30th ANNUAL FIRST CONFERENCE
KUALA LUMPUR
June 24-29, 2018
30 YEARS OF INCIDENT HANDLING

Multi-dimensional malware similarity will let you catch up with malware developers

Kunihiko Yoshimura, Koji Yamada, Toshitaka Satomi, and Ryusuke Masuoka

Fujitsu System Integration Laboratories

FIRST Improving Security Together

FUJITSU
shaping tomorrow with you

Outline

- **Introduction**
  - Difficulty with Malware Analysis Operations
  - Similarity Tool to Rescue
  - A Single Similarity Tool Is No Match
- **Road to the Proposed Solution**
  - Initial Struggles
  - Bunch of Similarity Tools
  - Three Dimensions for Human Analysts
  - Sample Similarity Scoring System (S4)
- **S4 vs. Malware Families**
  - Match Rules
  - S4 Won All the Matches!
  - Exhibition Match: Olympic Destroyer
- **Conclusion**
  - Future Plan
  - Take Home Message

Outline

- **Introduction**
  - Difficulty with Malware Analysis Operations
  - Similarity Tool to Rescue
  - A Single Similarity Tool Is No Match
- **Road to the Proposed Solution**
  - Initial Struggles
  - Bunch of Similarity Tools
  - Three Dimensions for Human Analysts
  - Sample Similarity Scoring System (S4)
- **S4 vs. Malware Families**
  - Match Rules
  - S4 Won All the Matches!
  - Exhibition Match: Olympic Destroyer
- **Conclusion**
  - Future Plan
  - Take Home Message

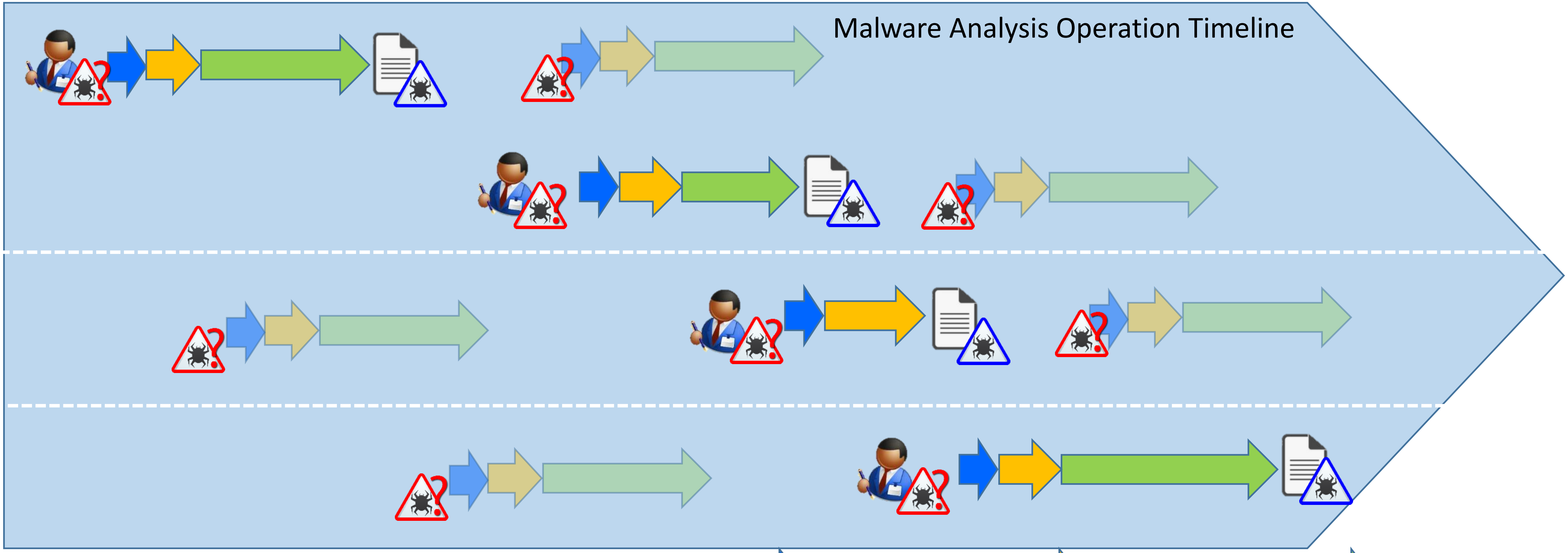
Difficulty with Malware Analysis Operations

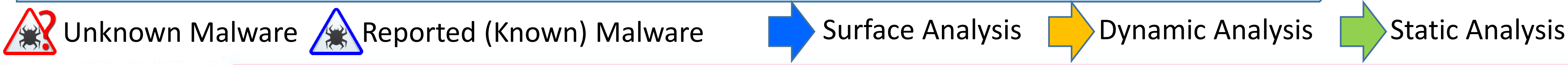
- Attackers continue to develop new pieces of malware at an alarming rate
  - “Conficker” ”CryptoWall" “Badrabbit” “HackerDefender” “Hiddad” “HummingBad” “Necurs” “Nivdort” “Sality” “Triada” “Zeus” “Locky” “CoinHive” “Ramnit” “Fireball” “Pushdo” …

