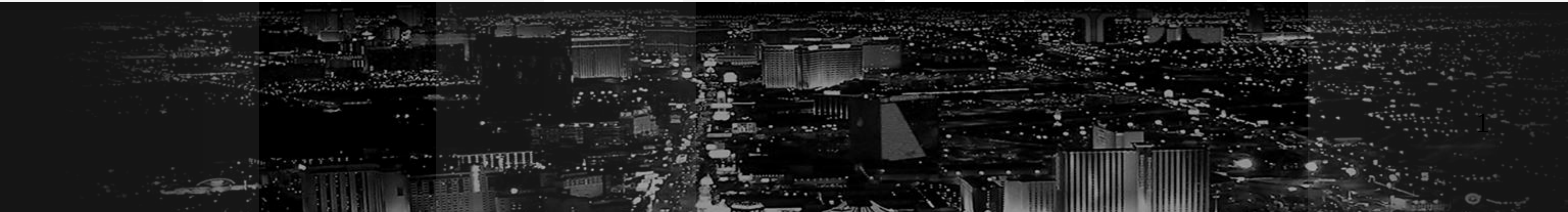
ZERO NIGHTS 2015

DISTRIBUTING THE RECONSTRUCTION OF HIGH-LEVEL INTERMEDIATE REPRESENTATION FOR LARGE SCALE MALWARE ANALYSIS

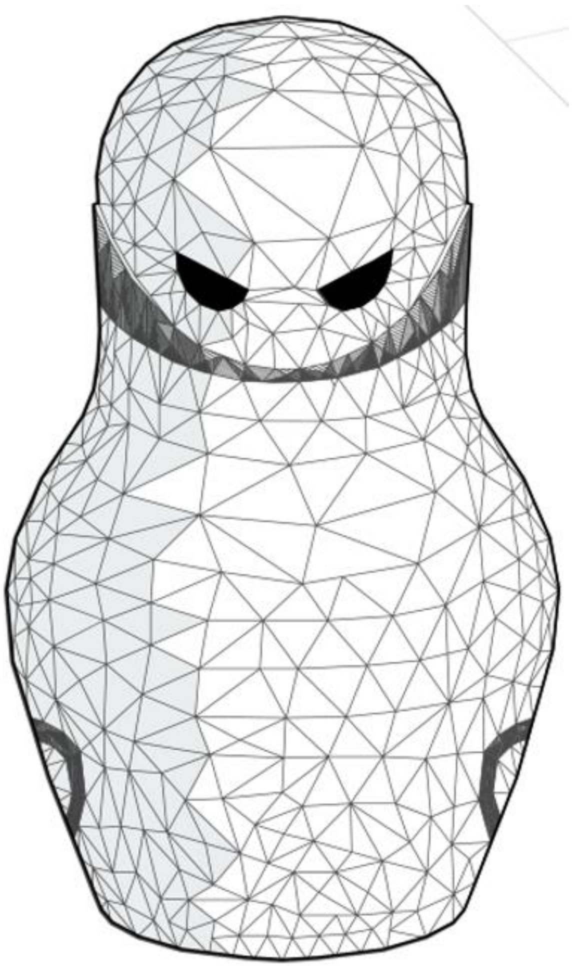
Alexander Matrosov (@matrosov)
Eugene Rodionov (@vxradius) [1]
Gabriel Negreira Barbosa (@gabrielnb)
Rodrigo Rubira Branco (@BSDaemon)

{alexander.matrosov || gabriel.negreira.barbosa || rodrigo.branco}
noSPAM intel.com
[1] rodionov *noSPAM* eset.com



