



2009
GENEVA



ANUBIS

ANalyzing **U**nknown **B**Inarie**S**

The automatic Way

Thomas Mandl, Ulrich Bayer, Florian Nentwich

25.09.2009, v1.0.02, EN

Virus Bulletin Conference 2009, Geneva

People behind ANUBIS – Who are we?

Ulrich Bayer

- Currently PhD student at Vienna University of Technology
- Main developer and architect of ANUBIS

Florian Nentwich

- Senior malware analyst at Ikarus labs and maintainer of commercial ANUBIS version

Thomas Mandl

- Former CTO of Ikarus, now CEO of his own information security consulting company in Austria
- Still contributing to the ANUBIS project

ANUBIS' Academic Research Members

Engin Kirda

- Assistant professor at EURECOM Communication Systems
- Former assistant professor at Vienna University of Technology
- <http://www.eurecom.fr/people/kirda.en.htm>

Christopher Kruegel

- Assistant professor at UCSB, Dept. of Computer Science
- Former assistant professor at Vienna University of Technology
- <http://www.cs.ucsb.edu/~chris/>
- Development of Wepawet Tools

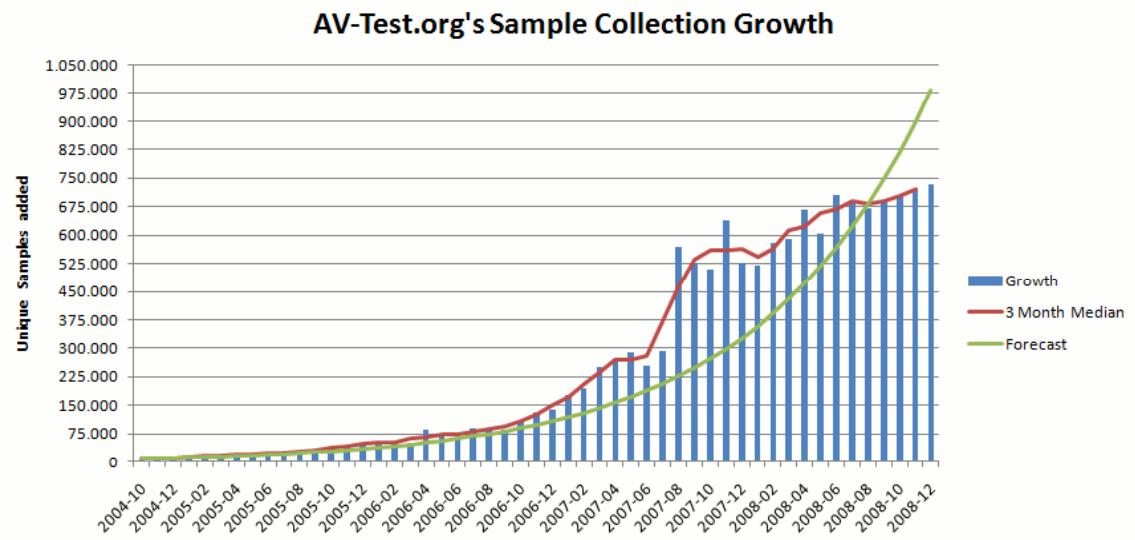
See also

- <http://www.iseclab.org/people.html>

Automated Malware Analysis: Why?

Too many new malware samples per day

- ~25k - 35k samples per day (unique MD5) (peak up to 50k)
- Increasing number of malware uses runtime packers/code obfuscation methods to trick pattern matching AV
- Increasing FP rate, nobody can handle this load manually!
- Almost no in-house incident response process/RE due to its complexity
(at least in Austria)
- Among others, this was our primary motivation to create ANUBIS!



A traditional Analysis Approach

- 
- ~35k samples/day
 - Manual analysis takes up to several days

- Limited human expert resources
- Experts should concentrate on novel malware

- Response time for signature creation is crucial
- How can we speed up this process?

What is ANUBIS?

Framework of several tools for **dynamic** code analysis

- We run a binary in an emulated PC environment (WinXP/SP3)
- We monitor its actions (SysCalls, Windows API functions, ...)
- We generate a detailed report of the sample's behavior
- Fully automatically within 4 min. (**no human interaction**)
- Based on an ANUBIS report, a human expert can decide whether to manually analyze a sample in depth or not.