- Analysts cannot keep up with the pace

Similarity Tool to Rescue

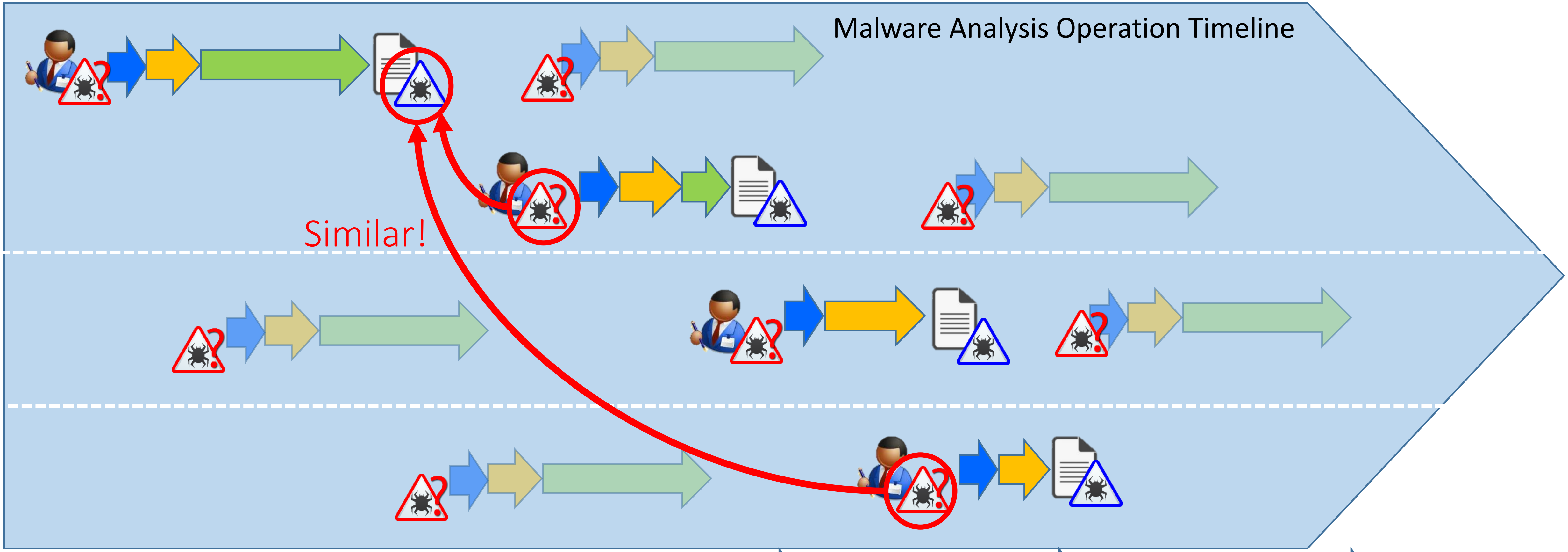
Before…





Similarity Tool to Rescue, but…

After



Unknown Malware | Reported (Known) Malware | Surface Analysis | Dynamic Analysis | Static Analysis

A Single Similarity Tool Is No Match

Similarity Tool	Evasion Technique
• Fuzzy hashing (ex. ssdeep, SDHASH)	<= XOR cipher
• Static Analysis (ex. Section Matching, BinDiff)	<= Packers
• Dynamic Analysis (ex. Techniques using Created Processes, APIs/DLL Calls)	<= Anti-Sandbox

→ A Single Similarity Tool Can Be Easily Evaded

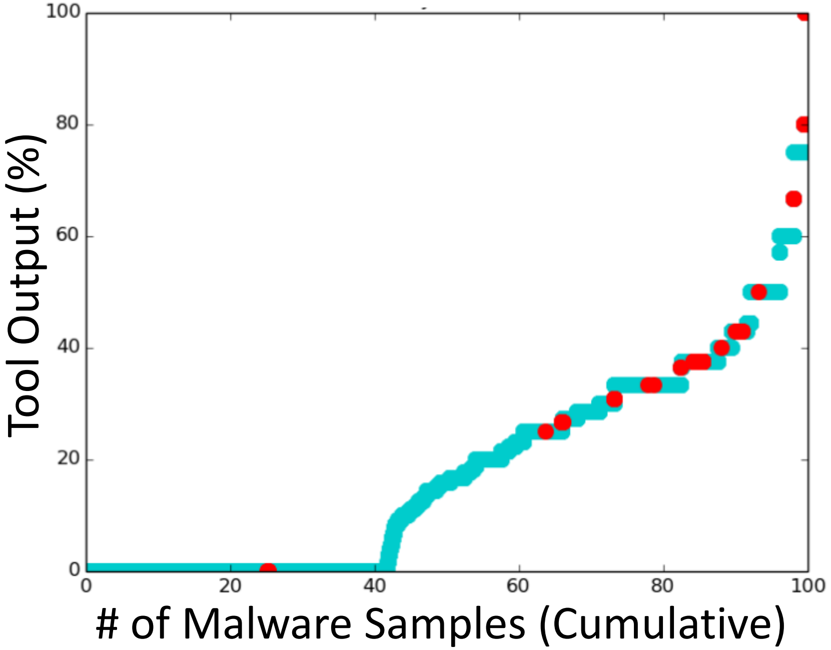
Outline

- Introduction
  - Difficulty with Malware Analysis Operations
  - Similarity Tool to Rescue
  - A Single Similarity Tool Is No Match
- **Road to the Proposed Solution**
  - Initial Struggles
  - Bunch of Similarity Tools
  - Three Dimensions for Human Analysts
  - Sample Similarity Scoring System (S4)
- S4 vs. Malware Families
  - Match Rules
  - S4 Won All the Matches!
  - Exhibition Match: Olympic Destroyer
- Conclusion
  - Future Plan
  - Take Home Message