Disclaimer

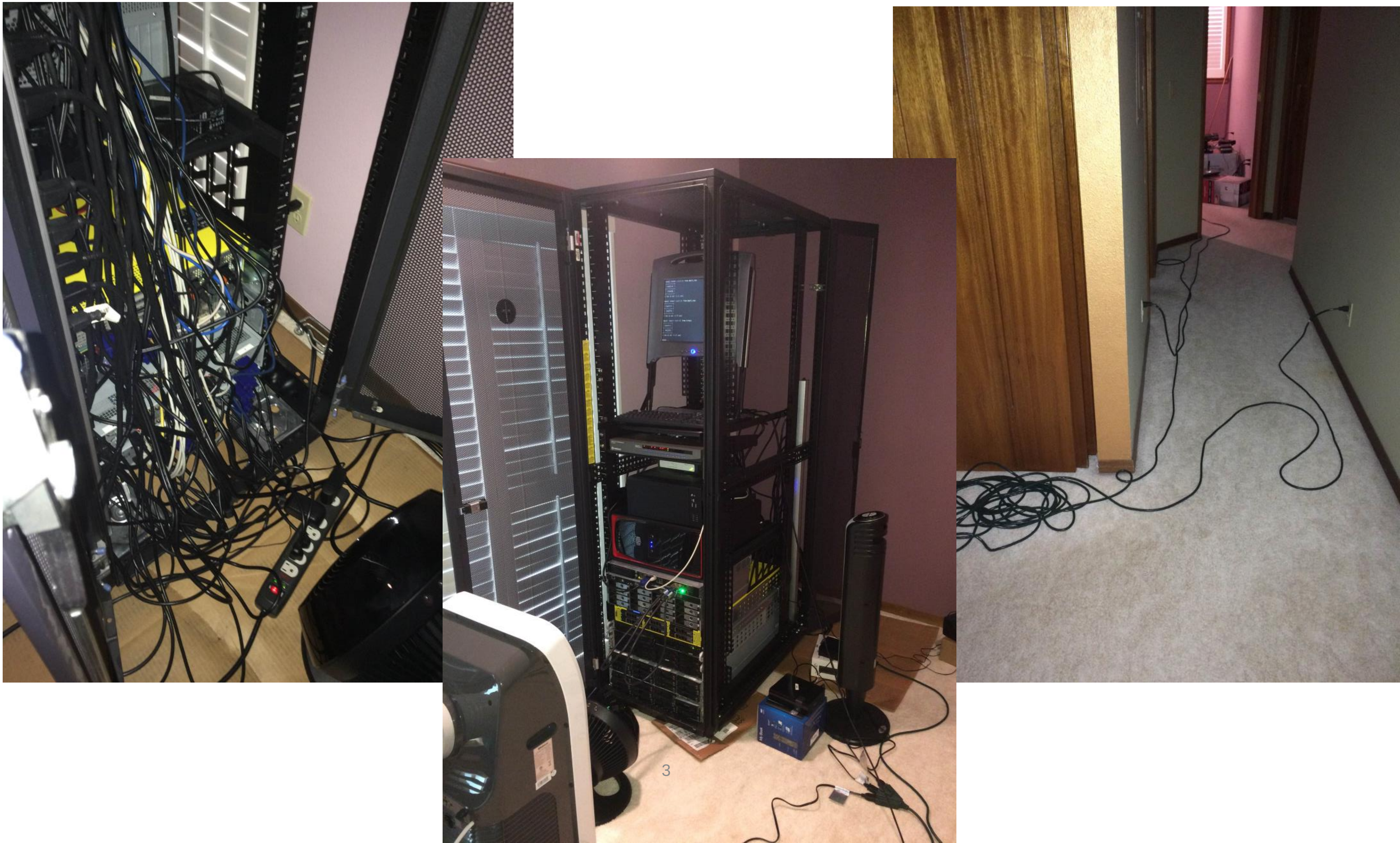
We don’t speak for our employer. All the opinions and information here are of our responsibility (actually no one ever saw this talk before).



So, mistakes and bad jokes are all

OUR responsibilities

Thanks to the smoke and fire detection mechanism :)



Introduction / Motivation

- Number of new malware samples grows at an absurd pace

- We still see words such as ‘many’ instead of the actual number of analyzed samples

- Assumptions without concrete data supporting them

- **INDUSTRY-RELATED RESEARCH NEEDS RESULTS, THUS NOT PROMISING POINTS ARE NOT LOOKED AFTER**

Objectives

- **Demonstrate** the possibility of in-depth large-scale malware analysis

- **Distribute and scale** IDA Pro (with Decompiler) to leverage its functionalities for automated malware analysis

- **Share with the community** the obtained results:
  - ✓ IDA Pro IDBs, plugins and scripts
  - ✓ Intermediate representation
  - ✓ MS Visual C++ reconstructed types
  - ✓ And more...

Methodology: Highlights

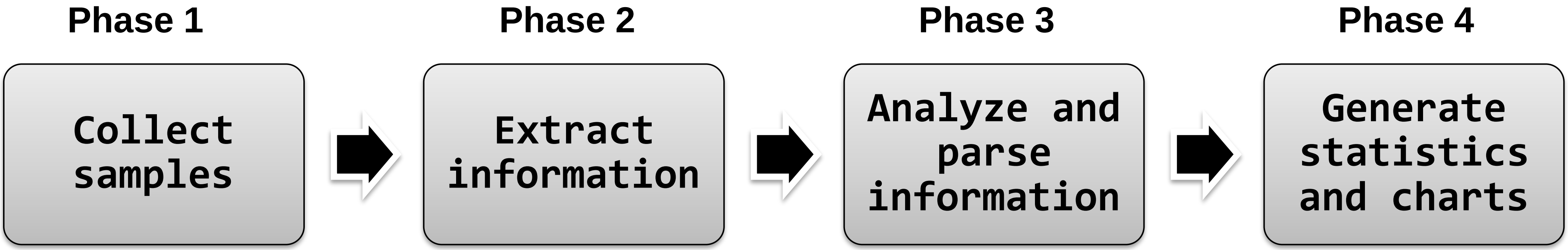
- **Analyzed 32-bit and x86-64-bit PE not-packed samples from public sources**

- **No malware size limitations at all**

- **Preference on MS Visual C++ samples because of HexRaysCodeXplorer OO types reconstruction feature**

- **Details on the infrastructure already discussed in Black Hat Las Vegas 2012 presentation**

Methodology: Overview of the process



Pre-process samples and collect millions of 32-bit and x86-64-bit not-packed PE malware samples

Run different malware analysis algorithms on the collected samples and store results on the filesystem.

Parse and structure the results.

Generate statistics and charts based on structured information.