Benefits of dynamic code analysis with ANUBIS

- Scalable approach, unaffected by runtime packers, code obfuscation or anti-debug mechanisms of modern malware
- Can handle basic user interactions if required during analysis

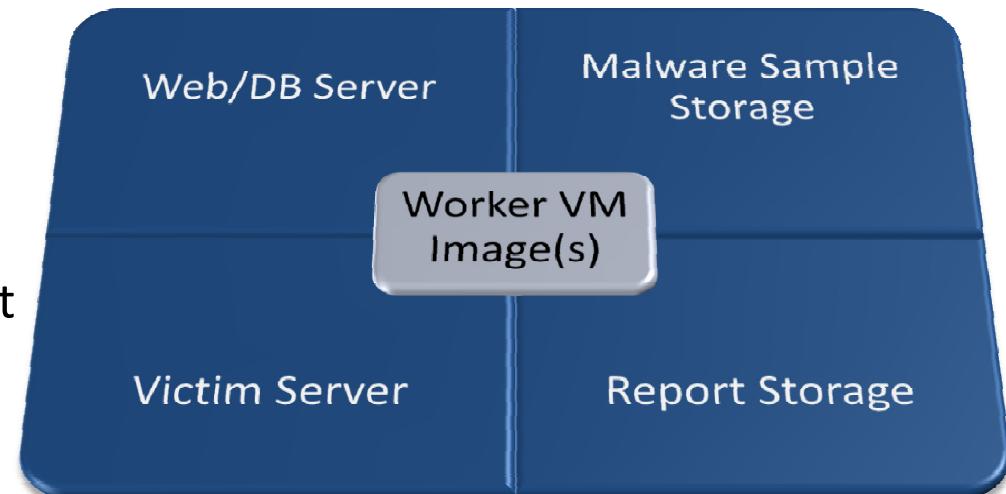
Community version heavily used by AV and AV researchers