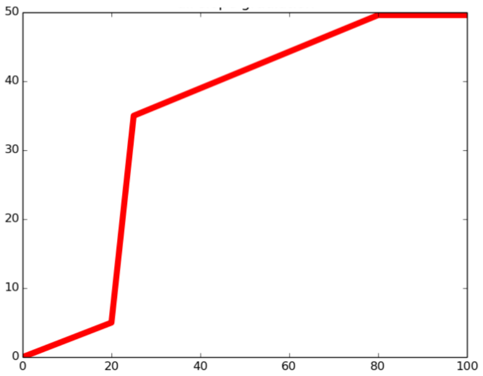
Score Transformation

- **Piecewise linear transformation (PLT)** to enhance malware family separation

DLL Jaccard



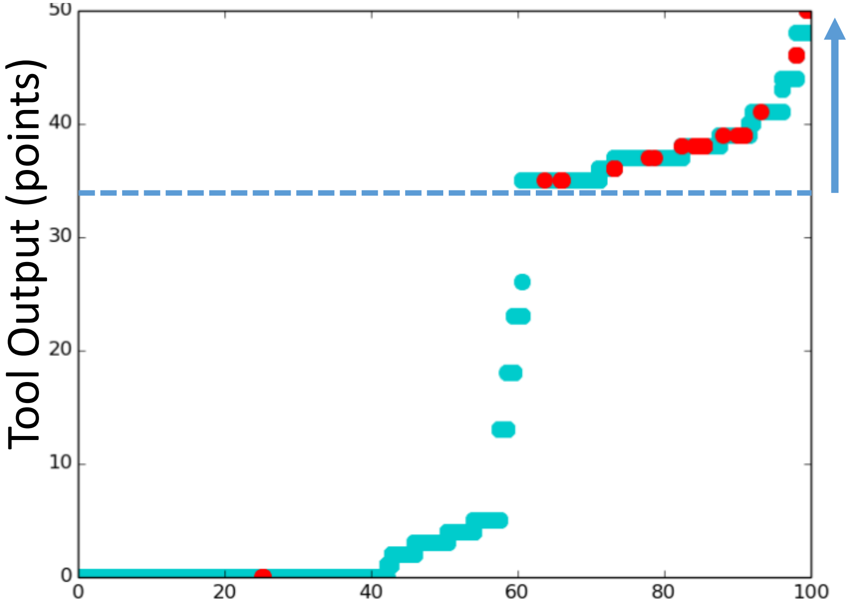
×



Piecewise Linear Transformation

⇒

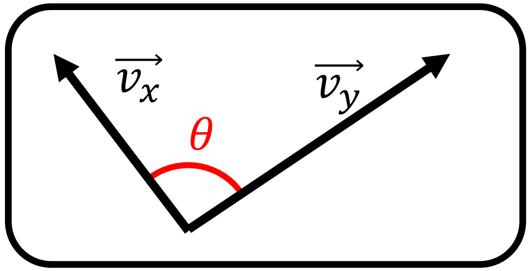
DLL Jaccard (Transformed)



Original Metrics by Cosine Similarity Tools

Ex. Call API Cosine

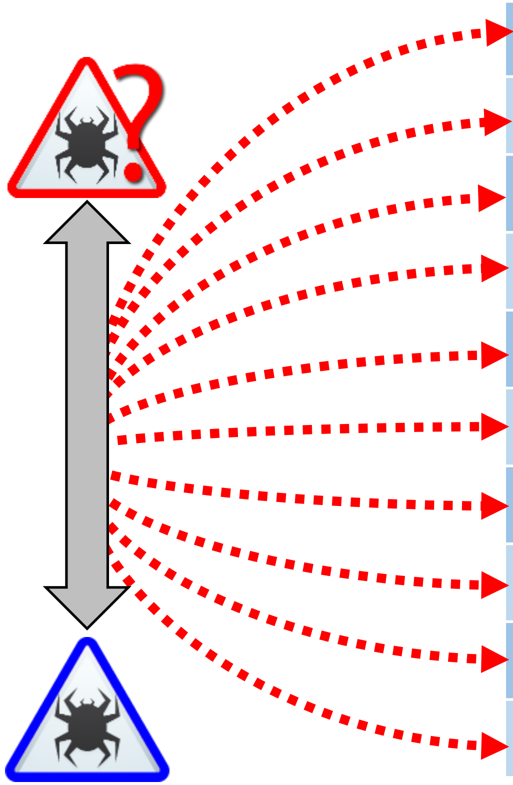
API	GetProcAddress	Connect	Hook	LoadLibrary	…	Feature Vector
Malware X	44	3	1	31	…	$\overrightarrow{v_x}$
Malware Y	22	11	1	30	…	$\overrightarrow{v_y}$



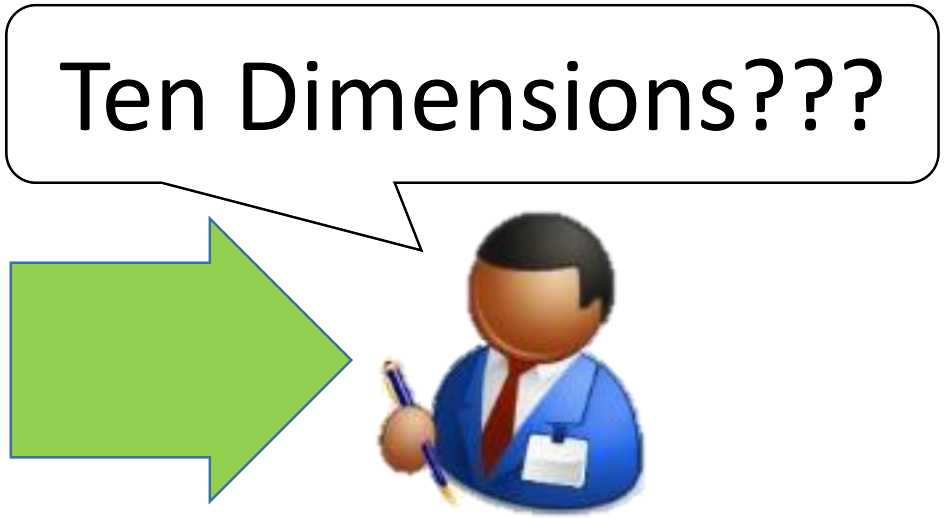
$$\cos\theta = \frac{\overrightarrow{v_x}\cdot\overrightarrow{v_y}}{|\overrightarrow{v_x}||\overrightarrow{v_y}|} = 0.8866$$