Methodology: Only static analysis

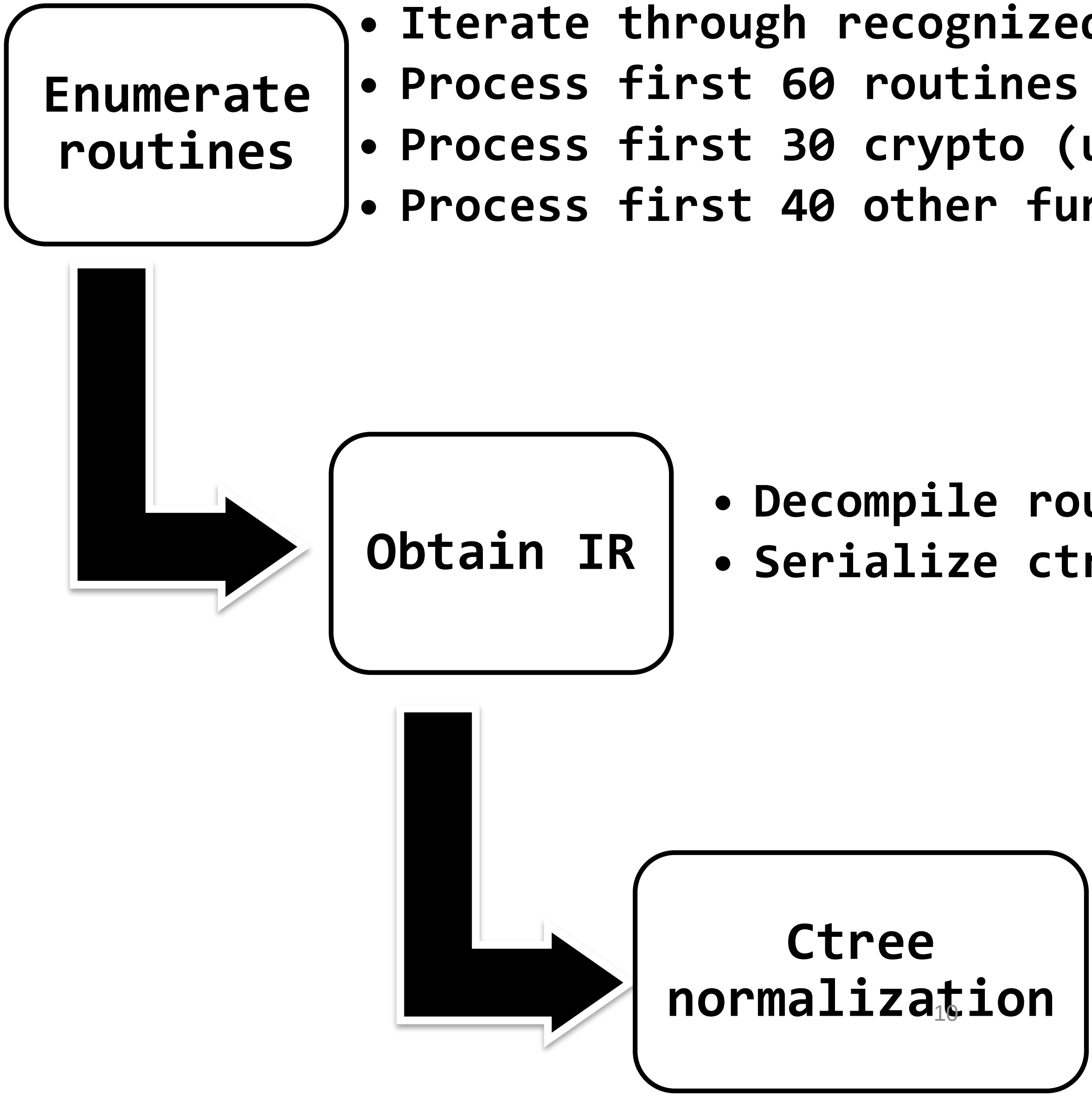
- We only used static analysis
- Not detectable by malware… unless it exploits the analysis environment!
- Prone to anti-disassembly tricks
- Has some limitations… but powerful tools and techniques are available
- IDA Pro rocks!! ☺

Methodology: Malware analysis algorithms

- **HexRaysCodeXplorer (by @REhints) used for:**
  - Ctrees* for some IDA-recognized functions
  - MS Visual C++ object-oriented types REconstruction
- **Ctrees depth analysis**
  - Highly-modified version of pathfinder by @devttyS0
- **OO “this” usage study**
- **Crypto usage detection based on IdaScope by @push_pnx**

** - ctrees is the intermediate representation in Hex-Rays decompiler*

Constraints and Limitations: Dumping Ctrees



Enumerate routines

- Iterate through recognized routines in idb
- Process first 60 routines of size larger than 0x160 bytes
- Process first 30 crypto (using AES-NI) routines
- Process first 40 other functions bigger than 0x60 bytes