- <https://anubis.iseclab.org> (public) with limited features

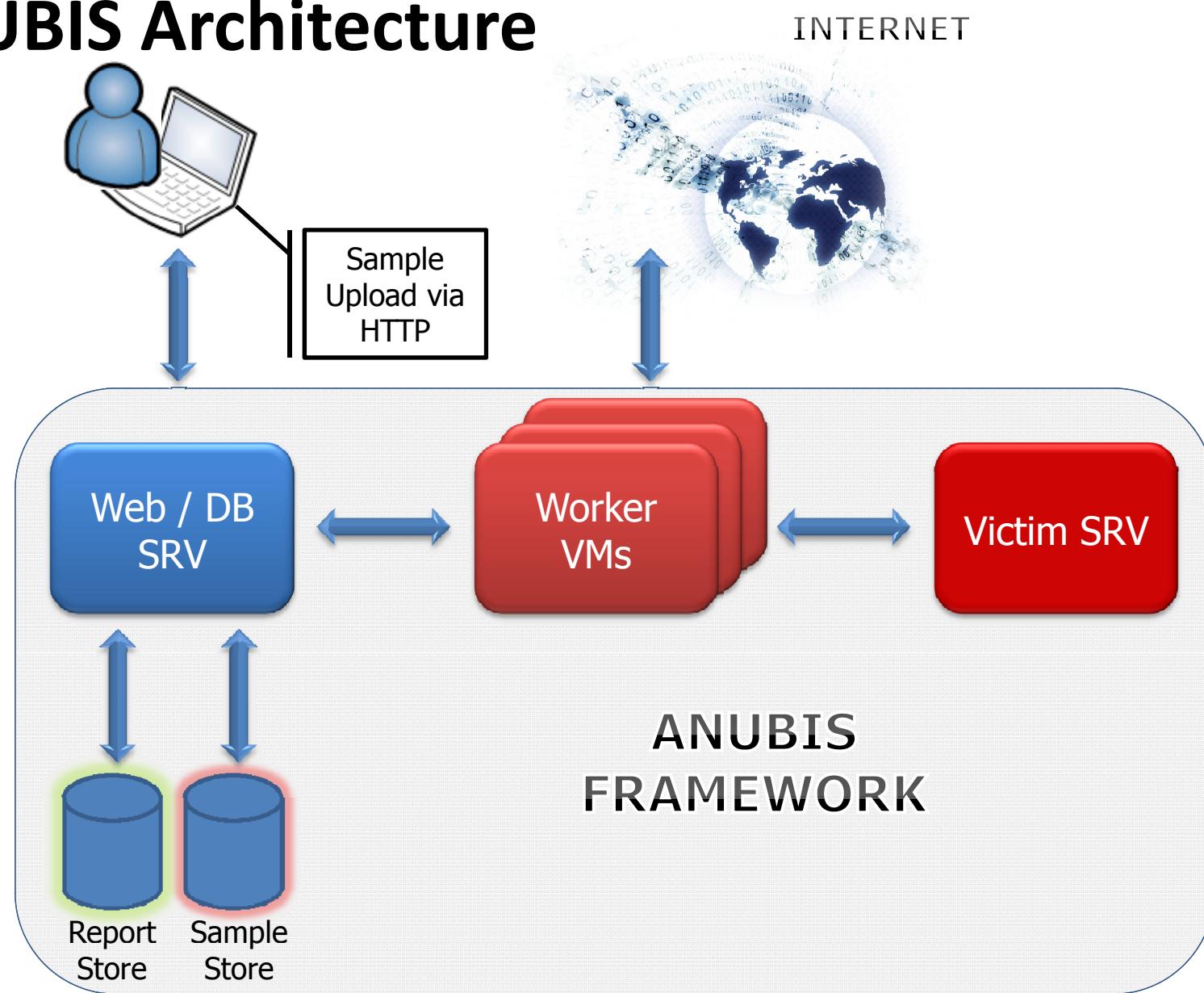
Architecture and Capabilities

ANUBIS has 5 primary building blocks

- Web/DB server/HTTP(s) frontend (upload/admin)
 - DB stores reports and references to samples (XML)
 - Enables us to generate **lots of statistics!**
- Malware sample storage
 - Archives uploaded and already analyzed samples
- Report storage
 - Archives report/result files (traffic dumps, downloaded files...)
 - Comprehensive Archive + **2nd stage malware!**
- Victim server
 - Acts as **local honey pot** for certain services and **keeps malicious traffic local!**
- Multiple Worker (VM)
 - Snapshot technology! Revert to known state in a second!



ANUBIS Architecture



Advanced Features of ANUBIS

Records and analyzes sample's network traffic

- HTTP, FTP, SMTP, IRC, ... are available as PCAP file

Storage of analysis reports in relational DB

- Servers contacted, files created, modified, deleted, RegKeys manipulated, and short threat summary

Several report formats

- XML, HTML, MHT, PDF, TXT
- Integrates also static analysis with AV scanner/PE scan



URL analysis (early development stage)

ANUBIS was designed to support human experts

- Gives quick overview of a sample's behavior within minutes
- What makes ANUBIS different from other sandbox solutions?



Anubis - Analysis Report



Analysis Report for nepenthes- 65c242c013045c678974e3be0796188d-index.html

[Comment on this report](#)

Summary:

Description	Risk
Creates files in the Windows system directory: Malware often keeps copies of itself in the Windows directory to stay undetected by users.	Orange
Performs Address Scan: The executable scans a range of IP Addresses. In most cases these scans identify more potential vulnerable targets.	Red
Performs File Modification and Destruction: The executable modifies and destructs files which are not temporary.	Red
Spawns Processes: The executable produces processes during the execution.	Yellow
Performs Registry Activities: The executable reads and modifies registry values. It may also create and monitor registry keys.	Yellow

Table of Contents

▼ expand all	collapse all ▲
General information	
nepenthes-65c242c013045c678974e3be0796188d-index.html	
urdvxc.exe	
urdvxc.exe	
services.exe	
urdvxc.exe	
urdvxc.exe	

Report – Static findings

SHA-1:	b616dcf0c05e539b317edd9d279a267a6fad01e
File Size:	131584 Bytes
Command Line:	"C:\nepenthes-65c242c013045c678974e3be0796188d-index.html"
Process-status at analysis end:	dead
Exit Code:	0

+ Load-time DLLs

+ Run-time DLLs

- SigBuster Output

Allapple_Polymeric_Packer vna SN: 1647

- Ikarus Virus Scanner

Net-Worm.Win32.Allapple.b (Sig-Id:158175)