Similarity Metric

Bunch of Similarity Tools

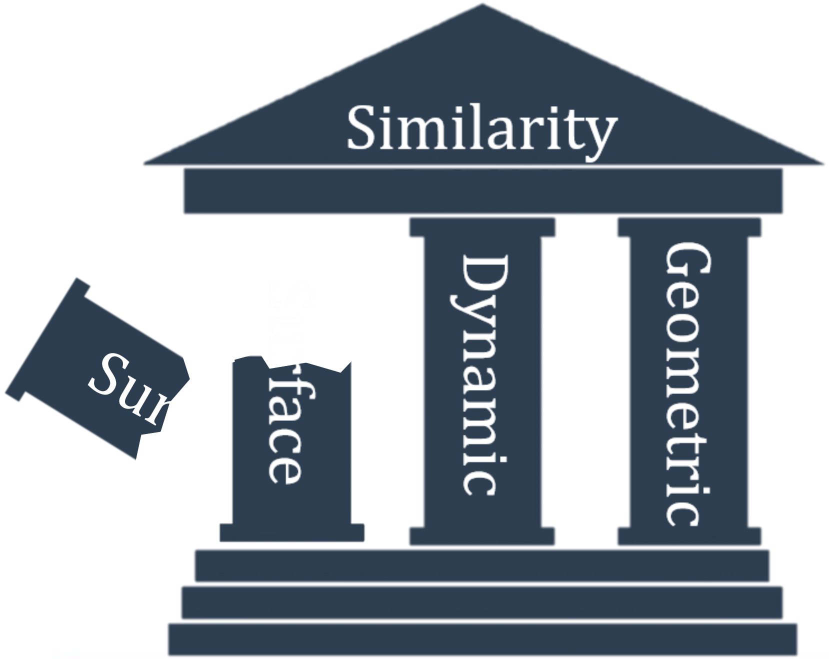


	Transformed Score
ssdeep	0/10
SDHASH	0/10
entropy	45/135
Section Match	0/10
Import DLL	42/50
Call API	31/50
Call DLL	39/50
Process Tree	30/30
API Cosine	42/50
API n-gram Cosine	31/50



Three Dimensions for Human Analysts

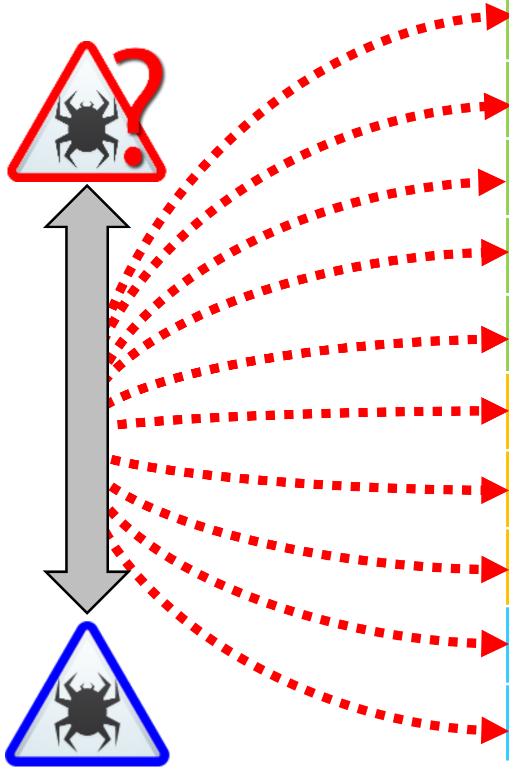
- Higher the Dimensions, Harder to Evade
- Highest # of Dimensions
  for Human to Handle -> 3!