Obtain IR

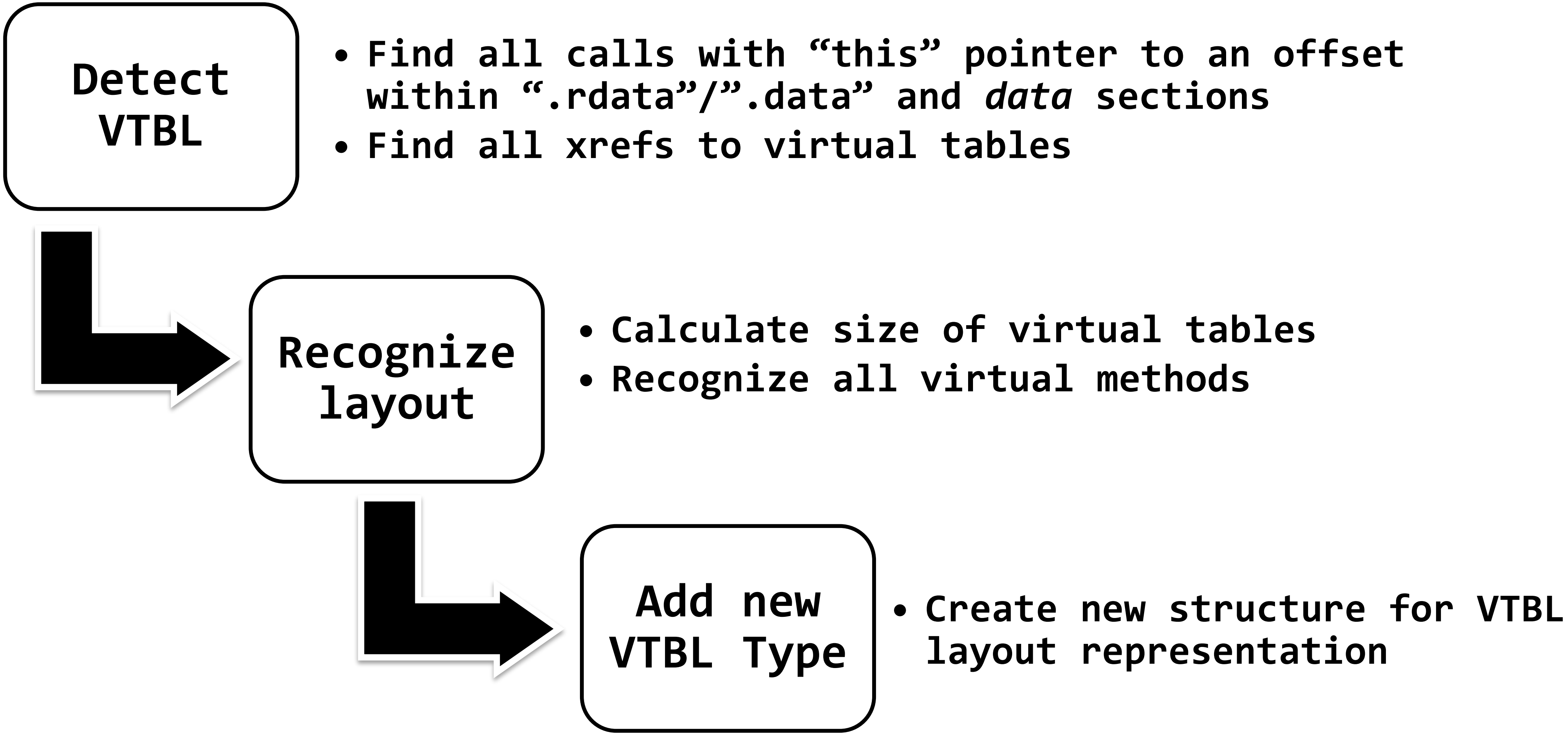
- Decompile routine to get ctree (IR)
- Serialize ctree to string

Ctree normalization

- See implementation of *ctree_dumper_t::filter_citem()*
- Use normalized ctree for comparison

Constraints and Limitations:

VTBL reconstruction algorithm



Detect VTBL

- Find all calls with “this” pointer to an offset within “.rdata”/”.data” and *data* sections
- Find all xrefs to virtual tables

Recognize layout

- Calculate size of virtual tables
- Recognize all virtual methods

Add new VTBL Type

- Create new structure for VTBL layout representation

Constraints and Limitations :
Complex types REconstruction algorithm

Detect Type

- Find pointers to possible type instances
- Find initialization routine entry point

