert dll		5	4 1 2		2	5	3.0	3.2		22	2.0	3.6	43	4.8	2.2
cdfview dll		2	1 07	2 25	10	0	13	1.1		1.6	2.9	1.8	0.9	1.7	0.8
dunzip32 dl1		3	8 0.7	3.9	1	3	51	23		4.0	2.1	2.0	3.4	3.6	1.6
hotfix exe		3	0 07	32	1	1	23	85		1.3	2.0	1.4	3.8	3.1	1.6
iexplore exe		3.	6 0.5	5 2.2	1	6	4.0	1.3	1	30.2	9.3	1.6	2.4	1.5	7.5
mobsync exe	8	2	8 0.7	2.9	2	2	2.1	2.0		9.3	61	1.5	2.3	2.7	1.6
msrdc_dll	9	2	2 0.6	3.6	1	8	2.0	1.4		1.6	1.5	4.5	1.4	2.0	0.9
nyrshu dll	10	3	3 0.3	4.3	0	9	3.4	3.8		2.4	2.3	1.4	32.9	6.2	2.1
pacman exe	11	5.	1 0.9	4.8	1	7	3.6	3.1		1.5	2.7	2.0	6.2	15	2.2
policytool exe	12	2.	3 0.6	5 2.2	0	8	1.6	1.6		7.5	1.6	0.9	2.1	2.2	2.6
powerpnt exe	13	3.	5 0.4	2.2	1	1	3.1	1.5	4	1.2	5.8	1.4	2.6	2.2	4.6
rtinstaller32 exe	14	3.	4 0.9	4.1	2	0	3.6	2.0		1.6	2.3	2.2	2.3	2.8	1.2
safrsly dll	15	1.	9 0.9	2.2	1	1	1.2	1.6		1.1	1.1	0.7	2.0	2.0	1.0
tabulate drive data pv 16		0.	1 0.1	0.1	0	1 (	0.1	0.1	1	0.2	0.1	-	0.1	-	0.3
typeaheadfind_dl1	17	0.	9 0.6	5 1.3	0	7 0	0.4	0.3		0.4	0.5	0.7	0.1	0.7	0.5
udlaunch exe	18	2.	9 0.4	1 3.3	1	1 3	2.5	-	1	1.3	1.7	1.8	3.3	3.0	-
vsplugin_dll	19	3.	0 0.6	5 3.4	1.	0 3	2.0	2.5	1	4.0	1.8	1.2	3.2	3.0	1.6
webclnt_dll	20	3.	3 1.0	3.6	1.	1 3	2.3	1.3	1	1.8	1.8	1.5	2.2	2.8	1.0
winprint_dl1		0.	8 0.5	5 0.9	0.	4 (	0.6	0.5	1	0.4	0.5	0.5	0.4	0.6	0.6
wmplayer_exe 22		3.	1 0.4	3.1	0	9 3	2.4	2.0	1 2	21.7	3.6	1.3	3.0	2.8	2.4
xrxwiadr_dll 2		11.	5 0.8	3 12.1	2.	5 5	9.2	4.1		3.2	4.6	3.3	12.9	13.0	3.8
Family		n	13	14	15	16	1	7	18	19	20	21	22	23	Ī
a0003775_dll		1	3.5	3.4	1.9	0.1	0.	9 2	2.9	3.0	3.3	0.8	3.1	11.5	1
bthserv_dll	i	2	0.4	0.9	0.9	0.1	0.	6 (	0.4	0.6	1.0	0.5	0.4	0.8	
ccalert_dl1		3	2.2	4.1	2.2	0.1	1.	3 3	3.3	3.4	3.6	0.9	3.1	12.1	
cdfview_dl1		4	1.1	2.0	1.1	0.1	0.	7 1	1.1	1.0	1.1	0.4	0.9	2.5	
dunzip32_dll		5	3.1	3.6	1.2	0.1	0.	4 2	2.5	2.0	2.3	0.6	2.4	9.2	
hotfix_exe		6	1.5	2.0	1.6	0.1	0.	3	-	2.5	1.3	0.5	2.0	4.1	
iexplore_exe		7	41.2	1.6	1.1	0.2	0.	4 1	1.3	4.0	1.8	0.4	21.7	3.2	
mobsync_exe		8	5.8	2.3	1.1	0.1	0.	5 1	1.7	1.8	1.8	0.5	3.6	4.6	
msrdc_dl1		9	1.4	2.2	0.7	-	0.	7 1	1.8	1.2	1.5	0.5	1.3	3.3	
nvrshu_dl1		10	2.6	2.3	2.0	0.1	0.	1 3	3.3	3.2	2.2	0.4	3.0	12.9	
pacman_exe		11	2.2	2.8	2.0	-	0.	7 3	3.0	3.0	2.8	0.6	2.8	13.0	
policytool_exe		12	4.6	1.2	1.0	0.3	0.	5	-	1.6	1.0	0.6	2.4	3.8	
powerpnt_exe		13	76.0	1.5	1.0	0.2	0.	2 1	1.5	2.8	1.7	0.3	12.6	8.2	
rtinstaller32_exe		14	1.5	13.4	1.1	0.1	0.	4 3	3.1	1.9	2.0	0.6	1.7	6.3	
safrslv_dll		15	1.0	1.1	3.3	0.1	0.	8	-	1.4	1.2	0.6	1.1	2.6	
tabulate_drive_data_	_ру	16	0.2	0.1	0.1	2.8	0.	1	-	0.2	0.2	-	0.1	0.3	
typeaheadfind_dll		17	0.2	0.4	0.8	0.1	2.	3 (	J.2	0.5	0.8	0.4	0.2	0.6	
udlaunch_exe		18	1.5	3.1	1.4	-	0.	2		2.1	0.9	0.5	2.2	3.6	
vsplugin_dl1		19	2.8	1.9	1.4	0.2	0.	5 2	2.1	5.2	1./	0.5	2.7	3.5	
webcint_dll		20	1.7	2.0	1.2	0.2	0.	8 0	1.9	1.7	3.8	0.7	1.5	5.0	
winprint_dii		21	12.6	0.0	0.0	-	0.	4 0	1.5	0.5	0.7	1.1	0.4	0.6	
wmplayer_exe		22	12.6	6.2	1.1	0.1	0.	2 2	2.2	2.7	1.5	0.4	9.1	3.7	
xrxwiadr_dll		25	8.2	0.5	2.0	0.3	0.	0 .	0.0	3.5	5.0	0.0	3./	15.9	1