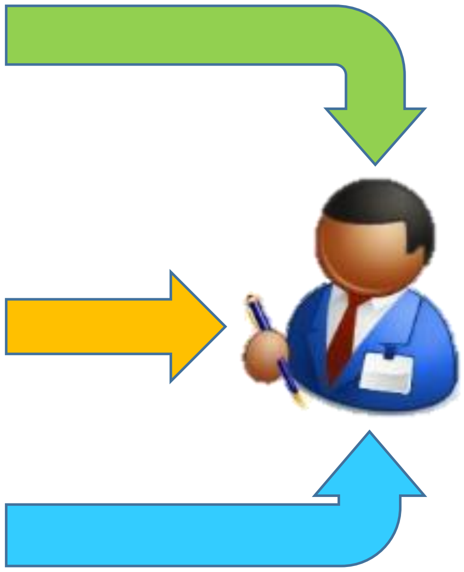
Two Dimensions

Three Dimensions

Similarity Metrics into Three Dimensions

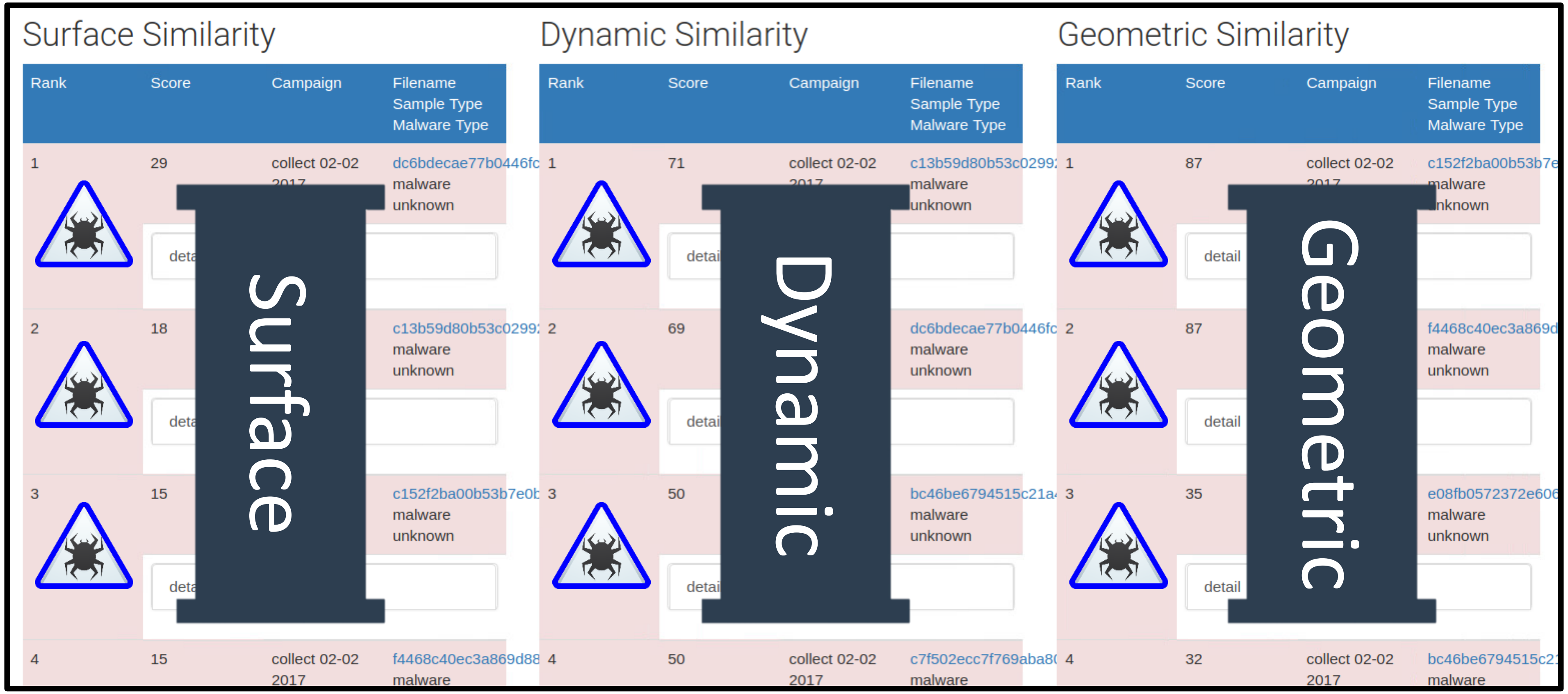


	Transformed Score	Integrated Score
ssdeep	0/10	**Surface Score: 40** (87/215)
SDHASH	0/10	
entropy	45/135	
Section Match	0/10	
Import DLL	42/50	
Call API	31/50	**Dynamic Score: 77** (100/130)
Call DLL	39/50	
Process Tree	30/30	
API Cosine	42/50	**Geometric Score: 73** (73/100)
API n-gram Cosine	31/50	



Sample Similarity Scoring System (S4)

For a given unknown malware → Similar malware ranking in each dimension



Surface Similarity

Rank	Score	Campaign	Filename Sample Type Malware Type
1	29 deta	collect 02-02 2017	dc6bdecae77b0446fc malware unknown
2	18 deta		c13b59d80b53c0299: malware unknown
3	15 deta		c152f2ba00b53b7e0b malware unknown
4	15	collect 02-02 2017	f4468c40ec3a869d88 malware

Dynamic Similarity

Rank	Score	Campaign	Filename Sample Type Malware Type
1	71 detai	collect 02-02 2017	c13b59d80b53c0299: malware unknown
2	69 detai		dc6bdecae77b0446fc malware unknown
3	50 detai		bc46be6794515c21a malware unknown
4	50	collect 02-02 2017	c7f502ecc7f769aba8(malware

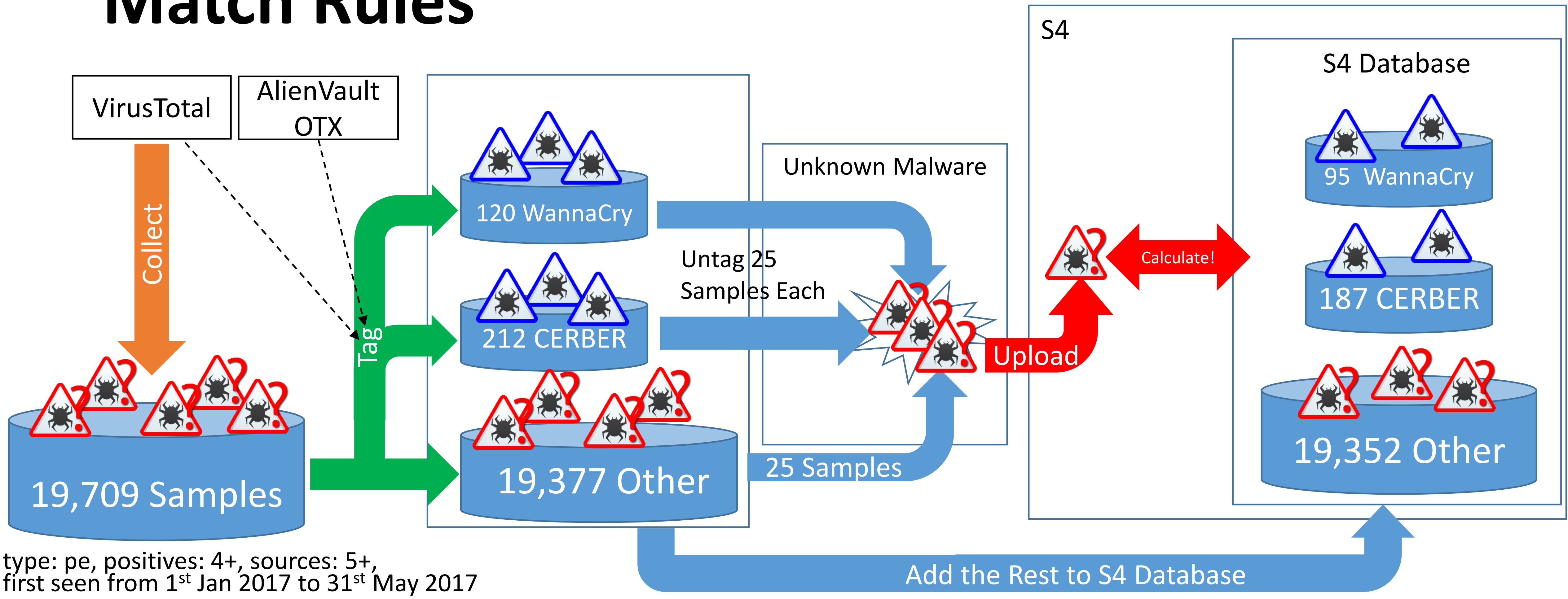
Geometric Similarity

Rank	Score	Campaign	Filename Sample Type Malware Type
1	87 detail	collect 02-02 2017	c152f2ba00b53b7e alware nknown
2	87 detail		f4468c40ec3a869d malware unknown
3	35 detail		e08fb0572372e606 malware unknown
4	32	collect 02-02 2017	bc46be6794515c2 malware

Outline

- Introduction
  - Difficulty with Malware Analysis Operations
  - Similarity Tool to Rescue
  - A Single Similarity Tool Is No Match
- Road to the Proposed Solution
  - Initial Struggles
  - Bunch of Similarity Tools
  - Three Dimensions for Human Analysts
  - Sample Similarity Scoring System (S4)
- **S4 vs. Malware Families**
  - Match Rules
  - S4 Won All the Matches!
  - Exhibition Match: Olympic Destroyer
- Conclusion
  - Future Plan
  - Take Home Message

Match Rules



type: pe, positives: 4+, sources: 5+,
first seen from 1st Jan 2017 to 31st May 2017

S4 Wins When It Shows Ones from the Same Family ≧ 50% in Those with Scores ≧ 90

Similar Malware Ranking for "unknown01.exe"