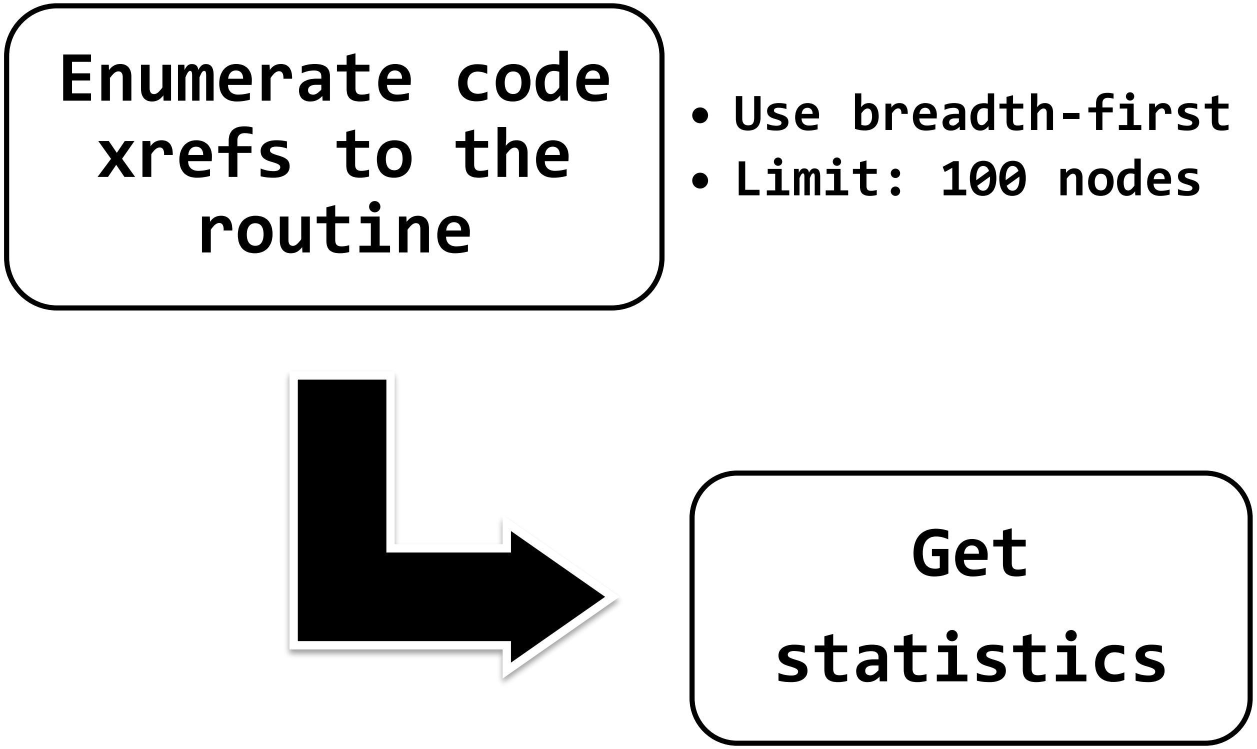
Recognize Type layout

- Find all references to possible type address space
- Find all xrefs to the attributes of the identified type
- Reconstruct data flow for the identified type

Add new Type definition

- Create new local type if it has more than 3 attributes

Constraints and Limitations:
Ctrees Depth Analysis



Enumerate code xrefs to the routine

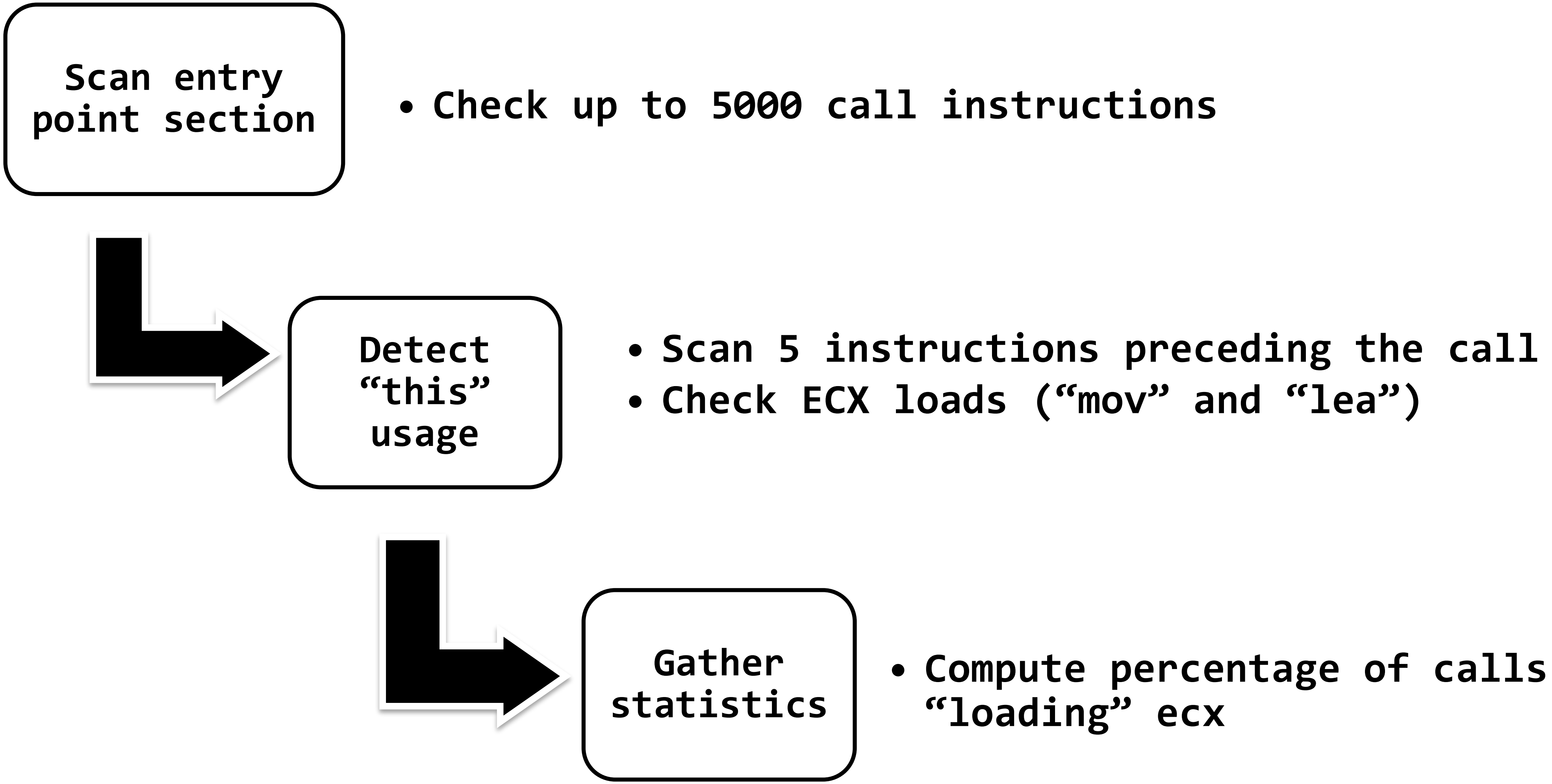
- Use breadth-first search algorithm
- Limit: 100 nodes