Report – Windows Services

3.c) urdvxc.exe - Windows Service Activities

- Services Created:

Name	Type	Path
MSWindows	SERVICE_AUTO_START	"C:\WINDOWS\system32\urdvxc.exe" /service

- Services Changed:

MSWindows
MSWindows

7.c) urdvxc.exe - Network Activity

- ICMP Traffic:

ICMP Echo Requests sent to 26 hosts

ICMP Echo Replies received from 26 hosts

Scanned a Subnet: 61.229.0.0/16

- Unknown TCP Traffic:

from ANUBIS:1328 to 61.229.113.109:445

State: Connection established, not terminated - Transferred outbound Bytes: 172 - Transferred inbound Bytes: 0

Data sent:

```
0000 00a8 ff53 4d42 7200 0000 0008 0140      .....SMBr.....@  
0000 0000 0000 0000 0000 0000 0000 8000
```

- TCP Connection Attempts:

from ANUBIS:1040 to 61.229.113.109:139

from ANUBIS:1039 to 61.229.82.160:139

from ANUBIS:1038 to 61.229.54.57:139

from ANUBIS:1041 to 61.229.118.248:139

from ANUBIS:1042 to 61.229.218.221:139

Data Tainting in Anubis

Powerful technique for tracing data flows of a program

- E.g. how network data is processed by a program
- E.g. it enables us to find out if malware uses random file names for infection only during one single analysis run

How does tainting work?

- Performed on hardware level, invisible for analyzed malware
- Data elements of interest are labeled (tainted)
- When memory values are copied, taint labels (information) are maintained allowing us to **identify the data flow** process

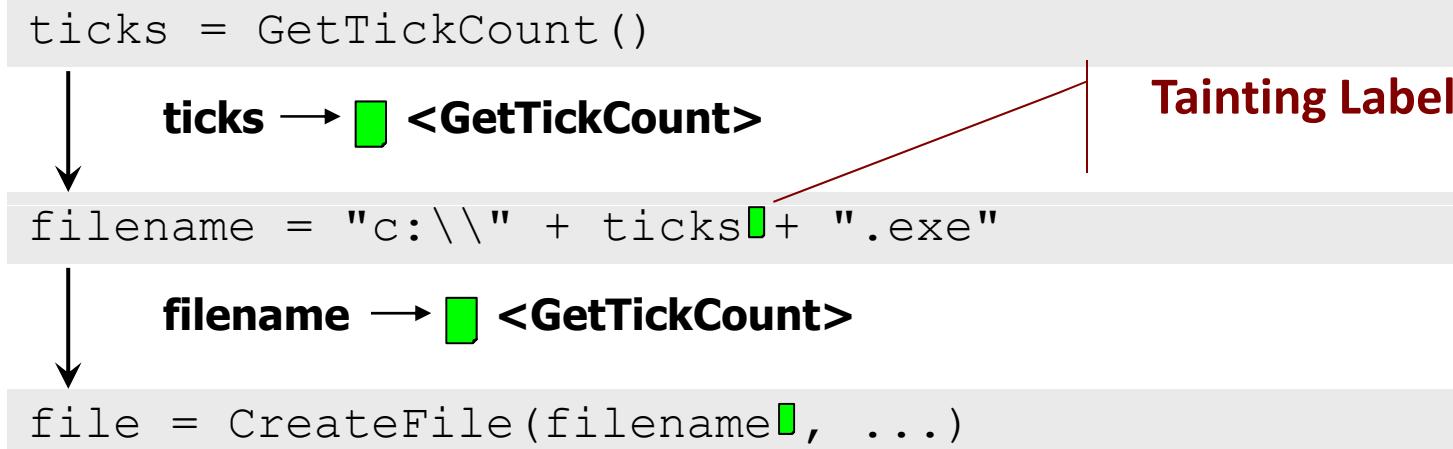
Memory Tainting Example

Consider the following code fragment

```
ticks = GetTickCount()  
filename = "c:\\\" + ticks + ".exe"  
file = CreateFile(filename, ...)
```

Creates file with random name

Enhanced with tainting information



=> `CreateFile` is called with a random filename

Resume so far

By now we have achieved the following

- We can automatically analyze single malware samples
- We know within 4 min. if this sample is malicious or not
- We can provide a non-obtrusive view from outside on our malware's behavior