Rank	Surface Similarity	Dynamic Similarity	Geometric Similarity
1	g568.x86.ca.1000.exe (Other): 79	read.php (CERBER): 98	<sha256>.bin (CERBER): 99
2	04958pg.jpeg.exe (Other): 79	read.exe (CERBER): 98	<MD5>.virus (CERBER): 99
3	SETUP-VW.EXE (Other): 78	<sha256>.bin (CERBER): 95	2.exe (CERBER): 99
4	CmbShowHis.EXE (Other): 77	rigamfu.exe (CERBER): 92	voperseanx.exe (CERBER): 99
5	AutoCAD_Setup.exe (Other): 76	DW20.Exe (CERBER): 90	<MD5> (CERBER): 99
6	tpad109.exe (Other): 76	user.phpf1.gif.exe (CERBER): 90	cerber.exe (CERBER): 99
7	your.exe (Other): 76	<MD5>.virus (CERBER): 90	cerber2.exe (CERBER): 99
8	your.exe (Other): 76	zzz.exe (CERBER): 90	dsconfig.exe (CERBER): 99
9	M3Apnda2.exe (Other): 76	1.EXE (CERBER): 90	exe1.exe (CERBER): 99
10	f5aauicn.exe (Other): 75	003.exe (CERBER): 89	exe1.exe (CERBER): 99

≧ 90

< 90

All (100%) of 19 Samples with Scores ≧ 90 Are CERBER Family

-> S4 Wins the Match!

S4 Won All the Matches!

Malware	# of Samples	# of S4 Wins	Winning Rate
CERBER	25	25	100%
WannaCry	25	25	100%
Other	25	25	100%

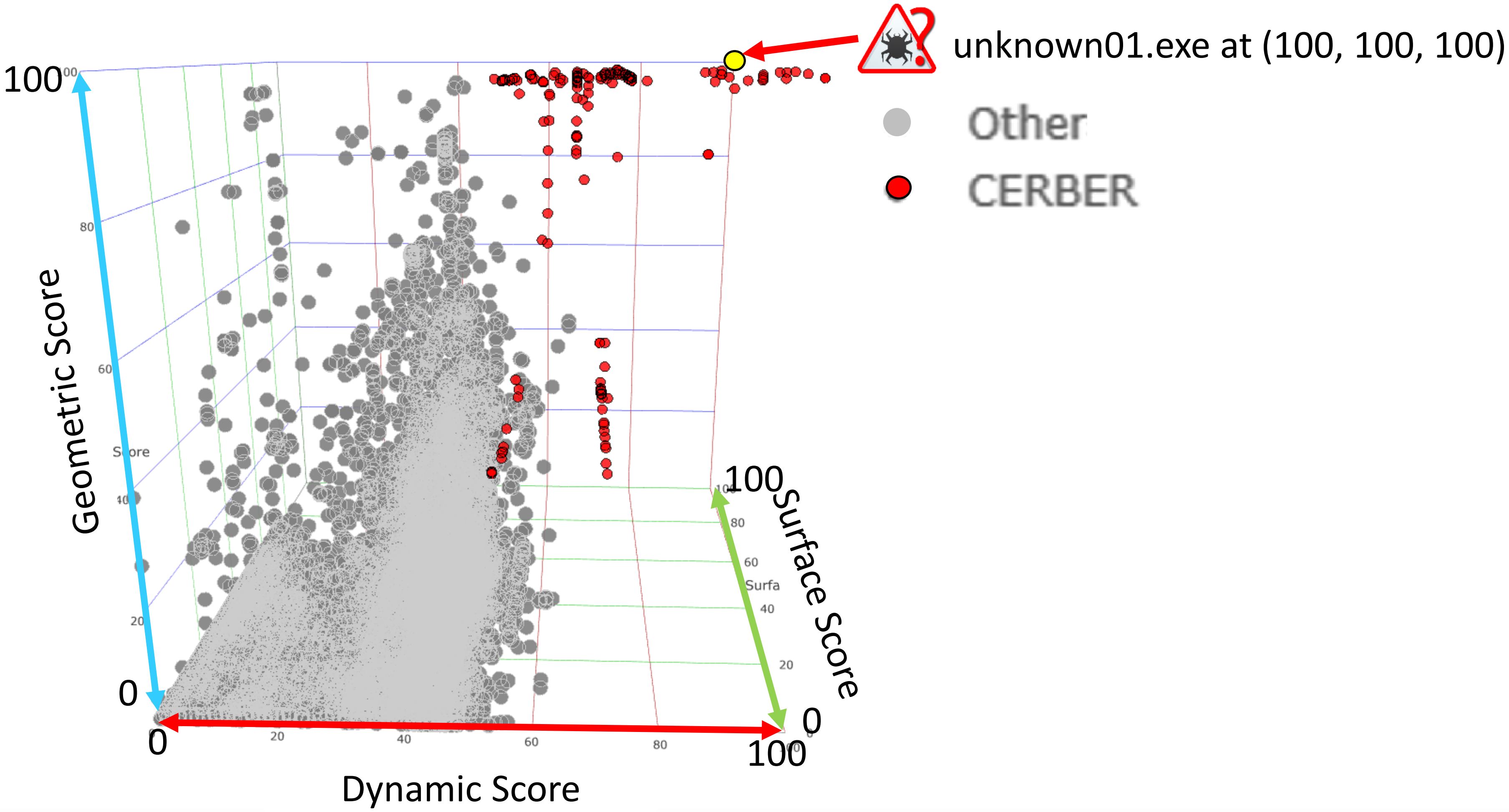
Similar Malware Ranking for “unknown01.exe”