Get statistics

- Distance from entry point
- depth counter
- number of xrefs

Constraints and Limitations:

C++ “this” usage study



Scan entry point section

- Check up to 5000 call instructions

Detect “this” usage

- Scan 5 instructions preceding the call
- Check ECX loads (“mov” and “lea”)

Gather statistics

- Compute percentage of calls “loading” ecx

Distributing IDA Pro: Highlights

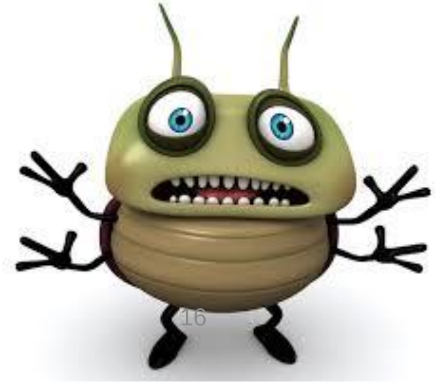
- **Unexpected performance benefits on IDA because the information is structured**
  - ✓ But we also came across some disadvantages: SDK is complex, function signatures change from version to version and is not fully documented

- **Good performance in commodity hardware**

- **C-based plugins are usually not compatible with Linux/Mac**
  - ✓ Portability efforts are required

Distributing IDA Pro: Highlights

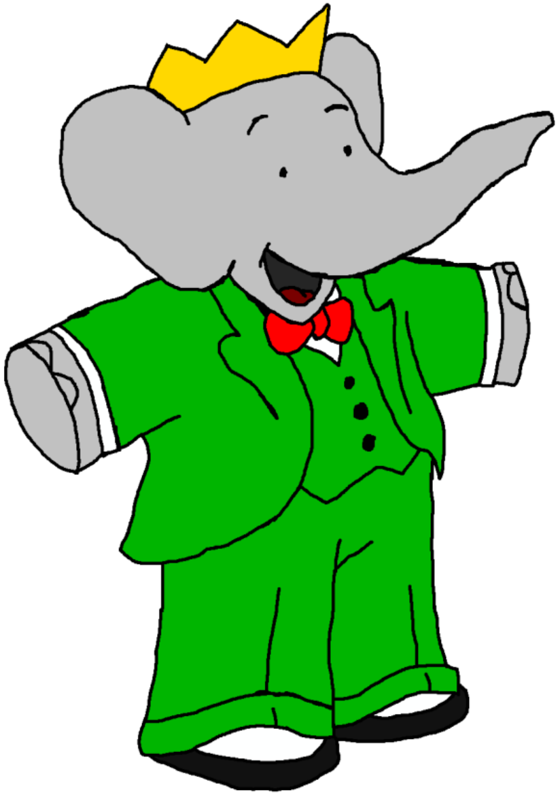
- IDA plugins are usually not made to scale
  - Target single-sample analysis
  - Focus on users interacting with IDA Pro interface
- Automated malware analysis exercises much more the internal plugin flows than manual analysis
  - ✓ As a result, corner cases and bugs were identified in many plugins including HexRaysCodeXplorer