But we still have the following challenges

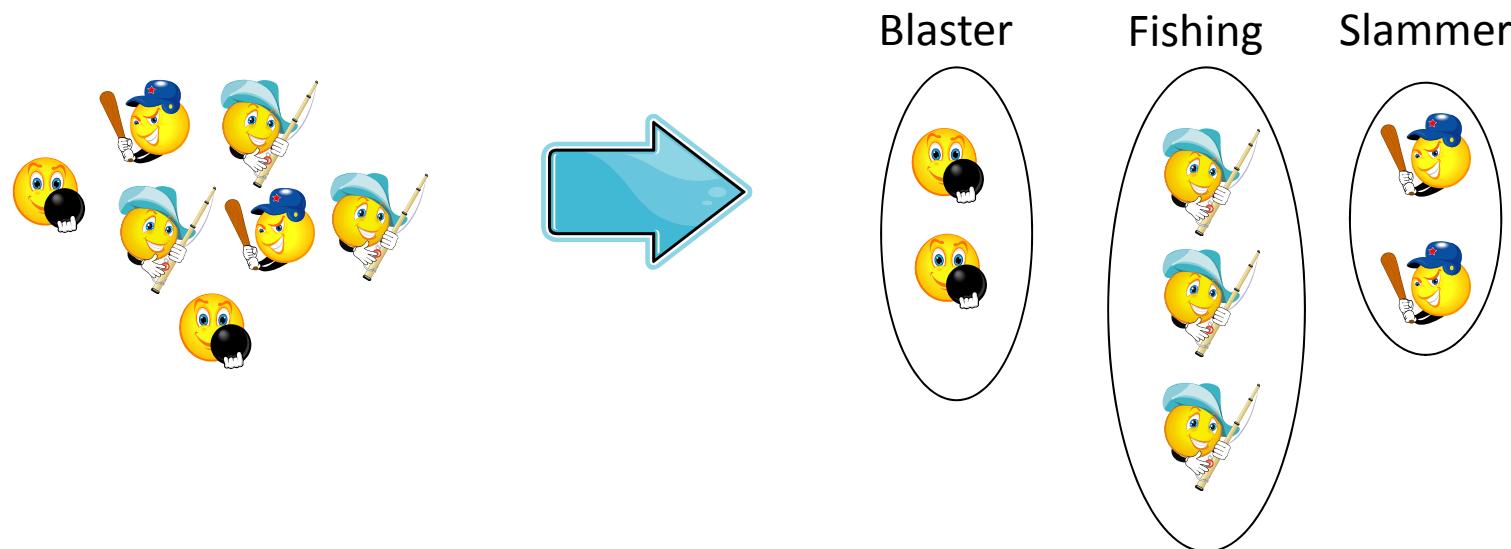
- How to structure thousands of generated analysis reports?
- Wouldn't it be nice to know (for every new incoming sample) if it belongs to a well-known malware family?

ANUBIS can also provide this additional information

- This feature is called "clustering"

Scalable, Behavior-Based Malware Clustering

Malware Clustering: Find a partitioning of a given set of malware samples into subsets so that subsets share some common traits (i.e., find “virus families”)



Malware Clustering – Features

Behavior-based

- Samples are clustered according to their behavior exhibited at runtime
- Requires prior analysis by Anubis

Scalable

- Use of LSH (Locality Sensitive Hashing) allows us to avoid computing all $n^2/2$ distances
- Suitable for clustering real-world malware collections