Rank	Surface Similarity	Dynamic Similarity	Geometric Similarity
1	g568.x86.ca.1000.exe (Other): 79	read.php (CERBER): 98	<sha256>.bin (CERBER): 99
2	04958pg.jpeg.exe (Other): 79	read.exe (CERBER): 98	<MD5>.virus (CERBER): 99
3	SETUP-VW.EXE (Other): 78	<sha256>.bin (CERBER): 95	2.exe (CERBER): 99
4	CmbShowHis.EXE (Other): 77	rigamfu.exe (CERBER): 92	voperseanx.exe (CERBER): 99
5	AutoCAD_Setup.exe (Other): 76	DW20.Exe (CERBER): 90	<MD5> (CERBER): 99
6	tpad109.exe (Other): 76	user.phpf1.gif.exe (CERBER): 90	cerber.exe (CERBER): 99
7	your.exe (Other): 76	<MD5>.virus (CERBER): 90	cerber2.exe (CERBER): 99
8	your.exe (Other): 76	zzz.exe (CERBER): 90	dsconfig.exe (CERBER): 99
9	M3Apnda2.exe (Other): 76	1.EXE (CERBER): 90	exe1.exe (CERBER): 99
10	f5aauicn.exe (Other): 75	003.exe (CERBER): 89	exe1.exe (CERBER): 99

≧ 90

< 90

Behind the Scene - 3D Visualization (CERBER)



Similar Malware Ranking for “unknown40.exe”

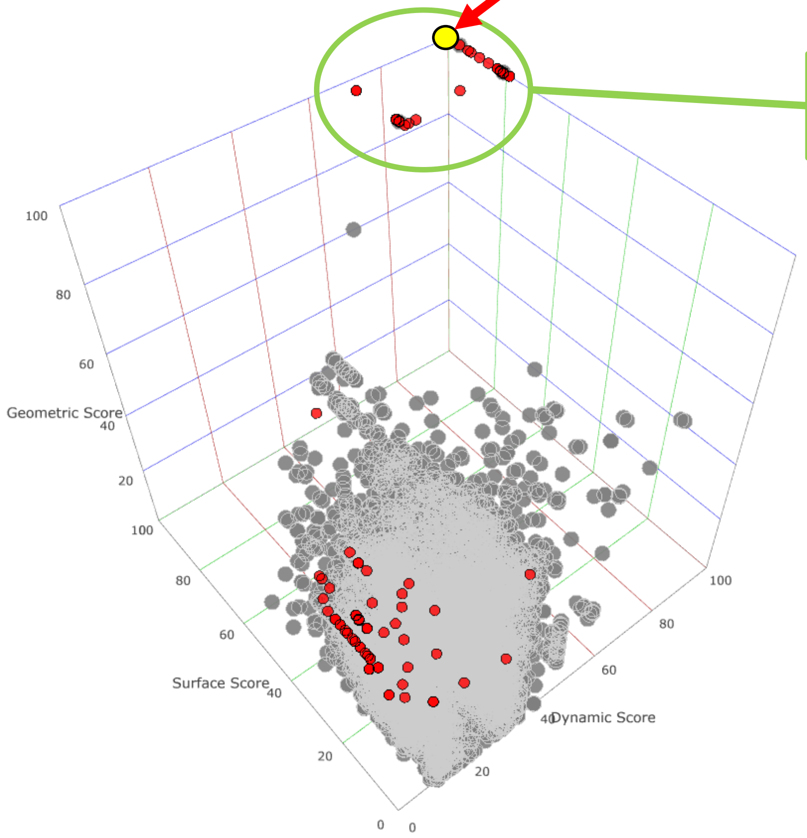
Rank	Surface Score	Dynamic Score	Geometric Score
1	mssecsvc.exe (WannaCry): 98	mssecsvc.exe (WannaCry): 100	mssecsvc.exe (WannaCry): 99
2	mssecsvc.exe (WannaCry): 95	<MD5hash>.virus (WannaCry): 100	lhdfrgui.exe (WannaCry): 99
3	<MD5hash>.virus (WannaCry): 94	lhdfrgui.exe (WannaCry): 100	lhdfrgui.exe (WannaCry): 99
4	lhdfrgui.exe (WannaCry): 94	lhdfrgui.exe (WannaCry): 100	lhdfrgui.exe (WannaCry): 99
5	mssecsvc.exe (WannaCry): 94	mssecsvc.exe (WannaCry): 100	<MD5hash>.virus (WannaCry): 99
6	mssecsvc.exe (WannaCry): 93	<MD5hash>.virus (WannaCry): 100	mssecsvc.exe (WannaCry): 99
7	mssecsvc.exe (WannaCry): 92	36318392.exe (WannaCry): 100	lhdfrgui.exe (WannaCry): 99
8	lhdfrgui.exe (WannaCry): 91	mssecsvc.exe (WannaCry): 100	<MD5hash>.virus (WannaCry): 99
9	lhdfrgui.exe (WannaCry): 91	mssecsvc.exe (WannaCry): 100	mssecsvc.exe (WannaCry): 99
10	36318392.exe (WannaCry): 89	<MD5hash>.virus (WannaCry): 100	<MD5hash>.virus (WannaCry): 99

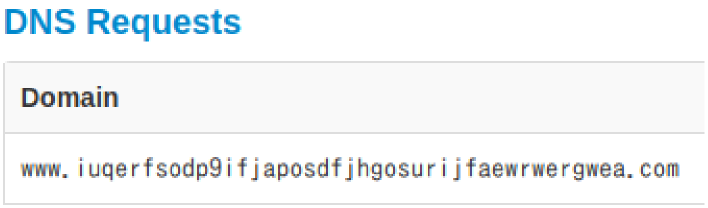
≧ 90

▨ < 90

Behind the Scene - Kill Switch of WannaCry

unknown40.exe at (100, 100, 100)





DNS Requests

Domain
www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com

Similar Malware Ranking for “unknown14.exe”

Rank	Surface Score	Dynamic Score	Geometric Score
1	P_SK001.exe (Other) : 93	AIRBNB Brute.exe (Other) : 100	AIRBNB Brute.exe (Other) : 99
2	P_SK003.exe (Other) : 92	Skype Resolver.exe (Other) : 100	Skype Resolver.exe (Other) : 99
3	Red_crypter.exe (Other) : 90	HackerClean.exe (Other) : 100	HackerClean.exe (Other) : 99
4	P_SK002.exe (Other) : 90	stm1.exe (Other) : 100	stm1.exe (Other) : 99
5	P_SK005.exe (Other) : 88	RebornBuddy.exe (Other) : 100	RebornBuddy.exe (Other) : 99
6	ReptileUI.exe (Other) : 87	updater.exe (Other) : 100	updater.exe (Other) : 99
7	HmpvInst.exe (Other) : 86	ProxyAlts Loader.exe (Other) : 100	ProxyAlts Loader.exe (Other) : 99
8	Stealth.exe (Other) : 86	PML_Alert.exe (Other) : 100	PML_Alert.exe (Other) : 99
9	google chrom.exe (Other) : 85	conhost.exe (Other) : 100	MmiStart.exe (Other) : 99
10	Application1.exe(Other) : 85	MmiStart.exe (Other) : 100	GITS-DE.exe (Other) : 99