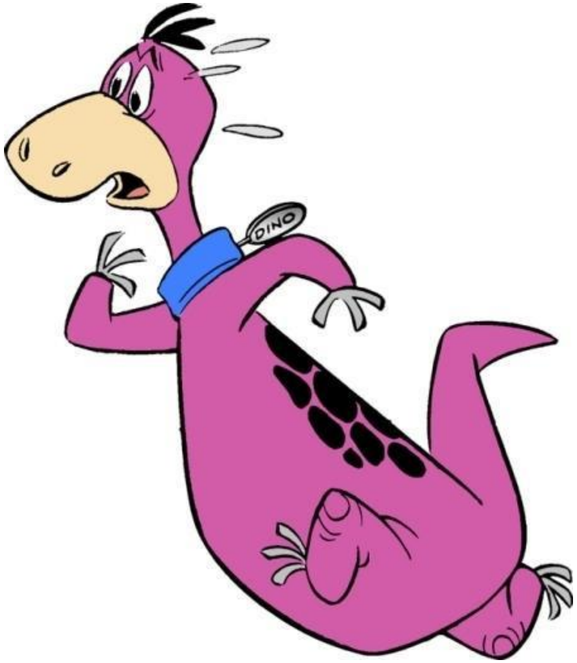


VALIDATING THE METHODOLOGY AND TOOLSET

ANALYSIS OF C++ TARGETED MALWARE

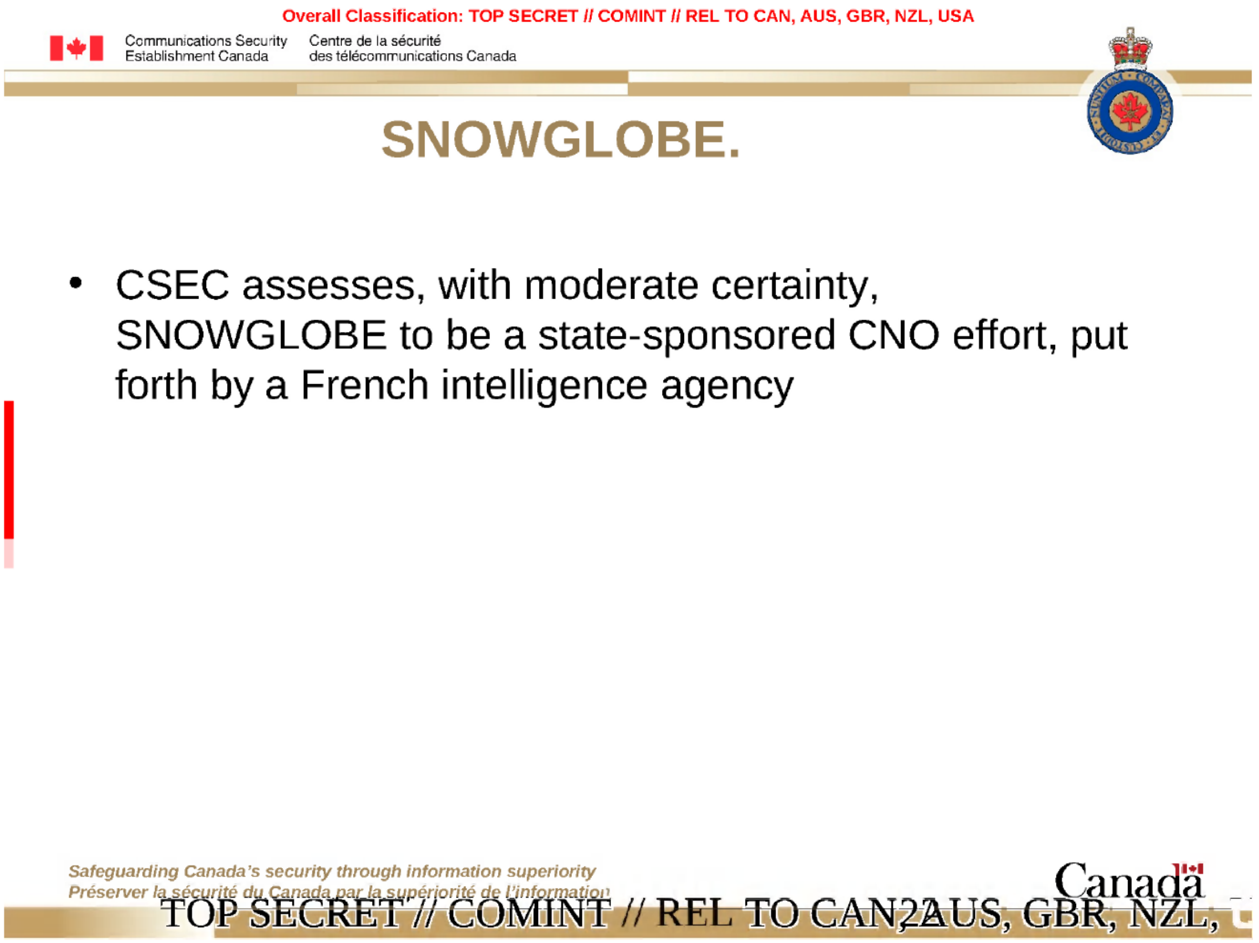
Animal Farm Case Study





Animal Farm* Case Study

- Discovered by CSEC as operation SNOWGLOBE
- Samples: NBOT, Dino, Babar, Bunny, Casper
- Written in MS Visual C++



Overall Classification: TOP SECRET // COMINT // REL TO CAN, AUS, GBR, NZL, USA

Communications Security Establishment Canada

Centre de la sécurité des télécommunications Canada

SNOWGLOBE.

- CSEC assesses, with moderate certainty, SNOWGLOBE to be a state-sponsored CNO effort, put forth by a French intelligence agency

Safeguarding Canada's security through information superiority
Préserver la sécurité du Canada par la supériorité de l'information

Canada

TOP SECRET // COMINT // REL TO CAN, AUS, GBR, NZL,

** - "Totally Spies", Joan Calvet, Marion Marschalek, Paul Rascagnères, http://recon.cx/2015/slides/recon2015-01-joan-calvet-marion-marschalek-paul-rascagneres-Totally-Spies.pdf*

Animal Farm: Shared C++ Types

	NBOT	Casper	Bunny	Babar	Dino
wmiException	X		X	X	
basic_AvWmiManager	X		X	X	
basic_WmiManager	X		X	X	
CTFC_HTTP_Form	X	X			X
CTFC_HTTP_Forms	X	X			X
CTFC_HTTP_Form_Multipart	X	X			X
CTFC_HTTP_Request	X	X			X
CTFC_AbstractSocket	X	X			X
CTFC_StandardSocket	X	X			X
RunKeyApi		X			X
RunKeyBat		X			X
RunKeyReg		X			X
RunKeyWmi		X			X
RunKeyDefault		X			X
AutoDelApi		X			X
AutoDelDel		X			X
AutoDelWmi		X			X
AutoDelDefault		X			X

Animal Farm: Shared C++ Types

	NBOT	Casper	Bunny	Babar	Dino
NBOT		6 shared custom types	3 shared custom types	3 shared custom types	6 shared custom types
Casper					15 shared custom types
Bunny				3 shared custom types	
Babar					
Dino					

Conclusions

- ➢ We demonstrated that IDA Pro scale really well and all its powerful features can be used in automated malware analysis systems
  - ✓ CALL TO ACTION: IDA Pro plugin developers to start adding batch mode switches and optimize the algorithms