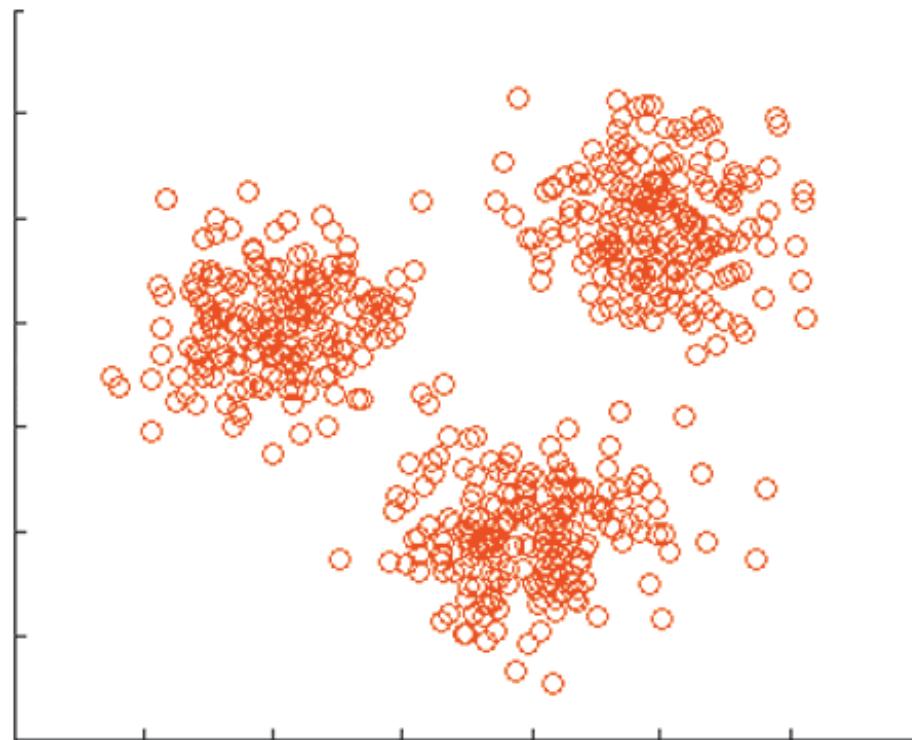
Details

- Ulrich Bayer, Paolo Milani, Clemens Hlauschek, Christopher Kruegel, and Engin Kirda: *Scalable, Behavior-Based Malware Clustering, NDSS 2009, San Diego, February 2009*

How about clustering 825k samples...

...in less than 8 hours?

- Most recent clustering run (August 16^h 2009):
- http://anubis.iseclab.org/?action=browse_clusters&task=299



Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730539

Number of Clusters:

91521

Local Sensitive Hashing Parameters

Distance Threshold	=	0.2
I	=	87
k	=	20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled-	2% trojan-dropper.agent	2% virus.win32.virut
6825011	92488	74% net-worm.win32.allapple	25% -unlabeled-	0% virus.win32.cheburgen
6823539	49148	61% -unlabeled-	14% trojan.generic	9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled-	10% trojan-clicker.html.iframe	8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled-	9% trojan-dropper.agent	7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple	17% -unlabeled-	0% backdoor.rbot
6830127	17440	98% -unlabeled-	0% trojan-downloader.win32.autoit	0% packed.win32.kclone
6797651	16042	58% backdoor.win32	40% -unlabeled-	
6818183	15483	69% -unlabeled-	25% trojan-downloader.win32	
6788535	14495	91% net-worm.win32.allapple	9% -unlabeled-	0% virus.win32.cheburgen

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:
827377

Anubis Tasks:
998505
Unique Behavioral Profiles:
730539
Number of Clusters:
91521

Local Sensitive Hashing Parameters

Distance Threshold = 0.2
l = 87
k = 20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled- 2% trojan-dropper.agent 2% virus.win32.virut		
6825011	92488	74% net-worm.win32.allapple 25% -unlabeled- 0% virus.win32.cheburgen		
6823539	49148	61% -unlabeled- 14% trojan.generic 9% not-a-virus:.webtoolbar		
6789759	31984	50% -unlabeled- 10% trojan-clicker.html.iframe 8% trojan-clicker.js.agent		
6798627	25494	47% -unlabeled- 9% trojan-dropper.agent 7% virus.worm.win32.socks		
6830115	24639	82% net-worm.win32.allapple 17% -unlabeled- 0% backdoor.rbot		
6830127	17440	98% -unlabeled- 0% trojan-downloader.win32.autoit 0% packed.win32.kclone		
6797651	16042	58% backdoor.win32 40% -unlabeled-		
6818183	15483	69% -unlabeled- 25% trojan-downloader.win32		
6788535	14495	91% net-worm.win32.allapple 9% -unlabeled- 0% virus.win32.cheburgen		

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730539

Number of Clusters:

91521

Local Sensitive Hashing Parameters

Distance Threshold	= 0.2
l	= 87
k	= 20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled-	2% trojan-dropper.agent	2% virus.win32.virut
6825011	92488	74% net-worm.win32.allapple	25% -unlabeled-	0% virus.win32.cheburgen
6823539	49148	61% -unlabeled-	14% trojan.generic	9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled-	10% trojan-clicker.html.iframe	8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled-	9% trojan-dropper.agent	7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple	17% -unlabeled-	0% backdoor.rbot
6830127	17440	98% -unlabeled-	0% trojan-downloader.win32.autoit	0% packed.win32.kclone
6797651	16042	58% backdoor.win32	40% -unlabeled-	
6818183	15483	69% -unlabeled-	25% trojan-downloader.win32	
6788535	14495	91% net-worm.win32.allapple	9% -unlabeled-	0% virus.win32.cheburgen

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730530

Number of Clusters:

91521

Local Sensitive Hashing Parameters

Distance Threshold	=	0.2
l	=	87
k	=	20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled- 2% trojan-dropper.agent 2% virus.win32.virut		
6825011	92488	74% net-worm.win32.allapple 25% -unlabeled-		0% virus.win32.cheburgen
6823539	49148	61% -unlabeled- 14% trojan.generic		9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled- 10% trojan-clicker.html.iframe		8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled- 9% trojan-dropper.agent		7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple 17% -unlabeled-		0% backdoor.rbot
6830127	17440	98% -unlabeled- 0% trojan-downloader.win32.autoit		0% packed.win32.kclone
6797651	16042	58% backdoor.win32 40% -unlabeled-		
6818183	15483	69% -unlabeled- 25% trojan-downloader.win32		
6788535	14495	91% net-worm.win32.allapple 9% -unlabeled-		0% virus.win32.cheburgen

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730539

Number of Clusters:

91521

Local Sensitive Hashing Parameters

Distance Threshold	=	0.2
I	=	87
k	=	20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled- 2% trojan-dropper.agent 2% virus.win32.virut		
6825011	92488	74% net-worm.win32.allapple 25% -unlabeled-		0% virus.win32.cheburgen
6823539	49148	61% -unlabeled- 14% trojan.generic		9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled- 10% trojan-clicker.html.iframe		8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled- 9% trojan-dropper.agent		7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple 17% -unlabeled-		0% backdoor.rbot
6830127	17440	98% -unlabeled- 0% trojan-downloader.win32.autoit		0% packed.win32.kclone
6797651	16042	58% backdoor.win32 40% -unlabeled-		
6818183	15483	69% -unlabeled- 25% trojan-downloader.win32		
6788535	14495	91% net-worm.win32.allapple 9% -unlabeled-		0% virus.win32.cheburgen

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730539

Number of Clusters:

91521

Local Sensitive Hashing Parameters

Distance Threshold	=	0.2
I	=	87
k	=	20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled-	2% trojan-dropper.agent	2% virus.win32.virut
6825011	92488	74% net-worm.win32.allapple	25% -unlabeled-	0% virus.win32.cheburgen
6823539	49148	61% -unlabeled-	14% trojan.generic	9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled-	10% trojan-clicker.html.iframe	8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled-	9% trojan-dropper.agent	7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple	17% -unlabeled-	0% backdoor.rbot
6830127	17440	98% -unlabeled-	0% trojan-downloader.win32.autoit	0% packed.win32.kclone
6797651	16042	58% backdoor.win32	40% -unlabeled-	
6818183	15483	69% -unlabeled-	25% trojan-downloader.win32	
6788535	14495	91% net-worm.win32.allapple	9% -unlabeled-	0% virus.win32.cheburgen

Anubis Clustering Task 299

Cluster Task Id:	299
Create Time:	2009-08-16 10:11:24
Start Time:	2009-08-16 10:15:38
End Time:	2009-08-16 18:09:23
Run Time:	07:53:45
Peak Virtual Memory Size:	21.53 Gb
Peak Resident Set Size:	18.74 Gb
Samples were submitted between:	2007-02-07 13:44:00 - 2009-08-16 16:58:53

Number of Samples:

827377

Anubis Tasks:

998505

Unique Behavioral Profiles:

730539

Number of Clusters:

91521

Local Sensitive Hashing Parameters

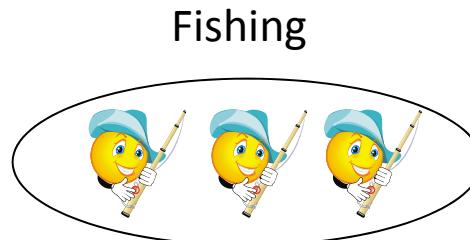
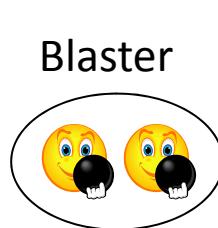
Distance Threshold	=	0.2
I	=	87
k	=	20

Anubis Families 1 - 10 of 91521

Family Id	# Samples	Top 3 A/V Labels		next»
6824791	116995	53% -unlabeled-	2% trojan-dropper.agent	2% virus.win32.virut
6825011	92488	74% net-worm.win32.allapple	25% -unlabeled-	0% virus.win32.cheburgen
6823539	49148	61% -unlabeled-	14% trojan.generic	9% not-a-virus:.webtoolbar
6789759	31984	50% -unlabeled-	10% trojan-clicker.html.iframe	8% trojan-clicker.js.agent
6798627	25494	47% -unlabeled-	9% trojan-dropper.agent	7% virus.worm.win32.socks
6830115	24639	82% net-worm.win32.allapple	17% -unlabeled-	0% backdoor.rbot
6830127	17440	98% -unlabeled-	0% trojan-downloader.win32.autoit	0% packed.win32.kclone
6797651	16042	58% backdoor.win32	40% -unlabeled-	
6818183	15483	69% -unlabeled-	25% trojan-downloader.win32	
6788535	14495	91% net-worm.win32.allapple	9% -unlabeled-	0% virus.win32.cheburgen

Clustering Workflow

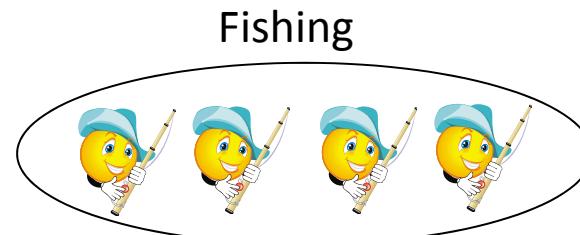
1) Periodic (e.g., weekly) full cluster runs:



2) Nearest neighbor search for each new sample:



nearest neighbor search



Handyman



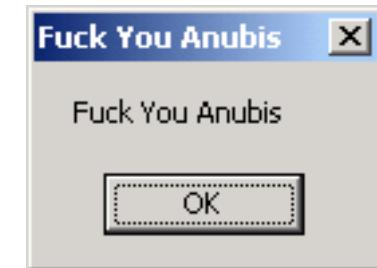
Lessons learned from 2 Years ANUBIS

Bot Analysis

- Bot analysis and IP address blacklisting become a problem
- Bot herders know IP range of public version of ANUBIS

ANUBIS Detection and Evasion

- Currently we've seen about 0,03% samples ITW with ANUBIS detection capabilities
- ANUBIS is capable of detecting if malware tries to evade ANUBIS



Some general Sandbox Problems

Timeout issues (general to automated sandbox analysis)

- Timeouts, how long shall the analysis run?
- Automatic analysis has to quit at some point (when?)

Most recent timeout problems

- Analysis of Mebroot malware resulted in empty ANUBIS logs
- Mebroot waits several minutes before infecting the system
- Watch out for empty logs!
- Timeout can not be altered in public online version (but in the in-house version this value is customizable)

Malware waiting for some user interaction

- Mouse movement/clicks, keystrokes, certain URL to be loaded

Packer with Anti-ANUBIS Features



Conclusion

ANUBIS offers technology to speed up malware analysis

- Automatic processing of incoming samples saves valuable time
- ANUBIS improves traditional analysis process flow with its features
- Clustering feature is unique to ANUBIS (AFAIK)
- Can offer **additional functionality for “in the cloud” services**
(already used in academic research projects like WOMBAT/SGNET)
See paper for more info on that.

Public version vs. commercial version

- Commercial version available on request
- Offers more features and **keeps your samples in-house**
- Offers **customization** (language, VM OS, 3rd party apps, ...)
- Offers integration into your existing pre-sorting process flow

Questions



**Thank you for your attention!
We'd be happy to answer all of your questions!**

**Please send your questions to: anubis@ikarus.at,
anubis@iseclab.org or thomas.mandl@mandl-itc.at**