≧ 90   < 90

Exhibition Match: S4 vs. Olympic Destroyer

“unknown2018_01.exe” (Found Feb 2018) against Feb 2018 Malware Set

Rank	Surface Score	Dynamic Score	Geometric Score
1	_bjv.exe (Olympic Destroyer) : 99	_bjv.exe (Olympic Destroyer) : 93	_bjv.exe (Olympic Destroyer) : 99
2	_bdm.exe (Olympic Destroyer) : 99	_bdm.exe (Olympic Destroyer) : 93	_bdm.exe (Olympic Destroyer) : 99
3	_rnk.exe (Olympic Destroyer) : 99	_rnk.exe (Olympic Destroyer) : 93	_rnk.exe (Olympic Destroyer) : 99
4	<MD5> (Olympic Destroyer) : 99	<MD5> (Olympic Destroyer) : 93	<MD5> (Olympic Destroyer) : 99
5	_jea.exe (Olympic Destroyer) : 92	_jea.exe (Olympic Destroyer) : 74	zeuspanda (Panda Banker) : 70
6	_ljy.exe (Olympic Destroyer) : 92	_ljy.exe (Olympic Destroyer) : 74	<MD5> (Other) : 70
7	_mpw.exe (Olympic Destroyer): 92	_nfc.exe (Olympic Destroyer) : 74	CFE_Factura.exe (Other) : 69
8	_qih.exe (Olympic Destroyer) : 92	_nka.exe (Olympic Destroyer) : 74	executable.1088.exe (Other) : 68
9	_nfc.exe (Olympic Destroyer) : 91	_mpw.exe (Olympic Destroyer) : 74	<MD5> (Olympic Destroyer) : 65
10	_nka.exe (Olympic Destroyer) : 91	_wun.exe (Olympic Destroyer) : 74	_nfc.exe (Olympic Destroyer) : 65

≧ 90   < 90

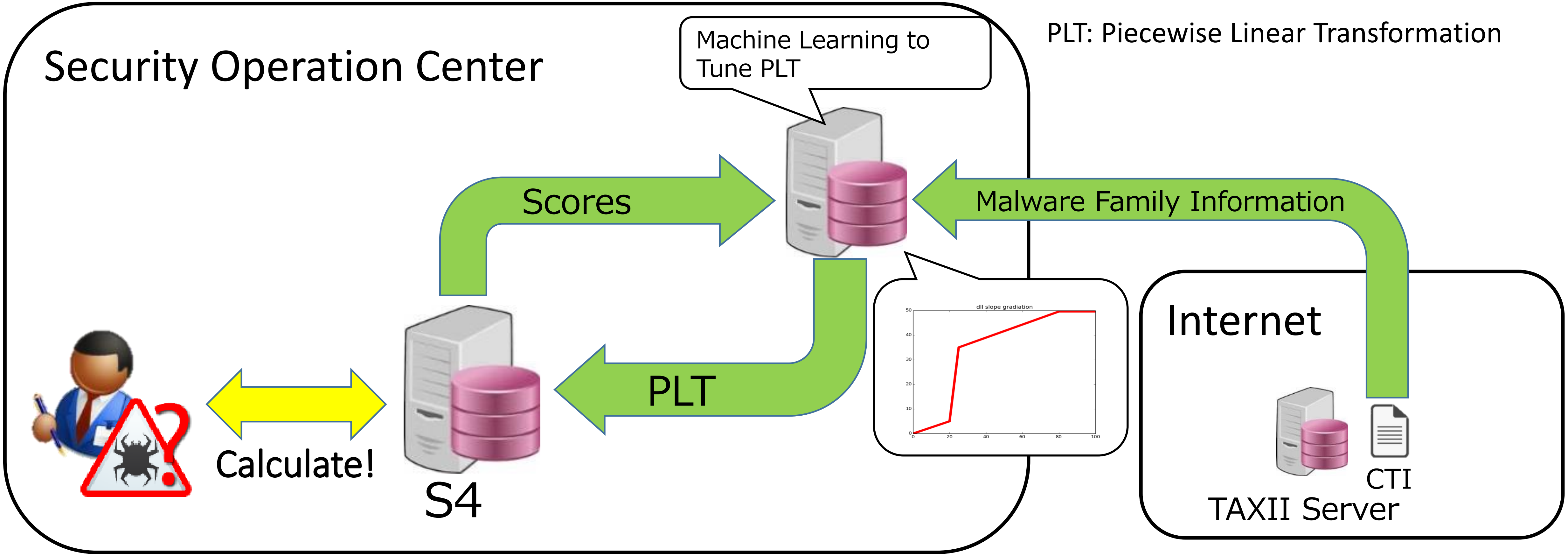
Outline

- **Introduction**
    - Difficulty with Malware Analysis Operations
    - Similarity Tool to Rescue
    - A Single Similarity Tool Is No Match
- **Road to the Proposed Solution**
    - Initial Struggles
    - Bunch of Similarity Tools
    - Three Dimensions for Human Analysts
    - Sample Similarity Scoring System (S4)
- **S4 vs. Malware Families**
    - Match Rules
    - S4 Won All the Matches!
    - Exhibition Match: Olympic Destroyer
- **Conclusion**
    - Future Plan
    - Take Home Message

Future Plan:

Automatic Score Transformation Tuning

PLT: Piecewise Linear Transformation



Take Home Message

There is □ with Malware Analysis Operations, a □ Tool comes to Rescue, But … A □ Similarity Tool Is No Match for evasion techniques.

Through Initial Struggles, we developed □ □, which put metrics from a □ of Similarity Tools Into □ Dimensions for Human Analysts for their easy understanding.

S4 □ All the Matches against two malware families.

→ Multi-dimensional Malware Similarity Will Let You Catch Up with Malware Developers

Q & A