#### **False Positives**

- Data regions
- High entropy

Texture Vector Similarity Version 1.0.0 Scale: 1:1 Step: 1 Section size: 500

SD Wt: 0.500, Mean Wt: 0.500, Mode Wt: 0.000, Mode Count Wt: 0.500, Entropy Wt: 0.500, Threshold: 5.000, Buckets: 434 Compensated statistics: SD: 11.7100, Mean: 3.8247, Max: 46, Sum: 1113, Max/Sum: 0.0413

File 1: /smallwork/bdallen/executable\_files\_500/a0003775\_dll/

Into 1.9 Junumous Jobal Constants in Construction of the second secon



### **Time-based Analysis**

- Similarity based on timestamp
  - Version inference
  - Virus injection

Similarity to node 410: Stay in group, Standard Deviation similarity > 1.000

Node 410: /smallwork/bdallen/executable\_files\_500/nvrshu\_dll/MY01-023\_WINDOWS\_system32\_nvrshu.dll.tmp Group: 'nvrshu\_dll', Size: 262144, Modtime: 2011-04-07 17:16:28, MD5: 920AE502475DD6E93FD47B58956B190C



## **Texture-Vector Similarity Distribution**

- Tools
  - calc\_tv.py calculates texture vectors
  - tv.py calculates and plots similarities
    between two files
  - tv\_browser.py graphs similarities by timestamp
- Dataset
  - The Similarity Graph data from the 1,134 files
- https://github.com/NPS-DEEP/tv\_sim





Similarity to node 410: Stay in group, Standard Deviation similarity > 1.00

de 410: /smailwork/bdailen/executable\_files\_500/nvrshu\_dll/MY01-023\_WINDOWS\_system32\_nvrshu\_dll.tmp pup: 'nvrshu\_dll', Size: 262144, Modtime: 2011-04-07 17:16:28, MD5: 920AE502475DD6E93FD47858956B190C



### Future Work

- Classify file types by their texture
- Evaluate similarity using texture trends
- Manage false-positives