- ➢ Want to run your IDA plugin on millions of malwares? Let us know! ☺

Resources

Presentation, code and instructions on how to download samples, IDBs and outputs will be available at:

https://github.com/REhints/BlackHat_2015

CodeXplorer v2.0 [BH Edition]

- Finally plugin support Linux/Mac/Windows

- Options for analysis in IDA batch mode
- Multiple bug fixes and code review
- Improvements for Types and VTBL’s reconstruction
- New Features:
  - ✓ dump Ctrees information for additional analysis
  - ✓ dump all reconstructed types information

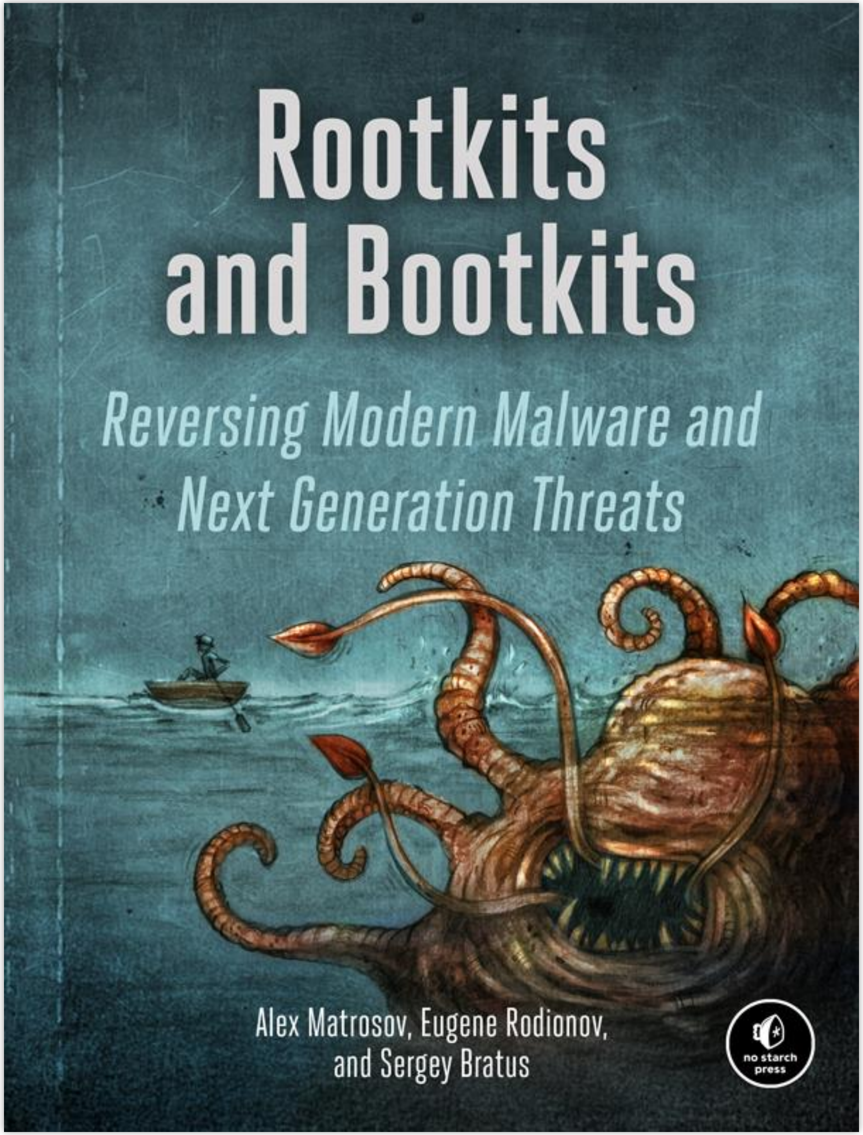
https://github.com/REhints/HexRaysCodeXplorer

Acknowledgements

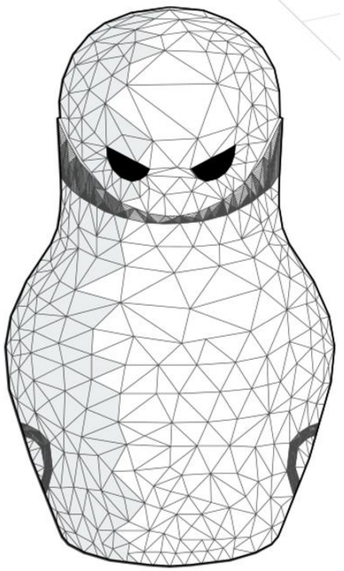
Personally to **Ilfak Guilfanov (@ilfak)** and **Hex-Rays team** for supporting this research

All the researchers releasing malware-related techniques!!!

The new RE book is coming soon!


Rootkits and Bootkits
Reversing Modern Malware and Next Generation Threats
Alex Matrosov, Eugene Rodionov, and Sergey Bratus
no starch press

https://www.nostarch.com/rootkits



THE END ! Really !?

Alexander Matrosov (@matrosov)
Eugene Rodionov (@vxradius) [1]
Gabriel Negreira Barbosa (@gabrielnb)
Rodrigo Rubira Branco (@BSDaemon)

{alexander.matrosov || gabriel.negreira.barbosa || rodrigo.branco}
noSPAM intel.com
[1] rodionov *noSPAM* eset.com